INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

TRANSCRIPTIONAL REPRESSION BY ADIPOCYTE ENHANCER BINDING PROTEIN 1

by

Peter J. Lyons

Submitted in partial fulfillment of the requirements for
the degree of
Doctor of Philosophy

at
Dalhousie University
Halifax, Nova Scotia

August, 2005

DALHOUSIE UNIVERSITY

To comply with the Canadian Privacy Act the National Library of Canada has requested that the following pages be removed from this copy of the thesis:

Preliminary Pages
- Examiners Signature Page (pii)
- Dalhousie Library Copyright Agreement (piii)

Appendices
- Copyright Releases (if applicable)

TABLE OF CONTENTS

	Page
List of Figures	viii
List of Tables	x
Abstract	xi
List of Abbreviations	xii
Acknowledgments	xiv
1. Introduction	1
1.1. Molecular mechanisms of adipogenesis	3
1.1.1. Transcriptional regulation of adipogenesis	3
1.1.2. Regulation of adipogenesis through MAP kinase	6
1.1.3. Regulation of adipogenesis through the extracellular matrix	8
1.2. Insights into the role of AEBP1 from knockout mouse models.	9
1.3. Mechanisms of Transcriptional Regulation	10
1.3.1. Nucleosome Remodeling	12
1.3.2. Variant Histones	12
1.3.3. Histone Post-Translational Modifications	13
1.3.3.1. Acetylation	14
1.3.3.2. Methylation	16
1.3.3.3. Ubiquitination and SUMOylation	17
1.3.3.4. Phosphorylation	18
1.3.3.5. N-linked phosphorylation	18
1.3.4. Possible Mechanism for Transcriptional Repression by AEBP1	19
1.4. The Metallopeptidase Family	20
1.4.1. Zincins	22
1.4.2. Carboxypeptidases	23
1.4.2.1. The A/B Carboxypeptidase Subfamily	23
1.4.2.2. The N/E Carboxypeptidase Subfamily	26

1.4.2.3. The Non-Traditional Carboxypeptidases 28
1.5. Goals of this Research 32
2. Materials and Methods 33
2.1. Computational Methods 33
2.1.1. Homology Modeling 33
2.1.2. Sequence Analysis 34
2.2. Recombinant DNA Methods 34
2.2.1. Plasmid DNA Construction 34
2.2.2. Transformation of Competent *E. coli* Cells 35
2.2.3. Site-directed Mutagenesis 41
2.3. Purification of Recombinant Proteins 43
2.3.1. AEBP1 43
2.3.2. GST fusion proteins 44
2.4. Mammalian Cell Culture Methods 45
2.4.1. Cell line Maintenance 45
2.4.2. Transfection 45
2.5. Protein Analysis 46
2.5.1. Preparation of Extracts 46
2.5.2. Protein Concentration Determination 46
2.5.3. SDS-PAGE 47
2.5.4. Protein Detection by Western Blotting or staining 47
2.6. Assays 48
2.6.1. β-Galactosidase Assay 48
2.6.2. Chloramphenicol Acetyl Transferase (CAT) Assay 48
2.6.3. Luciferase Assay 49
2.6.4. Electrophoretic Mobility Shift Assay (EMSA) 49
2.6.5. GST Pull-Down Assays 50
2.6.6. CaM-Agarose Pull-Down Assays 50
2.6.7. Kinase Assay 51
3. Results 52
3.1. Structural Analysis of AEBP1 52

3.1.1. DLD domain modeling 52

3.1.2. CP domain modeling 62

3.1.3. Secondary structure of the C-terminus and unmodelled loops within the CP domain of AEBP1. 72

3.1.4. AEBP1 orthologue analysis 74

3.2. Expression of Full-Length and Truncated AEBP1 89

3.2.1. Expression and purification of recombinant AEBP1 in a bacterial system 89

3.2.2. Expression of AEBP1 in mammalian cells. 101

3.3. Function of the CP domain of AEBP1 105

3.3.1. The CP domain of AEBP1 interacts with C-terminal motifs. 108

3.3.2. AEBP1 interacts with Ca^{2+}/calmodulin in a Ca^{2+}/calmodulin-dependent histone H3 arginine kinase complex. 111

3.4. Phosphorylation of AEBP1 116

3.4.1. AEBP1 is phosphorylated by MAP kinase in CHO cells upon EGF stimulation. 116

3.4.2. AEBP1 S668 mutation eliminates phosphorylation of AEBP1 by MAP kinase *in vivo*. 118

3.4.3. AEBP1 S668 mutation eliminates phosphorylation by MAP kinase *in vitro*. 122

3.4.4. Function of AEBP1 S668 phosphorylation by MAP kinase. 126

3.5. DNA binding and Transcriptional Repression by AEBP1. 127

3.5.1. Both N- and C-termini of AEBP1 are necessary for DNA binding. 127

3.5.2. DNA binding activities of AEBP1 are inhibited by MAP kinase phosphorylation site mutation. 130

3.5.3. Transcriptional repression abilities of AEBP1 are inhibited by MAP kinase phosphorylation site mutation in a cell-type specific manner. 131

3.5.4. AEBP1 may function as a co-repressor. 133

4. Discussion 139

4.1. Structure of AEBP1 139

4.2. Function of the CP domain of AEBP1 144

4.2.1. Enzymatic mechanism of AEBP1 144

4.2.2. Role of the CP domain in a histone H3 arginine kinase complex 148
4.3. DNA binding by AEBP1 153
4.4. Phosphorylation of AEBP1 by MAP Kinase 160
5. Conclusion 163

Appendix A. Multiple alignment of all identified AEBP1 orthologues. 167
Appendix B. Assembly of chicken AEBP1 EST clones. 170

References 175

LIST OF FIGURES

	Page
Figure 1. Domain structure of AEBP1 and ACLP.	2
Figure 2. Progression of 3T3-L1 preadipocyte differentiation.	5
Figure 3. Histone modifications on the nucleosome core particle.	15
Figure 4. Representative members of four classes of metallopeptidases.	21
Figure 5. Proposed tetrahedral transition state in peptide-bond hydrolysis by CPA and related carboxypeptidases.	25
Figure 6. A multiple alignment of representative active and inactive carboxypeptidases.	30
Figure 7. Alignment of the amino acid sequences of the DLD domain of AEBP1 and coagulation factor C2 domains.	54
Figure 8. Assessment of 1DLD model quality.	55
Figure 9. The DLD domain of AEBP1 is a β-barrel with two distinct faces.	60
Figure 10. Alignment of the amino acid sequences of the CP domain of AEBP1 and duck CPD-II.	63
Figure 11. Assessment of the quality of AEBP1 CP domain models.	65
Figure 12. The CP domain of AEBP1 as modeled by homology against duck carboxypeptidase D.	70
Figure 13. Secondary structure predictions for the C-terminal domain and two loops within the CP domain of AEBP1.	73
Figure 14. Multiple Alignment of AEBP1 homologues.	75
Figure 15. Phylogenetic comparison of AEBP1, CPX-1, and CPX-2.	85
Figure 16. Comparison of proteins encoded by pET16b-AEBP1 and pET21d-AEBP1.	90
Figure 17. Comparison of AEBP1 expressed from pET16b-AEBP1 and pET21d-AEBP1 in Origami B *E. Coli.*	93
Figure 18. Optimization of AEBP1 expression in *E. coli.*	94
Figure 19. Optimal concentrations of IPTG result in high level expression of AEBP1 in inclusion bodies.	96
Figure 20. AEBP1 and deletion derivatives expressed in and purified from *E. coli.*	98

Figure 21. Expression of wild-type and mutant versions of AEBP1 in CHO cells. 102

Figure 22. The active site cavity of AEBP1 is larger than that of CPDII. 107

Figure 23. AEBP1 interacts with PTEN through its CP domain. 109

Figure 24. AEBP1 interacts with the C-terminus of histone H3. 112

Figure 25. AEBP1 is the p85 calmodulin-interacting protein within a histone H3 arginine kinase complex. 113

Figure 26. AEBP1 interacts with Ca^{2+}/Calmodulin through its basic region. 115

Figure 27. AEBP1 is phosphorylated by ERK1/2 MAP kinase. 117

Figure 28. MAP kinase consensus phosphorylation sites within the C-terminus of AEBP1. 119

Figure 29. S668 is necessary for phosphorylation of AEBP1. 121

Figure 30. S668A mutation of AEBP1 eliminates all phosphorylation by ERK2 *in vitro*. 123

Figure 31. Truncation of AEBP1 affects *in vitro* phosphorylation by MAP kinase. 125

Figure 32. The basic domain of AEBP1 can be compared with that of the basic leucine zipper proteins. 128

Figure 33. Mutation of AEBP1 threonine 623 results in decreased DNA binding ability. 129

Figure 34. Mutation of AEBP1 threonine 623 results in decreased transcriptional repression ability in NIH/3T3 cells, but not in CHO cells. 132

Figure 35. AEBP1ΔSty is able to repress transcription of the aP2 promoter in CHO cells. 134

Figure 36. Wild-type AEBP1 binds DNA weakly through its basic region. 136

Figure 37. N- and C-terminally truncated AEBP1 repress transcription from the aP2 promoter. 138

Figure 38. Proposed model for the three-domain structure of AEBP1. 141

Figure 39. The basic helix of AEBP1 contains the sequence requirements for Ca^{2+}/calmodulin interaction. 149

Figure 40. Amino acids 386-405 of AEBP1 are similar to a protruding arm of CARM1. 151

Figure 41. A model of a possible mechanism for transcriptional repression by AEBP1. 166

List of Tables

	Page
Table 1. Plasmids used in this thesis.	36
Table 2. Primers used for plasmid construction.	40
Table 3. Primers used for site-directed mutagenesis of AEBP1.	42
Table 4. Ability of AEBP1 and deletions to interact with AE-1 duplex DNA.	156

ABSTRACT

Adipocyte Enhancer-binding Protein 1 (AEBP1) is a transcriptional repressor of the aP2 gene, which encodes the adipocyte lipid binding protein and is involved in the differentiation of preadipocytes into mature adipocytes. Other functions have also been ascribed to AEBP1 which play a part in its regulation of adipogenesis. These include the regulation of phosphorylation of the extracellular-regulated kinases 1 and 2 (ERK1/2) MAP kinases, the regulation of the stability of PTEN, a protein and lipid phosphatase, and the interaction with the γ5 subunit of a heterotrimeric G protein leading to an attenuation of the transcriptional repression activity of AEBP1. AEBP1 contains a conserved carboxypeptidase domain that is critical for its function as a transcriptional repressor.

Homology modeling and multiple alignment of AEBP1 homologues were performed to identify putative domains and critical residues that were then deleted or mutated in mouse AEBP1. Expression and purification of wild-type and mutant recombinant AEBP1 proteins were optimized. S668 in mouse AEBP1 was identified by site-directed mutagenesis as the residue critical for phosphorylation of AEBP1 by MAP kinase, both *in vitro* and *in vivo*. T623, however, is the only potential MAP kinase phosphorylation site conserved in human AEBP1, and mutagenesis of T623 followed by kinase and transcriptional assays suggested that this threonine might be phosphorylated and that phosphorylation status might affect DNA binding and transcriptional repression by AEBP1. While AEBP1 was able to bind DNA, electrophoretic mobility shift and luciferase reporter assays suggested that AEBP1 might function predominantly as a co-repressor, independent of DNA binding. A possible corepression mechanism was suggested by the identification of AEBP1 as a Ca^{2+}/calmodulin-interacting component of a histone H3 arginine kinase complex. Molecular modeling and pulldown experiments suggested the carboxypeptidase domain of AEBP1 might interact with the C-terminus of histone H3. A model for AEBP1 transcriptional repression through interaction with and subsequent phosphorylation of histone H3 in an arginine kinase complex, regulated by interaction of the C-terminus of AEBP1 with calmodulin, is presented.

LIST OF ABBREVIATIONS

ACE	angiotensin converting enzyme
ACLP	aortic carboxypeptidase-like protein
AEBP1	adipocyte enhancer binding protein 1
ALBP	adipocyte lipid binding protein
AML	acute myeloid leukemia
ARM	ATP-regulatory module
BSA	bovine serum albumin
C/EBP	CAAT/enhancer binding protein
CARM	coactivator associated arginine methyltransferase
CAT	chloramphenicol acetyltransferase
CHO	Chinese hamster ovary
CP	carboxypeptidase
DLD	discoidin-like domain
DMEM	Dulbecco's modified eagle medium
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
EGFR	epidermal growth factor receptor
EGTA	ethyleneglycol-bis(β-aminoethyl)tetraacetic acid
EMSA	electrophoretic mobility shift assay
ERK	extracellular-regulated kinase
ERα	estrogen receptor α
GEMSA	guanidinoethylmercaptosuccinic acid
GLUT	glucose transporter
GST	glutathione S transferase
HAT	histone acetyltransferase
HDAC	histone deacetylase
HMT	histone methyltransferase

IPNS	isopenicillin N synthase
IPTG	isopropyl-beta-D-thiogalactopyranoside
LB	Luria-Bertani
LTA4H	leukotriene A4 hydrolase
MAPK	mitogen-activated protein kinase
MEK	mitogen-activated ERK kinase
MIX	methylisobutylxanthine
NCoR	nuclear receptor corepressor
NHE3	sodium-hydrogen exchanger 3
NP40	Nonidet P-40
PBS	phosphate-buffered saline
pcd	Purkinje cell degeneration
PCR	polymerase chain reaction
PEPCK	phoshoenolpyruvate carboxykinase
PGC-1α	PPARgamma coactivator 1α
PKA	protein kinase A
PPAR	peroxisome proliferator-activated receptor
PPRE	PPAR response element
PRMT	protein arginine methyltransferase
PTEN	phosphatase and tensin homolog
RIPA	radioimmune precipitation assay
RXR	retinoid X receptor
SDS-PAGE	sodium dodecylsulfate-polyacrylamide gel electrophoresis
SMRT	silencing mediator of retinoid and thyroid hormone receptors
TAFI	thrombin-activatable fibrinolysis inhibitor
TBP	TATA-box binding protein
TEMED	N,N,N',N'-tetramethylethylenediamine
TGFβ	transforming growth factor β
TLC	thin layer chromatography
TZD	thiazolidinedione
VSMC	vascular smooth muscle cell

ACKNOWLEDGMENTS

I would like to acknowledge the help of many people who have made my work here over the past five years that much more productive and enjoyable. I would first like to thank my supervisor, Hyo-Sung, for his help and guidance and for giving me the flexibility to pursue my particular interests in this project. I would like to thank the many members of the Ro lab who have come and gone over the past five years. I would like to especially thank Shannon Reidy, whose scientific expertise, pleasant attitude, and love of eating out made life much more enjoyable in the lab. I also thank Lei, whose wealth of scientific knowledge has kept us in good stead, and Tara, Amin, and Chris, for their helpful and stimulating discussions, within and without science.

I am indebted to Dr. Neil Mattatall at the NRC-IMB, who was extremely helpful in guiding me through the construction of plasmids, purification of proteins, and homology modeling work. Without Neil's help this thesis would be very different. I would also like to thank Dr. Melanie Dobson for her critical reading of this thesis and advice throughout my graduate program, along with the other members of my supervisory committee, Drs. Rick Singer, Roger MacLeod, and Doug Hogue, who sadly passed away due to cancer early in my stay here. Other scientific help needing mention is the collaboration of Dr. Bassam Wakim from the Medical College of Wisconsin, whose initial work on p85/AEBP1 added a new and exciting dimension to this thesis and is greatly appreciated. Also, a graduate studentship from the Nova Scotia Health Research Foundation assisted me in pursuing my graduate research.

Finally, I would like to thank my family; my wife, Shari, for her friendship and patience with me in my late night stays at the lab, Tristan, for giving me a reason to play ball or a computer game and occasionally think of birds and tigers instead of proteins and DNA, and Brenna, whose 6-month-old grins and giggles always bring a smile to me as well. Also, my parents have always supported me in everything I undertake. My parents and other family members and friends outside the scientific world enable me to maintain a semblance of balance in life, without which the tedium of mini-preps and protein purifications would surely overwhelm.

Chapter 1

INTRODUCTION

The cell is an exquisitely controlled machine. Numerous levels of control regulate every part of its machinery, from gene transcription to protein degradation at the molecular level leading to cell growth, differentiation and death at the cellular level. Recent research actually suggests that it is not the cellular machinery itself that is most important in making each organism unique, but the control of its activity (*1*). This can happen through regulation of the transcription of a gene, through the regulation of the activity of the protein produced by the gene, or at any point in between. Many post-translational modifications, such as phosphorylation, acetylation, and ubiquitination, modify the function of a protein so that it serves its purpose only when needed. It is the combination of beautifully designed machinery and delicate control of this machinery that makes each cell function appropriately.

Adipocyte Enhancer-Binding Protein 1 (AEBP1) is a part of this cellular machinery that continues to surprise us with new functions and levels of control. AEBP1 is an 85 kDa protein with three domains, an N-terminal discoidin-like domain (DLD), a central carboxypeptidase (CP) domain, and a C-terminal highly charged and structurally uncharacterized domain (*2*) Figure 1A). It is an intracellular protein found in both the nuclear and cytoplasmic compartments (*3*). Also transcribed from the AEBP1 gene is Aortic Carboxypeptidase-like Protein (ACLP), which is identical to AEBP1, but with the addition of a 380-amino acid domain at its N-terminus (Figure 1B). This ACLP N-terminal domain contains a lysine- and proline-rich 11-amino acid repeating motif and a signal sequence (*4*), thus targeting ACLP to the extracellular matrix (*5*). Both AEBP1 and ACLP are known to be important for the proliferation of preadipocytes and inhibitory towards the differentiation of these cells into mature adipocytes (*2, 6*). Expression of the AEBP1/ACLP gene in mice is necessary for the normal growth of white adipose tissue, for the proper development of the mammary glands during pregnancy, and for male fertility (unpublished). AEBP1 is involved in the regulation of transcription (*2, 3*) and of cell signaling proteins (*7*), and is itself regulated by cell signaling proteins (*3, 8*).

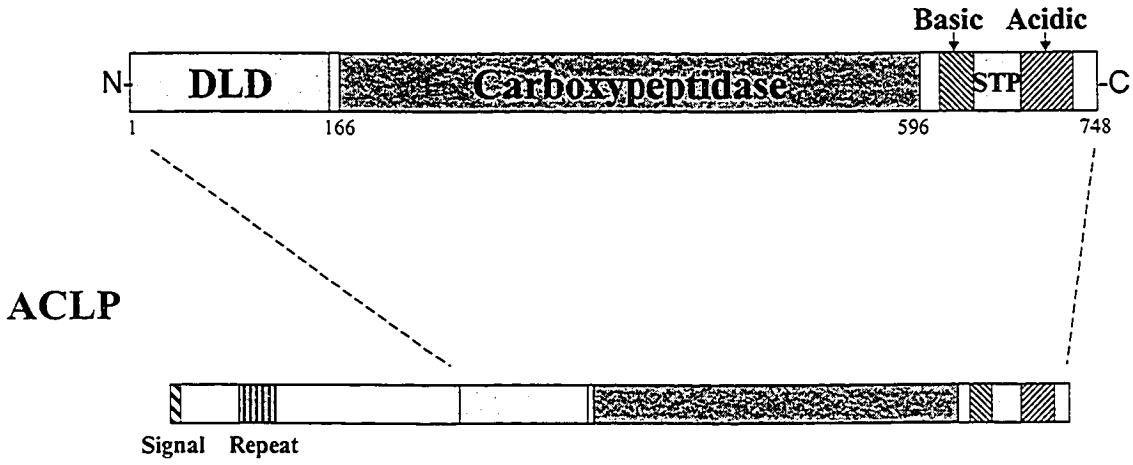


Figure 1. The domain structure of AEBP1 and ACLP. AEBP1 has three domains: the discoidin-like (DLD), the carboxypeptidase-like (CP), and the C-terminal domain. The C-terminal domain is composed of three distinct regions: the basic region, the serine-threonine-proline-rich (STP) region, and the acidic region. The numbers below the cartoon indicate the amino acid numbering at the domain boundaries. An additional 380 amino acids are present N-terminal to AEBP1 in the ACLP protein, made up of a signal sequence (signal) and a lysine- and proline-rich 11-amino acid repeating motif (repeat).

The effects of AEBP1 on fat cell growth leading to alterations in adipose tissue and mammary gland function suggest a role for AEBP1 in obesity and breast cancer. Many studies have shown that obesity is strongly correlated with heart disease, type 2 diabetes, and breast cancer (*9*). Because these diseases have a great impact on our society, many governments have placed renewed emphasis on obesity research. It is hoped that a clearer understanding of the genetic determinants leading to obesity, and hence heart disease, diabetes, and breast cancer, will arise from our study of AEBP1, leading to better treatments and preventative measures.

1.1. Molecular mechanisms of adipogenesis

Adipogenesis is the process by which partially differentiated preadipocyte cells develop into mature fat-filled adipocytes. It is the proliferation of preadipocytes and subsequent hypertrophy of differentiated adipocytes due to fat accumulation that eventually leads to obesity. The study of adipogenesis has been made easier through the use of partially differentiated preadipocyte cell lines derived from Swiss 3T3 mouse embryos (*10*). The 3T3-L1 and 3T3-F442A preadipocyte cell lines have been especially useful in the detailed characterization of the phenotypic and transcriptional alterations that accompany adipogenesis. The treatment of these preadipocyte cell lines at confluency with insulin and FBS, in the case of 3T3-F442A cells, or isobutylmethylxanthine and dexamethasone, in the case of 3T3-L1 cells, results in the initiation of adipogenesis characterized by an initial round of postconfluent cell division followed by growth arrest and fat droplet formation. Many adipocyte-specific genes are induced throughout this process and the most important of these will be described in the following section.

1.1.1. Transcriptional regulation of adipogenesis

The process of adipogenesis is very complex and a huge number of genes have been found to be differentially expressed upon differentiation of preadipocytes (*11*). However, many of the major factors regulating adipogenesis have been now identified. Adipogenesis has been reviewed in several comprehensive reviews (*12-16*) and is often abbreviated by describing the actions of four major players. At the early stages of preadipocyte differentiation, CCAAT

enhancer binding proteins (C/EBP) β and δ, two members of the large basic leucine zipper family of transcription factors, are transiently increased. This leads to the upregulation of C/EBPα as well as peroxisome proliferator activated receptor (PPAR) γ2, a nuclear hormone receptor, both of which cooperatively upregulate the expression of the other (Figure 2). Although ectopic expression of either C/EBPα or PPARγ2 in non-adipogenic mouse fibroblasts is able to induce adipogenesis (*17*), additional experiments have shown that PPARγ2 is the master regulator of adipogenesis. Ectopic expression of PPARγ2 in fibroblasts derived from C/EBPα null mice and activation with an appropriate ligand showed that PPARγ2 was sufficient for induction of adipogenesis (*18*). Conversely, fibroblasts obtained from PPARγ null mice were unable to differentiate into adipocytes even upon ectopic expression of C/EBPα (*19*). It appears that PPARγ2 is the dominant factor in the adipogenic hierarchy.

PPARγ2 is a member of a family of PPARs which also includes α, β/δ, and γ2 isoforms. PPARs contain two domains. The C-terminal ligand-binding domain has a transactivation function upon ligand binding. While there is some dispute over the relevant endogenous ligand, PPARs are known to be activated by the synthetic thiazolidinedione (TZD) drugs which are used as insulin sensitizers in the treatment of type 2 diabetes. The N-terminal DNA binding domain binds to DNA sequences called PPAR response elements (PPREs) as heterodimers with the retinoid X receptors (RXRs). The PPARγ1 and γ2 isoforms, expressed from the same gene through alternative promoter usage, are identical except for an additional 30 residues in the N-terminal DNA-binding domain of PPARγ2. Although these two proteins are very similar, PPARγ2 is specifically expressed in adipocytes while PPARγ1 is expressed in a variety of tissues (reviewed in (*20*)). The selective expression of PPARγ2 in adipocytes has suggested a unique role for this isoform, and further studies have indeed confirmed that PPARγ1 is unable to functionally replace PPARγ2 in adipogenesis (*21*). Ligand-bound PPARγ2 is known to regulate a vast number of genes important for adipocyte function (*22-24*), including genes encoding phosphoenolpyruvate carboxykinase (PEPCK) (*25*), the rate-limiting enzyme for glyceroneogenesis in the adipocyte, glucose transporters (GLUT) 1 and 4 (*26, 27*), critical for moving glucose into the adipocyte and maintaining insulin sensitivity, resistin (*28*), a hormone secreted by adipose cells and suggested to be a link between obesity and insulin resistance, and the aP2 gene (*29*), encoding adipose fatty acid binding protein (ALBP) which is

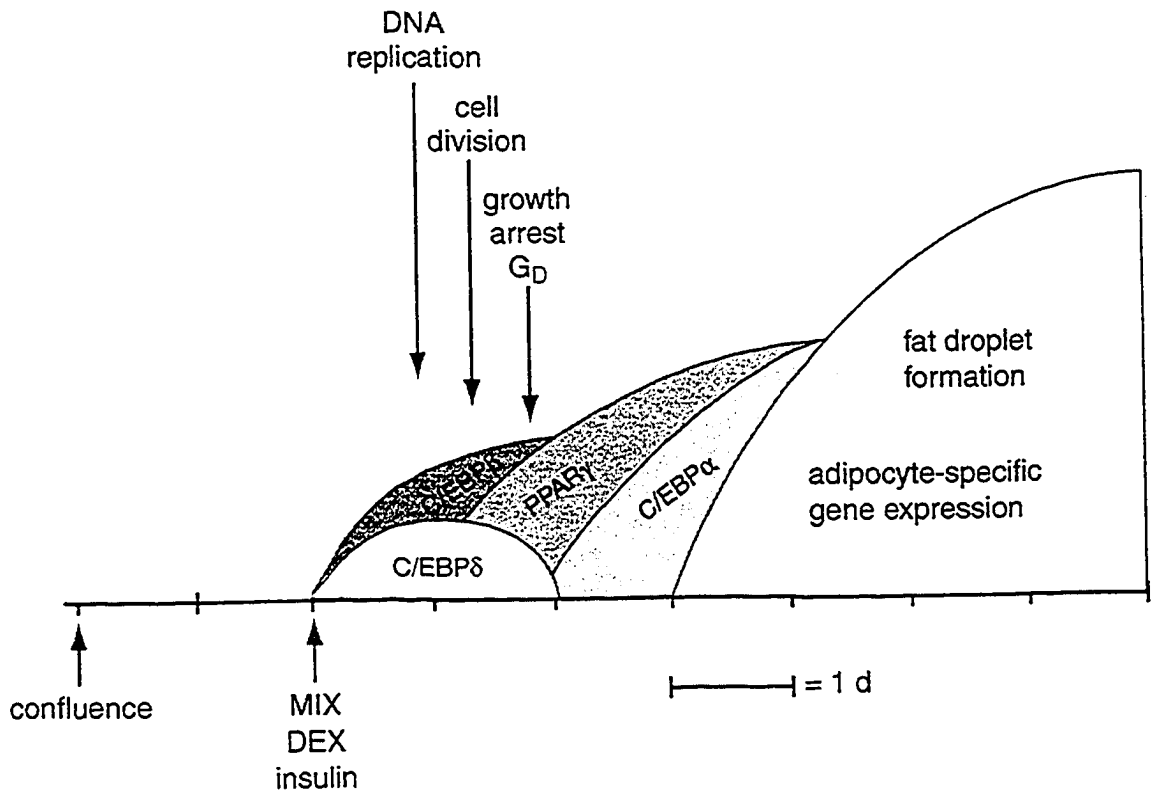


Figure 2. Progression of 3T3-L1 preadipocyte differentiation. A chronological record of the major events involved in 3T3-L1 adipogenesis is shown along with the important transcription factors upregulated at various time points. C/EBP, CCAAT/enhancer binding protein, PPAR, peroxisome proliferator-activated receptor; MIX, methylisobutylxanthine; DEX, dexamethasone. This figure is adapted from reference (10).

a critical protein for the transport of long-chain fatty acids within the adipocyte.

The aP2 gene is also regulated by other transcription factors through a region called the adipocyte enhancer-1 (AE-1) site within its promoter (*30, 31*). A screen for 3T3-L1 preadipocyte proteins interacting with the AE-1 site pulled out several clones, one of which was a partial cDNA of AEBP1 (*2*). AEBP1 was shown to specifically repress the transcription of the aP2 gene and bind *in vitro* to the AE-1 site (*2, 32*). The function of AEBP1 as a transcriptional repressor was further analyzed through a yeast two-hybrid screen for interacting partners (*3*). This screen discovered the γ5 subunit of a heterotrimeric G protein as an interacting partner for AEBP1. It was found that Gγ5 is localized to the nucleus in 3T3-L1 preadipocytes, but not NIH-3T3 cells, and was able to attenuate the transcriptional repression ability of AEBP1 in 3T3-L1 preadipocytes at the clonal expansion stage of adipogenesis. As 3T3-L1 differentiation proceeded, the levels of AEBP1 protein and mRNA were also shown to be downregulated (*2, 7*). Recently, others have also shown the transcription of the AEBP1 gene to be downregulated during adipogenesis. In a large-scale oligonucleotide microarray analysis, the relative abundance of 11,000 transcripts was compared at many different time points of 3T3-L1 preadipocyte differentiation as well as between adipocyte and stromal fractions isolated from mouse white adipose tissue (*11*). Expression of the AEBP1/ACLP gene steadily decreased throughout the initial stages of 3T3-L1 adipogenesis until approximately 4 days following induction of adipogenesis when its levels began to rise back to preinduction levels. A similar enrichment of AEBP1 expression in preadipocytes was found *in vivo*, when adipocytes isolated from white adipose tissue were compared with the stromal (preadipocyte) fraction. This suggests that AEBP1 plays a role in maintaining the preadipocyte phenotype.

1.1.2. Regulation of adipogenesis through MAP kinase

AEBP1 was also found to have non-transcriptional roles in adipogenesis through the regulation of the extracellular-regulated kinases 1 and 2 (ERK1/2) cell signaling molecules (*7*). These kinases are members of the mitogen-activated protein kinase (MAP kinase) family and are activated by dual phosphorylation following a cascade of upstream signaling events. The best characterized pathway leading to activation of ERK1/2 begins with activation of Ras, a

small GTPase, by a ligand-bound growth factor receptor. This leads to the activation of MEKK, a MAP kinase kinase kinase, which phosphorylates MEK, a MAP kinase kinase, which then phosphorylates ERK1/2, a MAP kinase. ERK1/2 is then translocated to the nucleus where it phosphorylates and modulates the activity of a variety of nuclear targets. The ERK1/2 MAP kinase pathway has been shown to be important in the process of adipogenesis. It was initially shown that expression of an oncogenic, constitutively active Ras oncogene enabled 3T3-L1 preadipocytes to differentiate into adipocytes without externally added insulin or insulin-like growth factor I (IGF-I), suggesting a role for downstream kinases such as ERK1/2 in adipogenesis (*33*). Experiments using an oligonucleotide-based antisense strategy to reduce expression of ERK1 and 2 clearly demonstrated that these proteins are necessary for 3T3-L1 cell adipogenesis (*34*). Further experiments, however, served to complicate the picture, as it was found that MAP kinase activation inhibits 3T3-L1 adipogenesis (*35*), possibly through PPARγ2, the so-called master regulator of adipogenesis. PPARγ2 activity is inhibited by phosphorylation by ERK (*36, 37*). While this contradicted the previous results, it made sense that the activation of ERK may be time-dependent – necessary during the initial clonal expansion phase of adipocyte differentiation to fulfill its role in proliferation, but then shut off before it is able to phosphorylate and inhibit the activity of PPARγ later in adipogenesis. This concept was substantiated by a recent study showing that inhibition of the MAP kinase cascade using a specific MEK inhibitor, U0126, inhibited the mitotic clonal expansion at the onset of adipogenesis and thus also inhibited adipogenesis as a whole (*38*).

AEBP1 has been found to interact with ERK through its N-terminus which contains a discoidin-like domain. The result of this interaction was the protection of ERK from dephosphorylation, leading to greater phosphorylation and activation of ERK, and an inhibition of adipogenesis (*7*). Based on the above-described effects of ERK on adipogenesis, it might be assumed that overexpression of AEBP1 in these cells results in a long-term activation of ERK, possibly leading to increased phosphorylation and inhibition of the activity of PPARγ. The ability of ERK1/2 to phosphorylate AEBP1 was also observed (*7*). Because phosphorylation by ERK is able to modulate the activity of PPARγ as well as a great number of other substrates identified to date (*39*), the activity of AEBP1 may also be regulated through this phosphorylation event.

1.1.3. Regulation of adipogenesis through the extracellular matrix

Many secreted extracellular proteins are important for regulating the process of adipogenesis. For example, deletion of osteonectin/SPARC, a matrix-associated protein, leads to an increase in adipose tissue due to lack of normal cell shape regulation (*40*). ACLP, an alternatively transcribed isoform of AEBP1 targeted to the extracellular matrix, also plays a role in adipogenesis. Similar to the function of AEBP1 in adipogenesis, Abderrahim-Ferkoune *et al.* (*41*) have shown that ACLP is downregulated upon differentiation of preadipocyte cell lines, and that overexpression of ACLP in these cells inhibited their ability to differentiate into adipocytes. This group also showed that ACLP is able to stimulate the transdifferentiation of these cells into smooth muscle-like cells. Gagnon *et al.* (*6*) found that ACLP was downregulated beginning with the postconfluent clonal expansion phase of 3T3-L1 adipogenesis, but that stable overexpression of ACLP did not affect the ability of these cells to differentiate into adipocytes. This same group recently recapitulated these experiments, as well as demonstrating a regulation of ACLP expression in 3T3-L1 preadipocytes by transforming growth factor β (TGFβ) (*42*). The exact reason for the conflicting results with respect to the effect of ACLP on adipogenesis is not apparent at this time, but may be due to differences in overexpression levels of ACLP or differences in cell lines used (3T3-L1 or 3T3-F442A).

It is apparent that the process of adipocyte differentiation is complex. Although it has been simplified through the use of partially-committed preadipocyte cell lines, small differences between these cell lines often make the picture appear even more complex. Additionally, these preadipocyte cell lines do not allow us to study the stages of commitment that lie prior to the preadipocyte stage. However, studies have elucidated many of the key players in the stages of preadipocyte differentiation, including C/EBPα, β, and δ, and PPARγ2. The role of AEBP1 in this process is also beginning to emerge as being inhibitory to differentiation through its effects on ERK1/2 and its transcriptional repression of an adipose specific gene.

1.2. Insights into the role of AEBP1 from knockout mouse models.

AEBP1/ACLP knockout mouse models have been made by two independent groups. Although one group focuses on the role of ACLP and one group on AEBP1, both mouse models eliminate the expression of both of these proteins. While both knockout mouse models were viable, the ACLP knockout mice were found to exhibit impaired abdominal wall development resulting in the majority of pups not surviving to birth (*5*). A similar low survival rate for AEBP1 knockout pups was also observed (*32*, data not published). Analysis of skin wounds and dermal fibroblasts from the ACLP knockouts indicated deficient wound healing due to a reduced proliferative ability. This, and unpublished work from our lab, suggests that AEBP1/ACLP is important for cell proliferation. Previously it had been shown that ACLP was highly expressed in the smooth muscle cells of the adult mouse aorta, and that its expression was upregulated in cultured aortic cells upon serum starvation (*4*). A subsequent study which confirmed a role for ACLP in cell proliferation found that the expression of ACLP was upregulated in the dedifferentiated and proliferating vascular smooth muscle cells (VSMC) of the neointima upon carotid injury. This expression in VSMCs was further substantiated using a transgenic mouse expressing an ACLP promoter-LacZ reporter gene (*43*).

While studies of ACLP and ACLP knockout mice seem to indicate a predominant expression of ACLP in vascular and abdominal smooth muscle cells, studies of the AEBP1 knockout mice have focused on the adipose and mammary gland tissue. The predominant phenotypes of AEBP1 knockout mice are the inability of females to lactate, the infertility of males, and an overall smaller size, including less adipose tissue (unpublished). The mammary gland dysfunction has been studied closely. Normal mouse mammary glands go through a complex process of development which can be separated into four distinct stages. During adolescence, weeks 4-6 in a mouse, ductal morphogenesis occurs as epithelial cells proliferate and fill the mammary fat pad. Upon pregnancy the mammary gland begins another phase of growth in which new ducts develop and lobuloalveoloar structures appear. Following parturition, hormonal stimuli initiate lactation. Finally, upon weaning of the pups the mammary glands undergo involution, in which the alveolar structures are destroyed through apoptosis and the mammary glands return to their pre-pregnancy state. AEBP1 knockout

mouse mammary glands develop normally until late pregnancy, when apoptotic marker proteins characteristic of involution begin to appear. Following the birth of the pups, the mammary glands can be seen to be prematurely involuting instead of lactating. Adipose-specific expression of an AEBP1 transgene driven by the aP2 promoter results in normal mammary gland development and the ability of these mice to nurse their pups, suggesting that the defect in mammary gland development is due to the lack of expression of AEBP1 in the mammary gland stroma, composed largely of a mixture of preadipocytes and adipocytes. This, of course, implicates AEBP1 in the signaling between stromal and epithelial cells within the mammary gland.

As well as having a role in the adipose tissue of the mammary gland, AEBP1 also plays a role in development of the white adipose tissue stores elsewhere in the mouse. AEBP1 knockout mice have reduced adipose tissue, suggested to be a result of decreased preadipocyte proliferation leading to fewer adipocyte precursors. Unpublished results suggest that decreased levels of AEBP1 result in increased levels of PTEN, a tumor suppressor molecule with lipid and protein phosphatase activity. This decrease in PTEN leads to increased preadipocyte proliferation.

The role of AEBP1 in the testes has been characterized on only a phenotypic level. Microscopy studies have indicated a dilated rete testes, the network of tubules that connects the seminiferous tubules in the testes with the efferent ductules in the epididymis, as early as six weeks of age with very low amounts of abnormal sperm present. This phenotype is very similar to that found in the estrogen receptor alpha (ERα) knockout mouse, which has been attributed to a reduction of sodium-hydrogen exchanger 3 (NHE3) expression which is important for fluid reabsorption within the efferent ductules (*44*). This reduction in fluid reabsorption results in fluid build-up and subsequent dilation of the rete testes.

1.3. Mechanisms of Transcriptional Regulation

The function of AEBP1 in transcriptional regulation requires a brief review of transcriptional mechanisms. Transcription is initiated by the binding of a large complex of general transcription factors and RNA polymerase at a promoter. Recent advances in structure determination have given us a clearer picture of the functions and positioning of these factors

leading to transcription initiation (*45*). In the case of TATA-box-containing promoters, the factor that initiates this sequence of events is the TATA-box binding protein (TBP), which interacts with the minor groove of the DNA at the TATA-box as part of a large multimeric general transcription factor called TFIID. Following the binding of this factor many other general transcription factors (TFIIA, TFIIB, TFIIE, TFIIF, TFIIH, and the Mediator complex) interact with each other and with RNA polymerase II to form a preinitiation complex. Each factor is necessary for a specific function, such as positioning of the complex on the DNA or unwinding of the DNA helix.

Each individual gene has specific sequences within its promoter and upstream enhancers which regulate its transcription. This gene regulation occurs through the binding of proteins called transcription factors, such as AEBP1, to specific 6-10 base-pair-long DNA sequences. A transcription factor binds to the DNA through a sequence-specific DNA binding domain. There are several different kinds of DNA binding domains which characterize different classes of factors. Most of these DNA-binding domains interact with DNA through an α-helix within a structure such as the helix-turn-helix motif, the homeodomain, the zinc finger motifs, the basic leucine zipper motif, or the helix-loop-helix domain. In many cases other parts of the transcription factor protein are also involved in the DNA interaction and sometimes the protein functions as a dimer in order to maximize specificity and binding strength. Each of these interactions can be modified through post-translational modifications such as phosphorylation. For example, PPARα, a close relative of PPARγ previously discussed in regards to its role in adipogenesis, is phosphorylated by protein kinase A (PKA) resulting in a stabilization of its interaction with DNA (*46*).

Following specific DNA binding, different mechanisms are employed by each transcription factor in order to regulate the transcription of a particular gene. A general theme that runs through all transcription factor mechanisms is the modification of the chromatin packaging of DNA in order to recruit or prevent recruitment of additional factors and/or basal transcription factors to the site of transcription initiation. Modifications that occur within the chromatin include repositioning whole nucleosomes by ATP-dependent remodeling complexes, replacing individual core histones with variant histones having different functions, and covalent modifications of histones by the addition of acetyl, methyl, phosphate, or ubiquitin groups. Transcription is activated or repressed depending on the modification

created and the context in which it lies. This context of additional modifications adds to the complexity and flexibility of transcriptional regulation and has been termed the "histone code".

1.3.1. Nucleosome Remodeling

Nucleosomes are inherently repressive. ATP dependent remodeling can be considered an enzymatic reaction in which the inherent thermal mobility of nucleosomes is enhanced by the actions of a remodeling complex, enabling DNA which was previously repressed within a nucleosomal structure to be freely accessed by the transcriptional machinery. These remodeling complexes all contain Snf2p-related ATP-hydrolyzing domains, similar to those found in helicases (reviewed in (*47*)), that function as ATP-dependent DNA translocases (*48*). One model proposed for the mechanism of nucleosome remodeling is that this helicase-containing complex remains attached to the histones while it moves the DNA as a wave around the histone octamer (*48*). This wave propagates due to induced changes in DNA twist providing the mechanical force needed to break DNA-histone interactions. Another model proposed is that of bulge diffusion, in which the DNA is removed from the histone octamer at the entry/exit point of the nucleosome and another segment some distance away subsequently binds, resulting in a bulge of DNA that is able to migrate around the nucleosome. Both of these models may play a role in nucleosome sliding and remodeling.

1.3.2. Variant Histones

In the process of remodeling and moving nucleosomes, histones may be removed or exchanged for variant forms resulting in changes in gene expression. A yeast multi-component complex containing a Swi2/Snf2-related ATPase named Swr1 was found to be responsible for exchange of the major histone H2A for a variant named H2A.Z in order to regulate a subset of genes (*49*). In most cases of histone exchange the mechanism is not well understood. However, many other variant histones have been found to be involved in transcriptional regulation.

Histone H1 is not a core histone, but interacts with nucleosomes at the position where DNA enters and exits the nucleosome, and is involved in higher-level chromatin organization.

It has several variant forms. One of these variants, histone H5, has been found to be depleted from active genes *in vivo*, while its deposition coincides with global transcriptional repression (*50*).

There are many variant forms of histone H2A. A variant of histone H2A with a very long C-terminal tail, MacroH2A, localizes specifically to the inactive X-chromosome and is thought to be involved in transcriptional repression (*51*). Two models have been suggested to explain this repressive effect, both involving the large C-terminal tail. One model suggests that the C-terminal tail may repress transcription enzymatically through a possible phosphoesterase activity of the macro domain towards ADP-ribosylated substrates (*52, 53*). Other groups suggest that it is the steric blockage of other transcriptional activators by the C-terminal tail that is important for its function (*54*). In contrast to MacroH2A, H2A-Bbd has a very small C-terminal tail and is localized to the active X-chromosome (*55*). Its presence is thought to destabilize the nucleosome and aid transcription (*56*). Much study has focused on the role of histone H2A.Z in transcription, but has resulted in seemingly conflicting results. While some studies suggest a role for H2A.Z in heterochromatin organization and gene silencing, other studies indicate involvement in gene induction, suggesting that there may be gene- or organism-specific differences in the way H2A.Z functions (*57*).

Histone H3.3 is a variant having only minor differences from the major histone H3 but generally is associated with transcriptional activity. In yeast histone H3.3 is the major H3 form. In *Drosophila*, McKittrick *et al.* (*58*) have shown that histone H3.3 makes up approximately 25% of total histone H3, and that this H3.3 is enriched in modifications that are generally associated with actively transcribed genes. Recently it was shown through chromatin immunoprecipitation experiments that H3.3 is preferentially found at active promoters (*59*).

1.3.3. Histone Post-Translational Modifications

Each nucleosome is composed of two H2A/H2B histone dimers and an $H3_2/H4_2$ histone tetramer. Positively-charged N-terminal tails of these histones extend from the core nucleosome and make contacts with the negatively-charged DNA phosphate backbone as well as neighboring nucleosomes (*60*). These histone tails can undergo post-translational

modifications that alter their interactions with DNA and/or their interactions with transcription factors and other components of the nuclear machinery (Figure 3 summarizes histone modifications known to date). Modifications of the histone tail can also affect other tail modifications, possibly through regulating the interactions of other modifying enzymes (*61*).

1.3.3.1. Acetylation

One of the best understood of the histone modifications is acetylation of lysines, which effectively neutralizes the positively charged lysines within the N-terminal tails and creates new binding surfaces for histone binding domains such as the bromodomain (*62*). Histone acetylation, performed by histone acetyltransferases (HATs), generally leads to gene activation, while deacetylation, performed by the histone deacetylases (HDACs), leads to repression. The first HAT to be discovered was the *Tetrahymena* homologue of the yeast Gcn5p (*63, 64*), while mammalian HDAC1 and yeast Rpd3p were the first deacetylases to be discovered (*65*). Many additional HATs, such as p300/CBP and TAFII250, and HDACs (HDAC1-11) have been subsequently discovered (see (*66*) for a list). These enzymes add and remove acetyl groups on specific lysine residues. For example, Gcn5p, as part of the SAGA complex, is specific to histone H3 and H2B and is able to acetylate five lysine residues on histone H3 and two on histone H2B (see Figure 3). On the other hand, Esa1p, as part of the NuA4 complex, is specific for acetylation of histones H2A and H4 (*67*). Many DNA-binding transcription factors recruit these enzymes to specific promoters as co-activators and co-repressors. For example, two prominent players in adipogenesis mentioned earlier are PPARγ2 and C/EBPα. PPARγ2 is a nuclear hormone receptor which activates transcription upon binding of a ligand such as one of the thiazolidinediones. In addition to ligand binding, coactivator interaction is also necessary. Two coactivators necessary for its function are PPAR gamma coactivator 1α (PGC-1α) and p300. PGC-1α was originally found to be a PPARγ specific coactivator within brown adipose tissue (*68*) but has subsequently been found to interact with many additional nuclear hormone receptors, as does p300. Recent work has shown that PPARγ and PGC-1α, interacting in a ligand-independent manner, form a complex with the TRAP220 subunit of

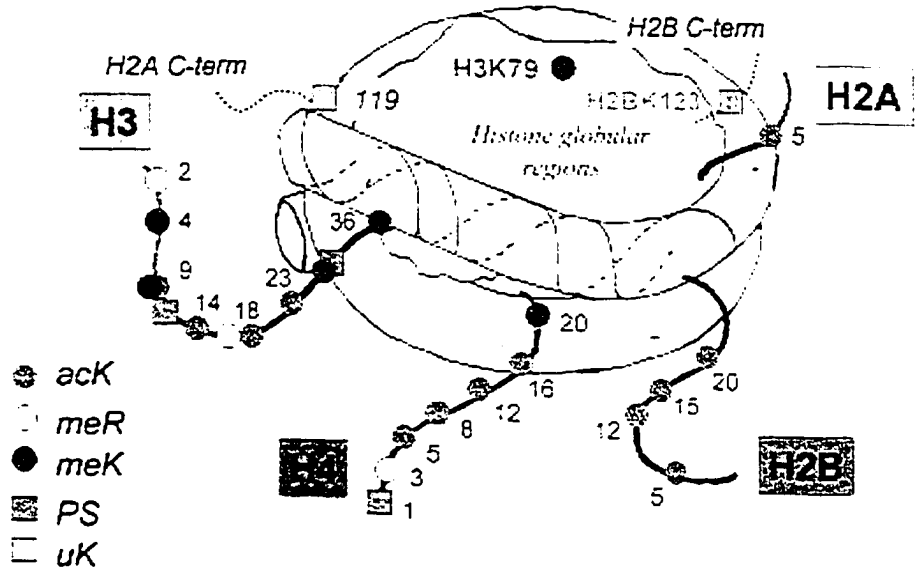


Figure 3. Histone modifications on the nucleosome core particle. A nucleosome is shown with 4 of the 8 core histone N-terminal tail domains and 2 C-terminal tails. Sites of posttranslational modification are indicated by colored symbols that are defined in the key; acK, acetyl lysine; meR, methyl arginine; meK, methyl lysine; PS, phosphoryl serine; and uK, ubiquitinated lysine. Residue numbers are shown for each modification. This cartoon is a compendium of data from various organisms, some of which may lack particular modifications (e.g., there is no H3meK9 in S. cerevisiae). Adapted from reference (61).

Mediator, thus stimulating transcription (*69*). This stimulation of Mediator-dependent transcription is also dependent on interaction of PPARγ with the p300 HAT and subsequent histone acetylation in a PPARγ ligand-dependent manner. Interestingly, PPARγ is also able to repress transcription through interaction with corepressors. In the absence of ligand, PPARγ interacts with the nuclear receptor corepressor (NCoR) and the silencing mediator of retinoid and thyroid hormone receptors (SMRT) in order to repress PPARγ-mediated transcriptional activation (*70*). These co-repressors mediate their repression through the activity of interacting HDAC enzymes (*71*). In a similar manner to PPARγ, C/EBPα also functions through interaction with p300 as a coactivator (*72*).

1.3.3.2. Methylation

Methylation is a histone post-translational modification which is able to lead to either transcriptional repression or activation. This modification occurs on the basic side-chains of lysines and arginines within histone H3 and H4 in at least 9 positions (*66*) also see Figure 3), and is performed by the histone methyltransferases (HMTs). While methylation of histones was known for many years, only recently were the first HMTs identified. The first HMT to be discovered was a nuclear receptor coactivator-associated protein, CARM1/PRMT4, a histone H3 arginine-specific methyltransferase involved in transcriptional activation through interaction with the p160 family of steroid receptor coactivators (*73*). Soon thereafter a human homolog of the *Drosophila* heterochromatic protein Su(var)3-9 was identified (SUV39H) as a histone H3 lysine-specific methyltransferase important for the repressive function of heterochromatin (*74*). Many more methyltransferases have been identified since within the PRMT family of methyltransferases as well as within the SET domain-containing family, of which the SUV39H methyltransferase is a part (*75*). One of these, Set9, catalyzes the methylation of histone H3 at lysine 4 and leads to transcriptional activation by inhibiting interaction of the NuRD chromatin remodeling and deacetylase complex with histone H3 (*76*). Set9 methylation of H3-K4 was additionally shown to inhibit the methylation of H3-K9 by SUV39H, just one example of combinatorial control of histone modifications (*76*). Until recently, a demethylation enzyme was lacking and many proposed that histone methylation was an irreversible mark. However, two mechanisms for demethylation have been recently found.

A peptidyl arginine deiminase, PADI4/PAD4, has been found to convert both arginine and methylated arginine to citrulline, thus opposing the histone arginine methylation reaction (*77, 78*). Most recently, a *bona fide* demethylase, LSD1, was discovered as a nuclear homogue of the amine oxidases and able to function as a histone demethylase and a transcriptional corepressor (*79*).

1.3.3.3. Ubiquitination and SUMOylation

Ubiquitination and SUMOylation, two common polypeptide protein modifications, have also been found to have roles in the regulation of gene expression through modifications of histones as well as many transcription factors. Ubiquitination and sumoylation have been found to have opposite effects on transcription, with ubiquitin modification of histones and transcription factors often resulting in activation while SUMO modification generally results in repression of transcription (reviewed in (*80*)). Early studies showed that ubiquitination of histone H2A was associated with transcriptionally active genes in *Drosophila* (*81*), although a recent study associated ubiquitination of H2A with gene silencing (*82*). Recently, the covalent attachment of ubiquitin to lysine 123 of yeast histone H2B was found to be absolutely required for methylation of histone H3 at lysine 4 as well as lysine 79, both of which are important for transcriptional activation and in these studies shown to be important for telomeric silencing and cell size control (*83-85*). Sumoylation was recently shown to occur on histone H4, although the exact location of this modification is not known (*86*). Several lines of evidence within this study point towards a role for SUMO modification in transcriptional repression. Targeting a SUMO E2 conjugating enzyme to the DNA resulted in an increase in SUMO at the promoter and a repression of transcription. Additionally, fusion of SUMO to histone H4 increased the amount of interacting HDAC1 and HP1, two known co-repressors. Thus, these modifications give us another example of the combinatorial nature of histone modifications, in which one modification may have an effect on the subsequent modification of other histone residues.

1.3.3.4. Phosphorylation

Phosphorylation of histones on serine and threonine residues is another case in which one modification affects another. While all histones have been shown to be phosphorylated (*66*), most research has focused on phosphorylation of the N-terminal tail of histone H3, specifically at serine 10. This modification was originally found to be associated with chromosome condensation and segregation during mitosis and meiosis (*87, 88*). Since that time evidence has accumulated that H3-S10 phosphorylation also plays a role in transcriptional activation and functions in concert with other H3 modifications. Phosphorylation of the H3 tail by Ipl1-aurora kinase was inhibited by prior methylation of lysine 9, and, vice versa, phosphorylation of serine 10 prevented methylation of lysine 9 by SUV39H (*74*). This would suggest that H3-S10 phosphorylation might activate transcription by preventing the repressive function of H3-K9 methylation. However, work by Lo *et al.* (*89*) indicated that phosphorylation of H3-S10 in yeast by the Snf1 kinase is followed sequentially by acetylation of H3-K14 by the acetyltransferase Gcn5, thus again implicating acetylation in transcriptional activation. Through all of the above work we can see that transcriptional regulation through histone modifications is not as simple as equating acetylation with activation and methylation with repression, for example, but the transcriptional state of each gene is regulated by its location within the chromatin and the sum total of all histone modifications.

1.3.3.5. N-linked phosphorylation

While most phosphorylation studies focus on serine, threonine, and tyrosine phosphorylation (O-linked), histidine, lysine, and arginine phosphorylation (N-linked) also exists and may also be involved in transcriptional regulation. This N-linked phosphorylation is alkali-stable and acid-labile and therefore has been neglected due to the acidic nature of many biochemical techniques. However, there is evidence that N-linked phosphorylation may comprise as much as 50% of total cellular phosphorylation (*90*). Early studies have revealed that two nuclear kinases from Walker-256 carcinosarcoma cells were able to produce acid-labile phosphates on lysine residues in histone H1 and on histidine residues in H4 *in vitro* (*91, 92*). Acid-labile phosphorylation was also found *in vivo* from nuclei of rat regenerating liver and

correlated with [^{3}H]-Thymidine incorporation into DNA, suggesting a possible role in DNA replication and mitosis (*93*). Histone histidine kinase activities have also been found in extracts from the slime mold *Physarum polycephalum*, from bovine heart, and from yeast. However, each of these studies used histones as an *in vitro* substrate and therefore histones may or may not be the endogenous substrates (*94, 95*). A large number of proteins other than histones have also been identified as containing phosphohistidine, and are reviewed elsewhere (*94, 96, 97*).

Arginine kinase activities are not as widely documented as histidine kinase activities, although several have been identified. A 34 kDa protein associated with rat liver DNA was identified as an arginine kinase capable of autophophorylation as well as phosphorylation of an 11 kDa unidentified chromosomal protein (*98*). Another arginine kinase activity was purified from calf thymus chromatin and activated by interaction with Ca^{2+}/calmodulin (*99*). This kinase was part of a larger complex containing two major proteins, 65 and 75 kDa in size. Separation of this complex into its component parts appeared to eliminate activity. A similar arginine kinase activity could be purified from mouse leukemia and quiescent rat heart endothelial cells, in addition to calf thymus, and analysis using cyanogen bromide cleavage and Edman degradation suggested that this arginine kinase phosphorylated histone H3 at arginines 2, 128, 129, and 131 *in vitro* (*100, 101*). *In vivo* experiments with rat heart endothelial cells showed that phosphoarginine could be found in histone H3 from quiescent cells, but not from actively dividing cells. The major component of this kinase complex was identified as a Ca^{2+}/calmodulin interacting protein of 85 kDa in size.

1.3.4. Possible Mechanism for Transcriptional Repression by AEBP1

AEBP1 was initially characterized as a transcriptional repressor of the aP2 gene through interaction with the AE-1 sequence within the promoter (*2*). Using a series of deletion mutants, this initial study correlated the transcriptional repression activity of AEBP1 with its carboxypeptidase activity, suggesting that C-terminal cleavage of proteins involved in transcription may lead to the repression of transcription. Subsequent work showed that AEBP1 had greater affinity for a C-terminal arginine than for lysine and that the CP activity of AEBP1 was activated by interaction with AE-1 DNA, suggesting that DNA binding might activate AEBP1 for reaction with a substrate also located at the DNA. As described above, the

histones are no longer considered passive bystanders in the nucleus, but active participants in nuclear processes such as transcription and DNA replication. Histones H1, H2A, and H2B all have a C-terminal lysine, while histone H3 has a penultimate arginine. Many of the variant H2A histones have altered C-terminal domains which play a role in chromatin compaction (*57*). The C-terminus of histone H1 has also been found to be involved in chromatin compaction (*102*). Few histone post-translational modifications occur at the C-terminus, except for the ubiquitination of histone H2A at K119 and of histone H2B at K123, the *in vitro* arginine phosphorylation of histone H3 at R128, R129, and R131, and the *in vitro* methylation of histone H3 at one or more of arginines 128, 129, 131, and 134 (*82, 83, 100, 103*). Little is known regarding the function of arginine phosphorylation of histone H3. However the ubiquitination of both H2A and H2B is known to be associated with transcriptional activity. It could be that the interaction of AEBP1 with and/or cleavage of a histone may lead to the transcriptional repression exhibited by AEBP1.

1.4. The Metallopeptidase Family

AEBP1 is a member of the zinc metalloproteases family. The number of known zinc metallopeptidases has increased dramatically in recent years, making a general understanding of these peptidases somewhat difficult and often confusing. They have been classified according to their function, their overall tertiary structure, and also according to the sequences surrounding the critical zinc-binding residues. Of course, these classifications are often identical, as structure usually underlies function. Hooper (*104*) summarized what is known regarding the zinc metallopeptidases and classified them into four categories. While the overall tertiary structures between these classes of enzymes may be quite different and the arrangement of metal binding residues within the primary sequence may vary, the basic mechanism of proteolysis is quite similar in all cases. I will summarize the structure and function of the two major classes of enzymes, the zincins, which can be subdivided into the gluzincins and the metzincins, and the carboxypeptidases, which can be subdivided into the digestive and the regulatory carboxypeptidases, with more focus placed on the regulatory carboxypeptidase class to which AEBP1 belongs. Figure 4 illustrates the structures of one member of each of these subfamilies of metalloproteases. The structural differences and slight

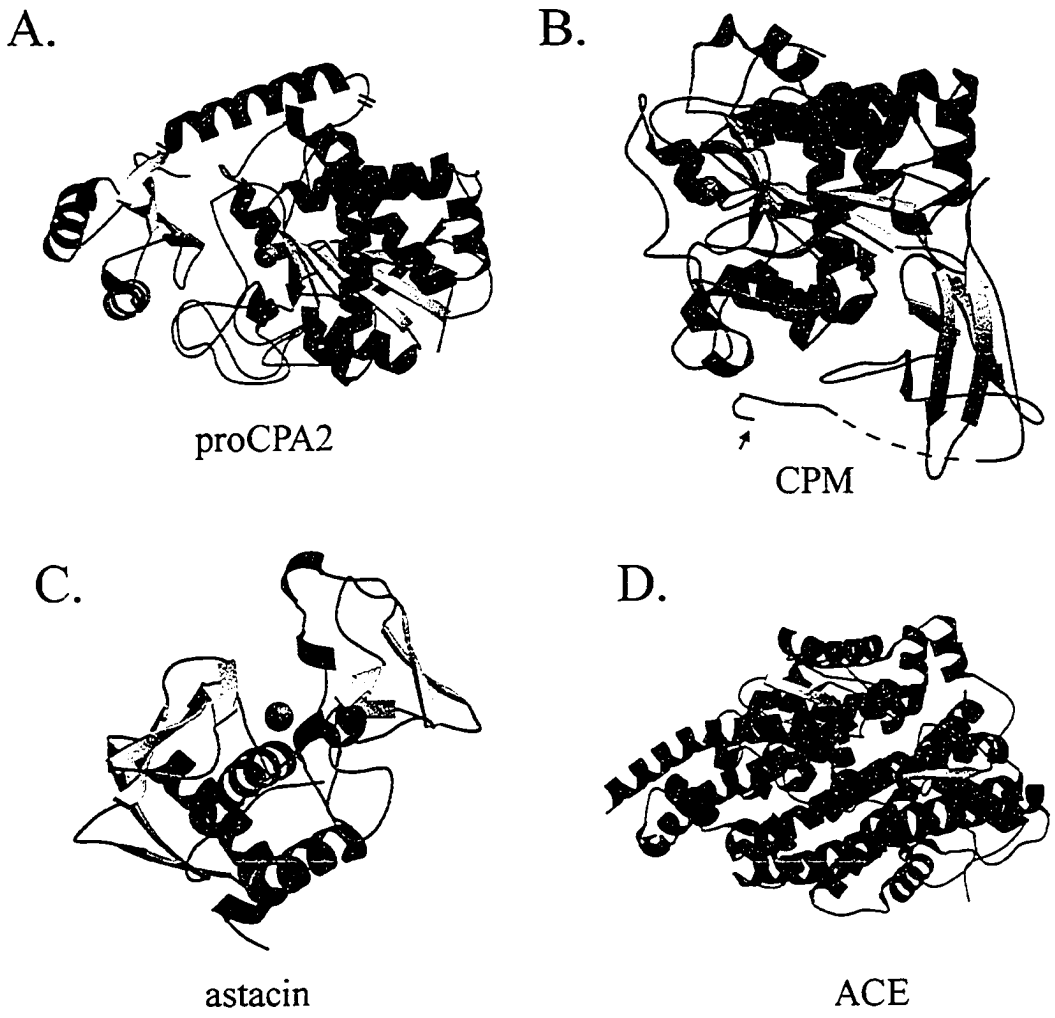


Figure 4. Representative members of four classes of metallopeptidases. (A) Human procarboxypeptidase A2 (proCPA2, PDB ID 1AYE) illustrates the structure of the digestive carboxypeptidases, with the site of propeptide cleavage shown by a double slash. (B) Human carboxypeptidase M (CPM, PDB ID 1UWY) is a membrane-bound member of the regulatory carboxypeptidase class. Four residues C-terminal from the C-terminal residue shown here (indicated by an arrow) is the GPI anchor site, which was removed for crystalization. (C) Astacin (PDB ID 1AST) from the crayfish Astacus astacus L. is an endopeptidase and the prototype of the metzincin class of proteases. (D) Human angiotensin converting enzyme (ACE, PDB ID 1O8A) is a dipeptidase member of the gluzincin class of enzymes. All structures contain a catalytic zinc ion shown as a gray sphere. This figure was made using Molscript and Raster3D.

variations in mechanisms demonstrate the variety of amino acid residues and motifs which are able to perform the same function.

1.4.1. Zincins

The zincin family consists of those peptidases having the HEXXH zinc-binding motif. This is a large group of peptidases having many diverse functions. It can be further subdivided into two subcategories, termed by some the gluzincins and the metzincins. The gluzincins include members of the thermolysin, endopeptidase-24.11, angiotensin converting enzyme (ACE), and aminopeptidase families. These proteases have the HEXXH motif, in which the two histidines are zinc ligands, with a glutamate at varying distances further C-terminal as a third zinc ligand. As a variation on this zinc-binding theme, there exists another class of proteases of the insulinase family, in which the zinc binding motif, HXXEH has had the glutamate and histidine swapped.

Angiotensin converting enzyme is a very prominent member of this family. The crystal structures of *Drosophila* (*105*) and human somatic (*106*) and testis (*107*) ACE have been solved (see Figure 4D). Recently, a novel relative of ACE has been discovered and termed ACE2. ACE2 has a very similar secondary and tertiary structure to ACE, as determined by X-ray crystallography (*108*), but has a different substrate specificity due to changes in several critical substrate binding residues. While ACE cleaves C-terminal dipeptides such as His-Leu, ACE2 cleaves single amino acids from the C-termini of substrates, thus making it structurally similar to ACE but functionally similar to the carboxypeptidases which will be discussed later. The attention that ACE has received is due to the role it plays in regulating blood pressure and heart function. ACE catalyzes the production of the hypertensive peptide, angiotensin II, and destroys the hypotensive peptide, bradykinin. Many inhibitors of ACE have been designed, such as captopril and lisinopril, and are widely used for the treatment of high blood pressure and other heart disorders.

The metzincin subcategory of peptidases includes the members of the astacin, serratia, reprolysin, and matrixin families (see Figure 4C). These proteases have a longer zinc binding motif, HEBXHXBGBXH, in which all three histidines are zinc binding residues and where "B" indicates a bulky, apolar residue. Metzincins also have a methionine-containing turn (the

Met-turn) which distinquishes them from the gluzincins. While the four subcategories of the metzincins have several differences, the most notable difference is in zinc coordination. The astacins and serratias are pentahedrally coordinated, with the three above mentioned histidines, a water molecule, and a tyrosine two residues C-terminal to the methionine in the Met-turn. The reprolysins and matrixin families are tetrahedrally coordinated, as they lack the tyrosine. This is a very large family of metalloproteases with a wide range of functions which is beyond the scope of this introduction.

1.4.2. Carboxypeptidases

The zinc metallocarboxypeptidase family of enzymes is characterized in part by an HXXE zinc binding motif (*109*). This large family is subdivided into two general classes which were originally named the digestive and regulatory subfamilies based on the functions of the first identified enzymes. Now t hese subfamilies are more accurately c lassified based on structural homology with the first identified members of each subfamily – 1) the A/B subfamily of CPs, including CPA (*110*) and CPB, which are primarily involved in the digestion of food, as well as TAFI and mast cell CPA, and 2) the N/E subfamily, including, CPN, CPE (*111, 112*), CPD (*113*), and CPM (*114*), which are involved in more specific cleavages important for peptide and protein biosynthesis. Three members of this latter class of regulatory carboxypeptidases, AEBP1 (*2*), CPX1 (*115*) and CPX2 (*116*), are also similar in that they each have an N-terminal discoidin domain similar to that found in the Factor V/VIII coagulation factors. However, all three of these carboxypeptidases lack several residues necessary for activity against standard carboxypeptidase substrates and, although low activity has been detected for AEBP1, they have been proposed by some to function as binding proteins without enzymatic activity (*117, 118*).

1.4.2.1. The A/B Carboxypeptidase Subfamily

The prototype carboxypeptidase upon which most work has been based is carboxypeptidase A. The X-ray crystal structure of this enzyme was solved by Lipscomb in 1967 (*119*), the third protein structure ever to be solved (see Figure 4A). Most of what we

know regarding the enzymatic mechanism of CPA, and similar CPs, comes from this work and subsequent work with transition state analogues and slowly hydrolyzed substrates. While several hypotheses as to the mechanism of CPA have been put forth over the years, the most likely of these is the promoted-water pathway (Figure 5, Ref.(*120*). In this mechanism the catalytic zinc ion, which is coordinated by histidine 69, glutamate 72, histidine 196, and a water molecule, activates the coordinated water molecule with the help of the carboxylate of nearby glutamate 270, causing this water molecule to act more like a hydroxide ion. Nucleophilic attack of this activated water molecule on the carbonyl carbon of the scissile peptide bond is accompanied by donation of a proton from the water molecule to the general base, glutamate 270, forming a negatively charged tetrahedral intermediate. This intermediate is stabilized by the positively charged side chain of arginine 127 and also by the zinc ion. Following formation of this intermediate, a proton is donated from glutamate 270 to the peptide NH group, resulting in cleavage of the peptide bond.

Zinc metallocarboxypeptidases cleave only specific C-terminal residues. They recognize the terminal carboxylate of substrates through three important interactions. In CPA, the side chains of tyrosine 248, arginine 145, and asparagine 144 all make hydrogen bonds with the terminal carboxylate group of the substrate. Tyrosine 248 is also striking due to the large 12Å swing it makes upon substrate binding which aids in occluding water from the active site. These three residues of CPA, along with the above mentioned His69, Glu72, His196, Glu270, and Arg127, constitute critical residues considered necessary for enzymatic activity in all members of this family of carboxypeptidases. In addition, the specificity of CPA for aliphatic or aromatic C-terminal substrate residues and CPB for basic C-terminal substrate residues is imparted by the shape, size, and charge of the pocket encompassing the side-chain of the substrate C-terminal residue. This, of course, varies depending on the specific CP.

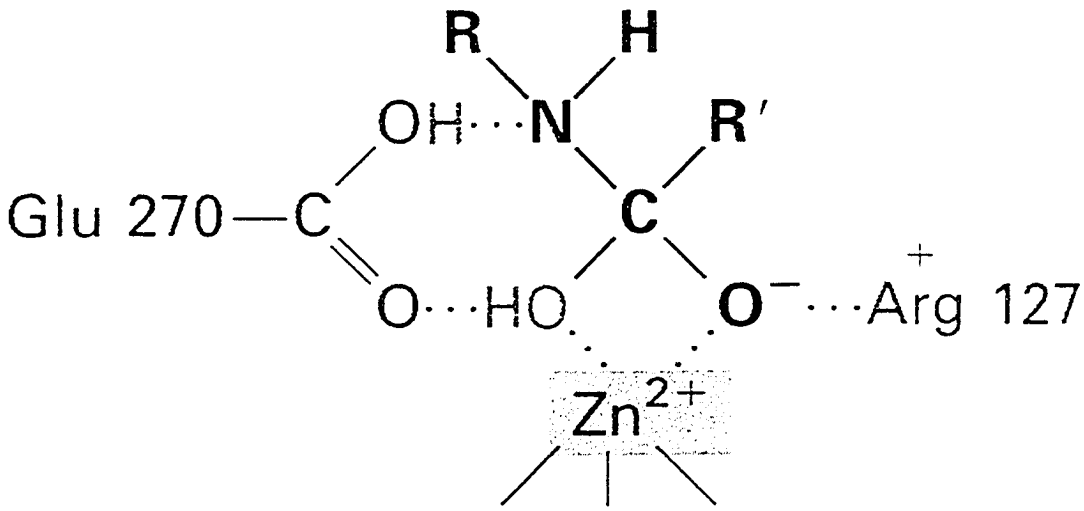


Figure 5. Proposed tetrahedral transition state in peptide-bond hydrolysis by CPA and related carboxypeptidases. In this "promoted-water" pathway, activation of water is accomplished through the combined efforts of the catalytic Zn^{2+} and glutamate 270. The nucleophilic oxygen of this activated water attacks the carybonyl carbon atom, glutamate 270 accepts a proton from the water, and a negatively charged tetrahedral intermediate is formed, stabilized by arginine 127. A proton from the COOH group of glutamate 270 is then transferred to the peptide NH group and the peptide bond is cleaved. This figure is taken from (120).

1.4.2.2. The N/E Carboxypeptidase Subfamily

The N/E carboxypeptidases have a similar structure and mechanism as the CPA/B enzymes. However, there are some noticeable differences in the overall structure. Generally, the N/E CPs have an additional C-terminal β-barrel-shaped subdomain. The function of this subdomain is presently unknown. Members of the CPA/B family usually have an N-terminal prodomain which is cleaved off in order to activate the enzyme, while the N/E CPs do not have a prodomain but instead seem to be regulated through localization and substrate specificity.

The first N/E CP to be discovered was CPN, also known as arginine carboxypeptidase or kininase I. CPN is found circulating in the plasma and is involved in the degradation of bradykinin and other peptide hormones circulating in the bloodstream. It is a tetramer, composed of two identical 83 kDa subunits and two identical 55 kDa subunits. The 55 kDa subunits contain the catalytic activity directed against peptides with lysine or arginine at their C-termini (*121*)

CPE, also known as carboxypeptidase H and enkephalin convertase, is a prohormone-processing enzyme present in the brain, pituitary gland, and many endocrine tissues. CPE removes C-terminal arginine and lysine from many peptides, such as insulins and enkephalins, within neuroendocrine secretory vesicles, leading to activation of these peptide hormones. A mutation in the CPE enzyme occurs naturally in the Cpefat/Cpefat mice, leading to inactivation of this enzyme and accumulation of many incompletely processed peptides (*122*). The major phenotype of these mice is severe obesity, along with altered levels and processing of many peptide hormones. A CPE knockout mouse model was recently generated and resulted in a similar phenotype of obesity, caused by an increase in food intake and a decrease in exercise (*123*). Altered levels of insulin and leptin and a wide variety of behavioural problems implicate CPE in more than just processing of cellular prohormones.

CPD is the only member of the carboxypeptidase family that contains multiple CP domains. CPD is a 180 kDa protein made up of three tandem ~390 amino acid CP domains (CPD-I, CPD-II, and CPD-III) followed by a transmembrane domain and a short cytosolic tail (*124*). While the third CP repeat lacks several critical catalytic residues and is inactive toward

standard CP substrates, the first two repeats have activity with different specificities. CPD-I is optimally active at pH 6.3-7.5 and prefers substrates with C-terminal arginine, whereas CPD-II is optimally active at pH 5.0-6.5 and prefers substrates with C-terminal lysine (*125*). This might serve to make CPD active toward a very broad range of substrates. CPD is found in the *trans* Golgi network in which the pH covers the optimums of CPD-I and –II, and it is likely to function in the production of receptors and growth factors processed there (*126*). CPD-II was the first of the regulatory CPs for which a structure was solved by X-ray crystallography, giving us much insight into the catalytic mechanism and the differences between CPA/B and this enzyme (*113, 127*). A novel prolactin-induced CPD (termed CPD-N) has recently been discovered in rat Nb2 T-lymphoma cells and found localized to the nucleus (*128*). The role of this CP within the nucleus is not known.

CPM is a membrane bound carboxypeptidase which is a marker of macrophage differentiation (*129*), although expressed in many different cell types. It cleaves C-terminal arginine and lysine from peptides such as bradykinin at neutral pH (*130*). CPM is attached to the cell membrane via a glycosylphosphatidylinositol (GPI) anchor, making it unique from other CPs such as CPE and CPD in that it may function in the control of peptide hormone activity at the cell surface. Its neutral pH optimum is also distinct from the acidic optimums of CPE and CPD. The crystal structure of CPM has also recently been solved, giving us greater understanding of the determinants of its specificity and of its orientation with respect to the cell membrane (*114*), also see Figure 4B).

CPZ was initially cloned and characterized as a novel metallocarboxypeptidase having relatively low activity against C-terminal basic amino acids (*118*). In addition to the CP domain, it contains a signal peptide and a 120-residue cysteine-rich domain with homology to the Frizzled proteins that are receptors for the Wnt family of signaling proteins, thus implicating CPZ in the Wnt pathway. CPZ is localized to the regulated secretory pathway and the extracellular matrix in many cell types (*131*). It is expressed more highly in embryonic tissues, and its expression pattern in embryonic development closely matches that of several *Wnt* genes (*126*). Recently, the role of CPZ in modulating the Wnt pathway has been confirmed. Moeller *et al.* (*132*) have shown that CPZ interacts with Wnt4 via its cysteine-rich domain *in vitro* and that CPZ, but not its inactive mutant form, enhances Wnt-dependent induction of *Cdx1*. Injection of a virus containing the chicken CPZ open reading frame into

the segmental plate of chick embryos in the presumptive wing region resulted in a loss of the scapular blade and of rostral ribs. Injection of an enzymatically inactive mutant of CPZ did not cause any skeletal abnormalities. The endogenous substrate for CPZ is still not known although the above work suggests that Wnt4 may be this substrate.

1.4.2.3. The Non-Traditional Carboxypeptidases

Within the family of zinc carboxypeptidases there exists a number of CPs which have the structure but not the function of the traditional CPs. CPX-1 and CPX-2 are two very similar proteins which lack any detectable CP activity towards traditional substrates. AEBP1 and its isoform, ACLP, both contain domain similarities to the CPX proteins and have low to no activity towards standard CP substrates. CPD-III, as mentioned above, is another CP domain lacking detectable CP activity but proposed to have a definite role due to amino acid conservation within this domain. Another protein called NnaI has low homology to carboxypeptidases and may also fall into this group.

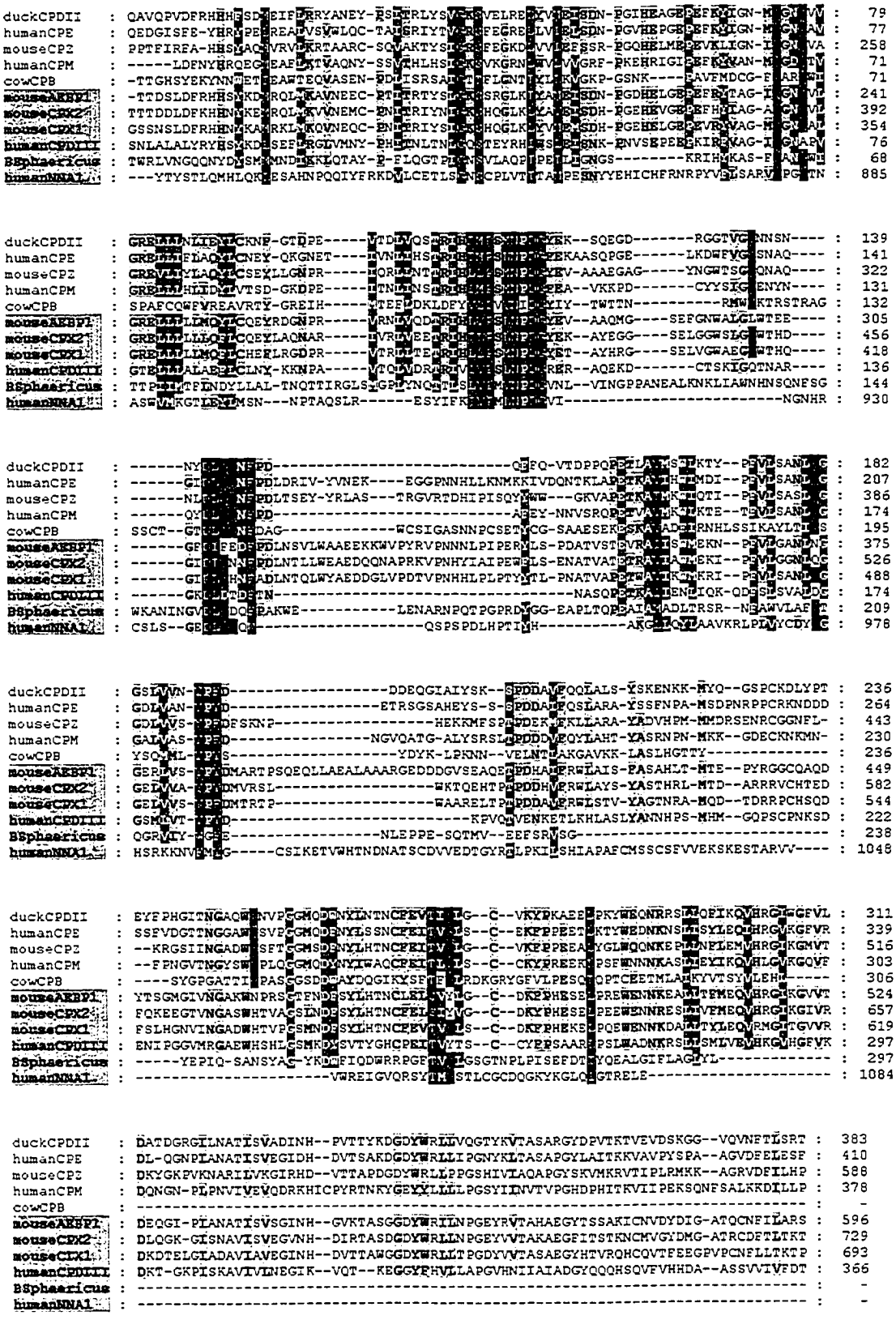
Little is known about CPX-1 and -2. They have been suggested to function as extracellular binding proteins, due to their lack of enzymatic activity and predicted signal sequences (*115, 116*). The N-terminus of these proteins consists of a 100-140 amino acid putatively unstructured region rich in proline and glycine in one part and basic amino acids in another. Next to this is a ~160 amino acid discoidin-like domain (DLD) having homology to Discoidin I, an extracellular lectin from the slime mold *Dictyostelium discoideum*, and the C2 domains of Factor V and Factor VIII coagulation proteins. The C-terminal half of both CPX-1 and -2 is made up of a carboxypeptidase domain followed by a short 30-35 residue highly basic stretch of amino acids. The CP domain of CPX-1 retains all three putative zinc-coordinating residues and the general base. However, histidines replace both CPA tyrosine 248 and arginine 145, both involved in substrate binding (Figure 6). CPX-2 is also lacking these substrate binding residues (Histidine replacing tyrosine 248 and asparagine replacing arginine 145, CPA numbering), as well as lacking one putative zinc-coordinating residue and the general base (glutamine replacing histidine 196 and tyrosine replacing glutamate 270 of CPA, see Figure 6). Although CPX-2 does not have activity against standard CP substrates, it is still able to bind

arginine-sepharose, suggesting that this and other similar CP-containing proteins may function through binding rather than enzymatic activity (*116*). A recent paper has implicated CPX-1 in osteoclastogenesis and immunofluorescence microscopy suggests that it may have an intracellular role (*133*). CPX-1 RNA levels are highest in testis and spleen adult tissues, with moderate levels detected in salivary gland, brain, heart, lung, and kidney. In embryonic tissues, CPX-1 was found by *in situ* hybridization to be more highly expressed in the nasal mesenchyme, meninges, and many skeletal structures (*115*). CPX-2 is expressed more highly in lung and kidney than brain and liver, as determined by Northern blot analysis (*116*).

CPD-III, the third CP repeat of carboxypeptidase D, has been shown to be inactive against standard substrates (*125, 134*), most likely due to the fact that most residues considered critical to catalytic function are altered in this domain (Figure 6). CPD-III retains a histidine equivalent to histidine 69 in CPA but no other residues predicted to be crucial for zinc coordination or substrate binding. However, the duck homologue of CPD-III has been shown to be necessary for interaction with the pre-S domain of the large envelope protein from hepatitis B virus (*134*). It is highly conserved between human and duck and so must have some necessary function, possibly that of protein-protein interaction.

One more carboxypeptidase with undetermined CP activity is Nna1. Nna1 is a protein induced upon axon regeneration following transection or crush injury (*135*). It contains a putative carboxypeptidase domain along with an ATP/GTP binding motif and nuclear localization signals and it is detected in both cytoplasmic and nuclear fractions of transfected neurons (*135*). There is very limited sequence identity between the carboxypeptidase domain of Nna1 and other carboxypeptidase domains. Depending on how the alignment is made, there may or may not be the critical residues necessary for CP activity. An earlier alignment of Nna1 to AEBP1 is likely inaccurate due to large insertions unique to AEBP1 (*135*). A more likely alignment, in which one of the putative substrate binding residues, arginine 135, is lacking, as well as the tyrosine located at position 248 in CPA, is illustrated in Figure 6. This suggests that Nna1 may fall into the class of "non-traditional" carboxypeptidases possibly involved in protein interactions rather than enzymatic activity. Subsequent to its cloning, a mutant Nna1 was found to be the cause of a classical recessive mouse mutant, *Purkinje cell degeneration (pcd)* (*136*). This mutant was originally reported in 1976 (*137*) and much has been

Figure 6. Multiple alignment of representative active and non-traditional carboxypeptidases. Shown in yellow are CPs with traditional CPB-like activity. Shown in pink are CPs with no activity or unknown/non-traditional CP activity. In the case where a protein has other domains or regions in addition to the CP domain, only the CP domain is shown. BSphaericus indicates Peptidase I from the indicated bacteria. Residues known to coordinate the catalytic zinc ion in most CPs and generally considered crucial to CP activity are highlighted in blue. Residues known to be important for substrate binding in most CPs are indicated in red. Proteins are taken from human or mouse, unless a crystal structure was solved of a protein from another organism. This figure was made using GeneDoc.



published since on the cellular and neurological phenotypes manifested in this mouse. Behaviourally, these mice exhibit moderate ataxia of gait detected around 22 days of age. While PCD mice at birth to about 15 days are similar in size to normal littermates, by 28 days they are reduced to only 63% the weight of WT mice. The females are fertile, although they have difficulty rearing the few litters they do have, and the males are sterile due to abnormally shaped and immotile sperm. Purkinje cells, located predominantly in the cerebellum, drastically decline in number from around day 20 until they are virtually entirely degenerated by day 30.

AEBP1/ACLP is possibly the most studied of the putatively non-tranditional carboxypeptidase proteins. The CP domain of AEBP1 has strong similarity to the regulatory CPN/E family of carboxpeptidases (the CP domain), although lacking several amino acids considered critical for activity against standard CP substrates. Of the previously mentioned residues that are critical for zinc-coordination and substrate binding in CPA, only histidine 69 and glutamate 72 are conserved in AEBP1 (Figure 6). The functions of AEBP1 in adipogenesis and various mouse models have been largely covered in the previous sections.

1.5. Goals of this Research

Most work on AEBP1 has assumed that one of its major roles is as a transcriptional repressor. However, little is known regarding the mechanism AEBP1 uses to repress transcription. Previous work has indicated that the carboxypeptidase domain of AEBP1 is important for this transcriptional repression (*2*). This is an exciting and novel role for a carboxypeptidase domain. However, much controversy exists regarding whether AEBP1 has CP activity, DNA binding activity, and/or transcriptional repression activity (*4, 41, 112, 118*). Several X-ray crystal structures have recently been solved for structures with high similarity to both the DLD and CP domains of AEBP1 (*113, 127, 138, 139*). My goal in this thesis has been to use homology modeling and rationally designed protein mutations and deletions to more fully understand the function and molecular mechanism of AEBP1 as a transcriptional repressor.

Chapter 2

MATERIALS AND METHODS

2.1. Computational Methods

2.1.1. Homology Modeling

The most critical step in protein modeling by homology is the alignment of the sequence of the protein to be modeled with that of a suitable template. In this work, suitable template candidates were known from the literature. However, PSI-BLAST (*140*) was used to search for any additional templates within the PDB library. This performs multiple iterative searches based on the consensus sequence of previously aligned matches, thus retrieving distant relatives as well as close matches.

Using the alignment produced by PSI-BLAST, the program Modeller (*141, 142*) was used to produce a molecular model based on satisfaction of spacial restraints. This program was run on a Linux desktop computer. Modeller was asked to produce 20 models, and the best of these models, based on the objective function produced by Modeller, were taken for further analysis of quality. Model quality was first assessed using ProsaII (*143*) to detect any energy clashes. ProsaII produces a plot indicating relative interaction energies within a defined window of amino acids (in this study, a window size of 10) at all points along the protein sequence. This plot is an indicator of proper protein folding, with a well-folded protein having the majority of the plot below zero. Procheck (*144*), which produces a Ramachandran plot, was also used to analyze the amino acid stereochemistry of the models produced. Finally, SCWRL3 (*145*) was used to predict side-chain placement for those amino acids not identical to the template.

When a suitable model was obtained, figures were produced using Molscript (*146*) and Raster3D (*147*). Other programs which aided in visualization and manipulation include VMD and Swiss PDB Viewer.

2.1.2. Sequence Analysis

The majority of AEBP1 homologue sequences were obtained from the Ensembl genome database, as well as some from the NCBI databases. Alignments were produced using ClustalW (*148*) and formatted for figures using GeneDoc (*149*). ClustalW also produced a .dnd file, which was used to produce a phylogenetic tree using TreeView (*150*). Sequences were analyzed for phosphorylation site motifs using the Scansite search algorithm (*151*), available on the World Wide Web. Secondary structure predictions were performed using PHDsec (*152*) or the Jpred server (*153*) which has recently incorporated a neural network prediction method called Jnet.

2.2. Recombinant DNA Methods

2.2.1. Plasmid DNA Construction

All plasmids were purified using Qiagen maxi-prep or mini-prep kits according to the manufacturer's protocol. Plasmids were quantitated by UV spectrophotometry (OD 260:280) when purified by maxi-prep or by comparison with High DNA Mass Ladder (Invitrogen) on an agarose gel when purified by a mini-prep kit.

Plasmids and their methods of construction are listed in Table 1. In those cases in which PCR was used to amplify the desired sequence for insertion into a plasmid, the reaction mix consisted of: 5 µl 10X cloned Pfu buffer, 10-20 ng template DNA, 125 ng primer 1, 125 ng primer 2, 1 µl 10 mM dNTP mix, 0.5 µl PfuTurbo DNA Polymerase (Stratagene), and water to a total volume of 50 µl. The PCR cycle was as follows: 30 s at 94°C, followed by 30 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 1 min/kb template. Following this cycle, a final extension of 7 min at 72°C was performed. Five microlitres of the reaction was run on an agarose gel to check for amplification, while the remaining 45 µl was digested with Proteinase K (Invitrogen) as follows: 45 µl PCR amplicon, 45 µl TE, 1 µl 0.5 M EDTA, 2.5 µl 20 mg/ml Proteinase K, and 5 µl 10% SDS were incubated at 37°C for one hour followed by DNA clean-up using the Qiaquick Gel Extraction Kit (Qiagen). This purified DNA was digested with the appropriate restriction endonucleases, the appropriate band purified from an agarose gel, and ligated into a similarly digested vector. Inverse PCR was done using the same protocol

as detailed above. All plasmid inserts constructed using PCR were verified for accuracy by sequencing. Primers used for constructing plasmids are listed in Table 2.

DNA fragments were run on 0.8 to 2.0 % agarose gels in 1X TAE buffer at 100V and purified using the Qiagen Qiaquick gel extraction kit according to the manufacturer's protocol.

DNA ligations were performed using the Rapid DNA Ligation Kit (Roche). Reactions containing 0.5 µl T4 DNA Ligase, DNA, and 5 µl 2X ligation buffer in a final volume of 10 µl were incubated for ten minutes at room temperature.

2.2.2. Transformation of Competent *E. coli* Cells

Fifty microlitres of competent *E. coli* (Invitrogen) were thawed slowly on ice. DNA was added and cells were left on ice for 30 minutes. Cells were then heat shocked at 42°C for 45 seconds and returned to ice. One hundred microlitres of Luria Broth (LB) was added to cells, which were subsequently plated on LB/ampicillin plates and grown overnight at 37°C. If a sensitive procedure was being performed, such as a ligation in which very little DNA was present, the cells were incubated at 37°C for one hour with 1 ml LB after heat shocking. DH5α *E. coli* were used for subcloning procedures and Tuner (DE3), Origami (DE3), or BL21 (DE3) cells were used for induction and subsequent purification of protein.

Table 1. Plasmids used in this thesis.

Plasmid Name	Construction Method or Plasmid Source	Features	Application
pET16b-AEBP1	Reference (*2*)	Encodes AEBP1 amino acids 27-730, N-terminal His6 tag	protein purification from *E. coli.*
pET16b-AEBP1ΔDLD	Reference (*8*)	Encodes AEBP1 amino acids 123-748, N-terminal His6 tag	"
pET21d-AEBP1	PCR amplified the AEBP1 ORF using primers AEBP1-1 and AEBP1-2. The amplicon was digested with NcoI and HindIII and ligated into the NcoI and HindIII sites of pET-21d(+). Also called pPL1 in lab notes.	Encodes entire AEBP1 protein, C-terminal His6 tag. NcoI at 5' and HindIII at 3'	"
pET21d-AEBP1ΔN	Same as above, but with primers AEBP1-2 and AEBP1-3. Also called pPL1ΔN in lab notes.	Encodes AEBP1 amino acids 166-748, C-terminal His6 tag	"
pET21d-AEBP1ΔC	Same as above, but with primers AEBP1-1 and AEBP1-4. Also called pPL1ΔC in lab notes.	Encodes AEBP1 amino acids 1-596, C-terminal His6 tag	"
pET21d-AEBP1ΔSTP	Same as above, but with primers AEBP1-1 and AEBP1-5. Also called pPL1-ΔSTP in lab notes.	Encodes AEBP1 amino acids 1-650, C-terminal His6 tag	"
pET21d-AEBP1-CP	Same as above, but with primers AEBP1-3 and AEBP1-4. Also called pPL1-CP in lab notes.	Encodes AEBP1 amino acids 166-596, C-terminal His6 tag	"
pET21d-AEBP1-C	Performed inverse PCR using pET21d-AEBP1 as a template and primers AEBP1-16 and AEBP1-17. Digested with NcoI, gel extracted vector fragment, and religated. Also called pPL1-C in lab notes.	Encodes AEBP1 amino acids 597-748, C-terminal His6 tag	"

pET21d-AEBP1-N	Performed inverse PCR using pET21d-AEBP1 as a template and primers AEBP1-16 and AEBP1-17. Digested with HindIII, gel extracted vector fragment, and religated. Also called pPL1-N in lab notes.	Encodes AEBP1 amino acids 1-165, C-terminal His6 tag. Does not express well.	"
pcDNA-AEBP1 WT, ΔN, ΔC, ΔSTP, CP, and C	pET21d-AEBP1 and derivative deletion mutants and pcDNA3.1/myc-HisA were digested with HindIII and XbaI and blunted with Klenow. The AEBP1 fragment was blunt-end ligated into the digested pcDNA plasmid. An in-frame stop codon within the XbaI site was utilized. Also called pPL2-WT, ΔN, etc., in lab notes.	Express same proteins as related pET21d plasmids (see above), but with no tag and an unrelated Lys-Leu at the C-terminus.	Overexpression of AEBP1 in mammalian cells.
pcDNA-AEBP1 Δ316-342	Inverse PCR using primers AEBP1-6 and AEBP1-7. Also called pPL2-Δ6:7 in lab notes.	Deletion of first large loop within the CP domain	"
pcDNA-AEBP1 Δ386-405	Inverse PCR using primers AEBP1-8 and AEBP1-9. Also called pPL2-Δ8:9 in lab notes.	Deletion of second large loop within the CP domain	"
pblue-12/11	AEBP1 ORF was amplified by PCR using primers AEBP1-11 and AEBP1-12. The amplicon was digested with EcoRI and XhoI and ligated into the EcoRI and XhoI sites of pBluescript II SK(+).	Encodes entire AEBP1 ORF including an N-terminal EcoRI site and a C-terminal stop codon and Xho-I site.	Subcloning
pPL3-WT	The EcoRI/BamHI 700bp 5' fragment from pBlue-12/11 and the BamHI/XhoI 1550bp 3' fragment from pPL1-WT were ligated into the EcoRI/XhoI sites of pcDNA3.1/myc-HisA (triple ligation).	C-terminally myc/His tagged AEBP1. (but has Zozak sequence problems – lower expression, and possibly multiple start sites.)	Overexpression of AEBP1 in mammalian cells.

pcDNA-AEBP1-Myc/His	The NdeI/BamHI (N-terminus of AEBP1) fragment from pcDNA-AEBP1 was used to replace the same fragment in pPL3. Also called pPL4 in lab notes.	Same Kozak sequence as pcDNA-AEBP1 and the myc/His tag from pPL3.	Overexpression of AEBP1 in mammalian cells.
paP2 (-168/+21)GL2	Digested pGL2 basic with XhoI and isolated vector. Digested aP2(-168/+21)CAT with SalI and isolated ~200 bp promoter fragment. Ligated. Obtained both orientations.	Luciferase reporter plasmid driven by the aP2 promoter.	Reporter assays
paP2 (+21/-168)GL2	See above	See above	See above
paP2 (3AE-1/-120) CAT	See reference (*7*)	CAT reporter plasmid driven by the aP2 promoter which includes 3 copies of the AE-1 site.	Reporter assays
pJ3H-AEBP1(-)	See reference (*7*)	AEBP1 ORF inserted in the reverse orientation	Negative control for overexpression experiments.
pJ3H-AEBP1	See reference (*7*)	Encodes AEBP1 amino acids 32-748, N-terminal HA tag	Overexpression of AEBP1 in mammalian cells.
pJ3H-AEBP1ΔSty	See reference (*7*)	Encodes AEBP1 amino acids 32-543, N-terminal HA tag	Overexpression of AEBP1 in mammalian cells.
Rc/CMV-EGFR	Obtained from G. N. Gill (*154*)	Encodes Epidermal Growth Factor Receptor	Overexpression of EGFR in mammalian cells.
pGEX2T-PTEN	Amplified the PTEN ORF using primers PTEN-1 and PTEN-2. Digested amplicon with BamHI and EcoRI and inserted into BamHI/EcoRI sites of pGEX2T	N-terminal GST tagged PTEN	recombinant protein purification in *E. coli*

pGEX2T-PTENΔ2C	Same as above, but using primers PTEN-1 and PTEN-3.	N-terminal GST tagged PTEN, lacking the two C-terminal amino acids.	recombinant protein purification in *E. coli*
pGEX2T-H3, H3m, ERK1, NHE3, PTEN, AQP2, CONTR	Complementary primers mH3(-) and (+), hNHE3(-) and (+), mERK1(-) and (+), mAQP2(-) and (+), hPTEN(-) and (+), mH3mut (-) and (+), control(-) and (+) were annealed and ligated into the BamHI and EcoRI sites of the pGEX2T plasmid.	GST with 12-amino acid C-terminal extension encoding the C-terminus of the protein stated.	recombinant protein purification in *E. coli*
pGEX2T-histone	Obtained from R.G. Roeder (*155*)	Four separate plasmids encoding the four core histones (H2A, H2B, H3 and H4), GST-tagged.	recombinant protein purification in *E. coli*

Table 2. Primers used for plasmid construction.

PRIMER	PRIMER SEQUENCE (5′→ 3′)
AEBP1-1	TATACCATGGAGTCACACCGCATTGA
AEBP1-2	TATAAAGCTTGAAGTCCCCAAAGTTCACTG
AEBP1-3	TATACCATGGTAACTACTGACAGCCTG
AEBP1-4	TATAAAGCTTGGATCGAGCCAGGATGAA
AEBP1-5	TATAAAGCTTATTGAGGCGACGCAGTCG
AEBP1-6	ATCTGGGAAGTCCTCGAAGA
AEBP1-7	CGTTACCTGTCCCCAGATG
AEBP1-8	GTCATAGGGATAAGACACAAGC
AEBP1-9	GAAGATGATGACGGGGTGTC
AEBP1-11	TATACTCGAGTCAGAAGTCCCCAAAGTTCAC
AEBP1-12	TATAGAATTCATGGAGTCACACCGCATTGA
AEBP1-16	TATAAAGCTTCACCTCATTCTGTGCGTAGT
AEBP1-17	TATACCATGGCAAACTGGAAGCGCATTCGGG
PTEN-1	TATAGGATCCATGACAGCCATCATCAAAGAG
PTEN-2	TATAGAATTCTCAGACTTTTGTAATTTGTGAATG
PTEN-3	TATAGAATTCTCATGTAATTTGTGAATGCTGATCTT
mH3(+)	gatccATCCAGCTGGCACGCCGTATCCGCGGGGAGCGGGCCTGAg
mH3(-)	aattcTCAGGCCCGCTCCCCGCGGATACGGCGTGCCAGCTGGATg
hNHE3(+)	gatccCGGCCCCCCGCCGCCCTCCCCGAGTCCACACACATGTGAg
hNHE3(-)	aattcTCACATGTGTGTGGACTCGGGGAGGGCGGCGGGGGGCCGg
mERK1(+)	gatccGAAGAGACTGCTAGATTCCAGCCAGGATACAGATCTTGAg
mERK1(-)	aattcTCAAGATCTGTATCCTGGCTGGAATCTAGCAGTCTCTTCg
mAQP2(+)	gatccCACTCTCCGCAGAGCCTGCCGCGCGGCAGCAAGGCTTGAg
mAQP2(-)	aattcTCAAGCCTTGCTGCCGCGCGGCAGGCTCTGCGGAGAGTGg
hPTEN(+)	gatccTTTGATGAAGATCAGCATACACAAATTACAAAAGTCTGAg
hPTEN(-)	aattcTCAGACTTTTGTAATTTGTGTATGCTGATCTTCATCAAAg
mH3mut(+)	gatccATCCAGCTGGCACGCCGTATCCGCGGGGAGGCCGCCTGAg
mH3mut(-)	aattcTCAGGCGGCCTCCCCGCGGATACGGCGTGCCAGCTGGATg
control(+)	gatccAGTCCCCCCACCGCCGGACCCGCTTCCACAGCAGGATGAg
control(-)	aattcTCATCCTGCTGTGGAAGCGGGTCCGGCGGTGGGGGGACTg

2.2.3. Site-directed Mutagenesis

All mutagenesis, except for AEBP1 T623A, was performed using the QuikChange site-directed mutagenesis kit (Stratagene). In this method, PCR is used to replicate both strands of a plasmid from mutagenic primers, resulting in a mutated plasmid with staggered nicks which are repaired in *E. coli* by DNA repair. The reaction mix consisted of 5 µl 10X cloned pfu buffer, 10-20 ng template DNA, 125 ng primer 1, 125 ng primer 2, 1 µl 10 mM dNTP mix, 1 µl PfuTurbo DNA Polymerase (Stratagene), and water to a total volume of 50 µl. The PCR cycle was as follows: 30 s at 95°C, followed by 12 cycles of 95°C for 30 s, 55°C for 1 min, and 68°C for 2 min/kb template. Ten µl was run on a 1% agarose gel to check for amplification and 1µl DpnI endonuclease was added to the remaining 40 µl in order to digest the methylated parental DNA template. A portion of this reaction was then transformed into competent DH5α *E. coli.* Individual colonies were picked, plasmid DNA purified, and the presence of the desired mutation was verified by sequencing. Mutagenesis of AEBP1 T623A was performed using the MORPH kit (Eppendorf-5 Prime) using the RO-P03 primer according to the manufacturer's intstructions. RO-P04 and RO-P04B primers were used for construction of the T623D AEBP1 mutant. The mutagenic function of all other primers, designed with the desired AEBP1 mutation in the middle, is indicated by name. Primers used for site-directed mutagenesis are listed in Table 3.

Table 3. Primers used for site-directed mutagenesis of AEBP1.

PRIMER	PRIMER SEQUENCE (5′→ 3′)
RO-P03	CCCCTCACGACCCATGGCCCCCCAGCAGCGGCGCATGCAG
RO-P04	CCCCTCACGACCCATGGACCCCCAGCAGCGGCGCATGCAG
RO-P04B	CTGCATGCGCCGCTGCTGGGGGTCCATGGGTCGTGAGGGG
S658A-sense	GCAGGCCCTGCCACAGCCCCCACTCCTGCC
S658A-antisense	GGCAGGAGTGGGGGCTGTGGCAGGGCCTGC
S658D-sense	GCAGGCCCTGCCACAGACCCCACTCCTGCC
S658D-antisense	GGCAGGAGTGGGGTCTGTGGCAGGGCCTGC
S668A-sense	CTTATGCCTCCCCCTGACCCTACACCAGCC
S668A-antisense	GGCTGGTGTAGGGGCAGGGGGAGGCATAAG
S668D-sense	CTTATGCCTCCCCCTGACCCTACACCAGCC
S668D-antisense	GGCTGGTGTAGGGTCAGGGGGAGGCATAAG
H236L-sense	CACAGCCGGGATCCTCGGCAATGAGGTGC
H236L-antisense	GCACCTCATTGCCGAGGATCCCGGCTGTG
E239A-sense	GGATCCACGGCAATGCGGTGCTAGGCCGAG
E239A-antisense	CTCGGCCTAGCACCGCATTGCCGTGGATCC
E363A-sense	CATTATTTCCTGGATGGCGAAGAACCCCTTTG
E363A-antisense	CAAAGGGGTTCTTCGCCATCCAGGAAATAATG
N374A-sense	GGGTGCAAATCTGGCCGGTGGTGAGCGGCTTG
N374A-antisense	CAAGCCGCTCACCACCGGCCAGATTTGCACCC
E377A-sense	GAACGGTGGTGCGCGGCTTGTGTCTTATCC
E377A-antisense	GGATAAGACACAAGCCGCGCACCACCGTTC
R388A-sense	CCCTATGACATGGCCGCGACACCTAGCCAGG
R388A-antisense	CCTGGCTAGGTGTCGCGGCCATGTCATAGGG
R442A-sense	GACGGAGCCCTACGCGGGAGGGTGCCAGGC
R442A-antisense	GCCTGGCACCCTCCCGCGTAGGGCTCCGTC
N470A-sense	GCTCTGGGACTTTCGCTGACTTTAGCTACC
N470A-antisense	GGTAGCTAAAGTCAGCGAAAGTCCCAGAGC

2.3. Purification of Recombinant Proteins

2.3.1. AEBP1

Two tubes of 3 ml LB (100 μg/ml ampicillin) were inoculated with BL21 (DE3) *E. coli* cells previously transformed with a pET-AEBP1 plasmid. This inoculation was either directly from frozen glycerol stock or from a freshly transformed bacterial colony. This inoculum was grown overnight at 30°C and shaking at 250 rpm. In the morning the entire 6 ml culture was transferred to 500 ml LB (containing 100 μg/ml ampicillin) in a 2 L flask and grown for approximately 3 hours at 37°C and 250 rpm until the OD_{600} was approximately 0.6. Expression of AEBP1 was then induced with 1-2 mM IPTG for 3.5 hours at 250 rpm and at 37°C. Cells were centrifuged at 6000g for 15 minutes at 4°C and frozen at -20°C at this point unless continuing further.

Cells were resuspended in 25 ml extraction/wash buffer (50mM Sodium Phosphate, pH 7.0, and 300mM NaCl) and lysed by passing through a French pressure cell twice at 14,000 psi. Immediately prior to lysis, 1mM phenylmethylsulfonyl fluoride (PMSF) was added. Following lysis, extracts were centrifuged at 15,000 x g for 20 minutes at 4°C. The supernatant was discarded and the pellet (containing insoluble inclusion bodies) was dissolved in 10 ml denaturing extraction/wash buffer (50 mM sodium phosphate, pH 7.0, 300 mM NaCl, and 6 M guanidine-HCl). This was centrifuged again at 15,000 x g for 20 minutes at 4°C to clarify.

His6-tagged protein was purified using Talon metal affinity resin (Clontech). A 0.75 ml bed volume of Talon resin was equilibrated twice with 10 ml denaturing extraction/wash buffer by mixing briefly and centrifuging at 700 x g for 2 minutes. The clarified protein sample was added to this resin and incubated for 30 minutes at room temperature on a rocking platform. This was then centrifuged at 700 x g for 5 minutes, and the resin washed twice with 10 ml of denaturing extraction/wash buffer, followed by additional washing (5-10 ml) on a 5 ml gravity-flow column until the $OD_{280} < 0.01$. When purifying pET16b-AEBP1, second batch wash and subsequent washing on the column was done with extraction/wash buffer at pH 6.7. Elution was performed with 3 ml imidizole elution buffer (45 mM sodium

phosphate pH 7.0, 5.4 M guanidine-HCl, 270 mM NaCl, 300 mM imidizole) and 0.5 ml fractions were collected.

Following analysis of protein concentration, fractions with similar concentrations were combined and dialyzed. Dialysis was performed stepwise in 50 mM sodium phosphate, pH 7.5, 100 mM NaCl, plus additives as follows:

Additive	Time	Volume
4 M urea, 1 mM β-mercaptoethanol, 0.2 mM $ZnCl_2$	2 hours	0.5 L
2 M urea, 1 mM β-mercaptoethanol, 0.1 mM $ZnCl_2$	2 hours	0.5 L
1 M urea	2 hours	0.5 L
0.5 M urea	2 hours	0.5 L
No additives	2 hours	0.5 L
10 % glycerol	overnight	1.5 L

2.3.2. GST fusion proteins

GST-PTEN was expressed in BL21 (DE3) *E. coli* and purified according to the procedure outlined in reference (*156*). A 500 ml culture was grown at 37°C until it reached mid-log phase as judged by an A_{600} = 0.6. Expression was induced by addition of IPTG to a concentration of 0.2 mM and growth was continued at room temperature overnight at 250 rpm. The bacteria were harvested by centrifugation and the pellet was resuspended in 5 ml cold 20 mM Tris (pH 8.0), 150 mM NaCl, and 5 mM EDTA. Just prior to lysis, PMSF was added to a concentration of 1 mM. Lysis was performed by passing the suspension through a French Press once at 14,000 psi. The lysate was cleared by centrifugation at 30,000 x g for 10 minutes and diluted with an equal volume of HBS (50 mM Hepes, 150 mM NaCl, pH 7.4). 300 µl Glutathione-Sepharose 4B (Pharmacia Biotech) was equilibrated with 7.5 ml HBS prior to addition of the lysate mixture. This was then incubated at 4°C on a rocking platform for 2 hours. The glutathione-Sepharose was washed 5 times with 10 ml HBS and the fusion proteins were eluted 3 times with 300 µl 20 mM glutathione, 50 mM Hepes, and 30% glycerol (pH 8.0). Fractions were combined and dialyzed in 50 mM Phosphate buffer, pH 7.5, 100 mM NaCl, 10% glycerol overnight.

GST-C terminal fusion proteins were induced and purified in the same manner as GST-PTEN, except that induction was carried out at 37°C for 3 hours. GST-histone was also induced at 37°C for 3 hours, but with 1 mM IPTG, and purified in the same manner as described above. Elution with the above-described GST elution buffer did not effectively elute GST-histones, and therefore these proteins were not eluted but used directly for GST pulldown assays.

2.4. Mammalian Cell Culture Methods

2.4.1. Cell line Maintenance

All cells were grown in an incubator at 37°C and containing 5% CO_2. NIH/3T3 cells, a mouse embryonic fibroblast cell line, were grown in DMEM (Gibco) supplemented with 10% calf serum and penicillin/streptomycin. Chinese Hamster Ovary (CHO) cells were grown in DMEM supplemented with 5% FBS, penicillin/streptomycin, and 37 μg/ml L-proline.

2.4.2. Transfection

Transfections were done in 60 mm dishes or 12-well plates with Polyfect (Qiagen) according to the manufacturer's protocol. Briefly, 3.0 X 10^5 NIH/3T3 cells were plated in a 60 mm plate and 8.0 X 10^5/1.5 X 10^5 (60 mm dish/12-well plate) CHO cells were plated the day before transfection. The day of transfection, 3/0.6 μg (60 mm dish/12-well plate) plasmid DNA was diluted to 150/30 μl with DMEM lacking any serum or antibiotics. 15/3 μl polyfect was added to diluted DNA and incubated for ten minutes to allow complexes to form. During this incubation period, cells were washed once with 4/1 ml PBS and 3/0.6 ml growth media added to cells. After incubation, 1/0.2 ml growth media was added to the DNA/polyfect mixture and all was transferred to cells. Cells were grown an additional 24-48 hours before harvesting.

2.5. Protein Analysis

2.5.1. Preparation of Extracts

Cell extracts were prepared in cold radioimmune precipitation buffer (RIPA buffer: 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 50 mM $Na_4P_2O_7$, 0.25% sodium deoxycholate, 0.1% Nonidet P-40, 1 mM Na_3VO_4, 1 mM NaF) containing 1 mM PMSF, 5 mM EDTA, 5 mM EGTA, and Complete Protease Inhibitor Cocktail (Roche).

Nuclear extracts for EMSA and subcellular fractionation studies were made for some experiments by mechanical fractionation as described in (*3*). Cells from a 60 mm dish were collected in 100 µl fractionation buffer (2 mM EDTA, 2 mM EGTA, 20 mM Tris-HCl, pH 7.5) supplemented with protease inhibitors, and passed through a 25 gauge needle 25 times to break the cell membranes. Cells were centrifuged for 5 minutes at 1000 x g and the pellet washed twice with fractionation buffer. The pellet (nuclear fraction) was then resuspended in RIPA buffer. The supernatant was additionally centrifuged for 1 hour at 19,000 x g to separate the cytoplasmic fraction from the plasma membrane-enriched fraction.

Some nuclear extracts were made using the Nuclear Extract Kit from ActiveMotif. Briefly, cells were washed with PBS/phosphatase inhibitors and harvested in the same. Cells were centrifuged and the pellet resuspended in a hypotonic buffer, incubated on ice for 15 minutes, after which a detergent was added and cells vortexed. This suspension was then centrifuged quickly at 14,000 x g to separate the supernatant (cytoplasmic fraction) from the pellet (nuclei). The nuclear pellet was then resuspended in 100 µl lysis buffer, incubated on a rocking platform for 30 minutes at 4°C, vortexed, and centrifuged to remove any cellular debris from the nuclear fraction.

2.5.2. Protein Concentration Determination

Protein concentration was measured using the Bio-Rad Protein Assay dye, which is based on the method of Bradford (*157*). 790 µl water was mixed with 10 µl sample and 200 µl Protein Assay Dye. After incubation for 5 minutes, absorbance was measured at A_{595} and compared with a standard curve made using known concentrations of bovine serum albumin (BSA, Sigma).

2.5.3. SDS-PAGE

Proteins were resolved on 6, 8.5, 10, or 12% polyacrylamide gels according to established methods. 10X running buffer consisted of 30.2 g Tris Base, 144 g glycine, and 10 g SDS in 1 liter. 6X SDS sample buffer (7 ml 4X stacking gel buffer, 3 ml glycerol, 1 g SDS, and 1.2 mg bromophenol blue) was added to samples which were then boiled for 3 minutes prior to loading on gel and run at 100-180 V, depending on composition of samples.

2.5.4. Protein Detection by Western Blotting or staining

Following SDS-PAGE, proteins were transferred to membrane by Western blotting or gels were washed three times with water and stained with GelCode Blue (Pierce) for 1 hour to overnight. Destaining was performed with water until complete. Gels were then soaked in coomassie destaining solution (5% methanol, 7% acetic acid, 88% water) overnight in order to fix proteins and avoid cracking of the gel, followed by drying between cellophane on a plastic frame for approximately one week.

In order to analyze resolved proteins by immunoblotting, supported nitrocellulose membrane (Bio-rad) and two pieces of filter paper cut to the size of the gel were soaked in transfer buffer (48 mM Tris-HCl, 39 mM glycine, 3.75 ml 10% SDS/litre, and 20% methanol) for 5 minutes. A sandwich was made of the gel, membrane, and filter papers and the proteins were transferred using a Trans-Blot SD Semi-dry Transfer Cell (Bio-Rad) for 50 minutes at approximately 400 mA.

Western blotting was begun by blocking the membranes with 5% non-fat milk powder in TBST (5X = 30.5 g Tris, 73 g NaCl, 10 ml Tween-20 in 1 L, pH 8.0), shaking either overnight at 4°C or for 1 hour at room temperature. Membranes were then incubated with primary antibody diluted in either 5% milk in TBST or 1% BSA in TBST (αHis-tag antibody, Qiagen) for either 1 hour at room temperature or 4°C overnight. The anti-p85 antibody used in Figure 25 was used as unpurified serum diluted 1:1000 in PBS. Membranes were washed three times with TBST (15, 5, 5 minutes) and incubated with secondary antibody for 1 hour at room temperature. Following three washes (15, 5, 5 minutes), proteins were detected using the ECL (Amersham) detection kit, or 5-bromo-4-chloro-3-indolyl-phosphate in conjunction with

nitro blue tetrazolium (BCIP/NBT) from Promega for colorimetric detection of alkaline phosphatase-conjugated secondary antibodies.

2.6. Assays

2.6.1. β-Galactosidase Assay

β-Galactosidase activity was assayed in order to normalize for transfection efficiency. 30 µl of extract was incubated with 3 µl Mg^{2+} buffer (10 µl 1M $MgCl_2$, 35 µl β-mercaptoethanol, 55 µl water), 66 µl ONPG (4 mg/ml in 0.1 M $NaHPO_4/Na_2PO_4$, pH 7.4), and 201 µl 0.1 M $NaHPO_4/Na_2PO_4$ buffer (pH 7.4) until color change was sufficient. Reactions were stopped by the addition of 0.5 ml 1 M Na_2Co_3, and enzyme activity measured spectrophotometrically at A_{410}.

2.6.2. Chloramphenicol Acetyl Transferase (CAT) Assay

Transfection of NIH/3T3 or CHO cells was generally performed with 3 µg pJ3H-AEBP1 or other pJ3H derivative expression plasmid, 200 ng paP2(3AE-1/-120)CAT reporter plasmid, along with 200 ng pHermes-lacZ, which expresses the lacZ gene under the control of the CMV promoter. Cells were harvested 48 hours after transfection. Cells were washed twice with cold PBS, collected in 1 ml PBS, and centrifuged at 6500 rpm for 5 minutes. The pellet was resuspended in 100 µl CAT/TE buffer (250 mM Tris pH 8.0, 5 mM EDTA) and cells were lysed by freezing in liquid nitrogen for 3 minutes and thawing at 37°C for 3 minutes. This freeze-thaw cycle was performed three times. The supernatant was collected following centrifugation at 14,000 x g at 4°C for 10 minutes.

CAT activity was assayed in a similar manner to that described by Gorman *et al.* (*158*). Cell extracts (after removing a portion for a β-galactosidase assay) were heated at 60°C for 10 minutes to inactivate any endogenous deacetylases in the extract. 70 µl of extract (or volume determined by β-galactosidase assay diluted to 70 µl with CAT/TE buffer) was incubated with 10 µl reaction buffer (10 reactions: 2 mg acetyl-CoA, 100 µl CAT/TE, 5 µl ^{14}C-chloramphenicol (2.5 µCi)) for 3-6 hours at 37°C. Reactions were stopped by adding 0.5 ml ethyl acetate, followed by vortexing for 30 seconds and centrifugation for 1 minute. The top

organic layer was collected, dried in a Speed Vac, and the residue redissolved in 10 µl ethyl acetate. This was spotted on a thin layer chromatography (TLC) silica plate (Sigma) and run for 30 minutes in a solvent of 95% chloroform/5% methanol. The plates were air dried and radioactivity quantified using a phosphorimager (Bio-Rad).

2.6.3. Luciferase Assay

Cells were transfected in a 12-well dish. 48 hours after transfection, cells were washed twice with 1 ml PBS and 150 µl Passive Lysis Buffer (Promega) was added. The plate was incubated on a shaker at room temperature for 20 minutes for lysis to proceed. 50 µl of lysate was transferred to a 96-well plate. Luciferase activity was measured by injection of 50 µl Luciferase Reagent (Promega) and measurement of luminescence by a FLUOstar Galaxy luminometer (BMG Labtechnologies).

2.6.4. Electrophoretic Mobility Shift Assay (EMSA)

Double-stranded DNA oligonucleotides were made by annealing complementary primers. In this study, AE-1 probe was made by annealing the 38mer AE1-sense (GATCCAGGGAGAACCAAAGTTGAGAAATTTCTATTAAA) and the AE1-antisense (GATCTTTAATAGAAATTTCTCAACTTTGGTTCTCCCTG) primers. This was done by mixing 37.6 µl of each 10 µM primer with 4.0 µl 1M Tris, pH 8.0, and 0.8 µl 1M $MgCl_2$ and placing in 90°C water. Primers were annealed by slowly cooling the water to < 37°C while constantly stirring.

AE-1 duplex oligonucleotide was radiolabelled by incubating one microlitre AE-1 duplex oligonucleotide with 2 µl 10X Klenow reaction buffer, 5 µl [α-^{32}P]dATP, 1 µl each of other 0.1 µmole/µl nucleosides, 8 µl water, and 1 µl Klenow enzyme at 37°C for 30 minutes. The labeled probe was purified using a Microspin G50 (Amersham) column according to the manufacturer's protocol. Radioactivity of one microlitre was measured by scintillation counting, and the probe subsequently diluted to the appropriate specific activity.

Recombinant proteins were incubated with DNA probe for 25 minutes at room temperature in binding buffer (10 mM Tris, pH 7.5, 10 mM KCl, 5 mM MgCl2, 1 mM DTT, 2.5 % glycerol). Samples were resolved on 5% 0.25x TBE mini-gels (2.5 ml 29:1

acrylamide:bisacrylamide, 375 µl 10X TBE, 95 µl 10% APS, 10 µl TEMED, 12.0 ml water) which were then dried and exposed to film. EMSA using nuclear extracts was performed as above, but with 2 µg nuclear extracts along with 0.1 µg poly dI:dC.

2.6.5. GST Pull-Down Assays

Ten microlitres Glutathione Sepharose 4B beads (Pharmacia Biotech) were washed with 1 ml HBS to equilibrate beads and remove ethanol. In order to bind GST or GST fusion proteins to beads, 30 µl phosphate buffer (50 mM sodium phosphate, 100 mM NaCl, pH 7.5) and 5-7 µg GST protein was added to the 10 µl of washed Glutathione Sepharose 4B beads and incubated at room temperature for 30 minutes on a rocking platform. Beads were washed twice with 300 µl HBS, and 150 µl HBS/0.1% Nonidet P-40 (NP40) and 5 µg recombinant AEBP1 were added and incubated on a rocking platform for 45 minutes at room temperature. In some cases the initial binding of GST or the GST fustion protein to the beads was omitted, and instead this protein was added to the pulldown step so that interaction of the GST moiety with the Glutathione Sepharose and AEBP1 interacting with the GST fustion protein occurred in one step. The beads were washed 5 times with 1 ml HBS/0.1% NP40 and 20 µl 1X SDS-PAGE sample buffer was added. The bound proteins were resolved by SDS-PAGE and the gels either transferred to nitrocellulose for immunoblotting or stained to visualize proteins.

2.6.6. CaM-Agarose Pull-Down Assays

Thirty microlitres calmodulin-agarose (Sigma) was washed with 0.5 ml HBS. 150 ul HBS/0.1% NP40 was added along with 5 µg recombinant AEBP1 and 2 mM $CaCl_2$ or 10 mM EGTA and incubated for 45 minutes at room temperature on a rocking platform. The beads were then washed four times with 1 ml HBS/0.1% NP40, 20 µl SDS-PAGE sample buffer was added, and bound proteins were resolved on an 8.5% SDS-PAGE gel. Proteins were visualized by staining with GelCode Blue (Pierce).

2.6.7. Kinase Assay

One half to 1.0 microgram recombinant AEBP1 was incubated with 50 units p42 MAP kinase (NEB) and 10 μCi [γ-^{32}P]ATP in the supplied MAP kinase reaction buffer (50 mM Tris-HCl pH 7.5, 10 mM MgCl2, 1 mM EGTA, 2 mM DTT, 0.01% Brij 35) for 30 minutes at 30°C. The reaction mixture was then analyzed by SDS-PAGE on an 8.5% gel followed by transfer to nitrocellulose membrane (Amersham) and autoradiography. Alternatively, the gel was dried and exposed to film.

Chapter 3

RESULTS

3.1 Structural Analysis of AEBP1

Molecular modeling *in silico* has become a useful approach for protein structure/function studies when methods such as X-ray crystallography are unavailable or have been unsuccessful. Modeling by homology is now widely utilized in structural characterization of proteins (*159, 160*). In order to learn more about AEBP1, I employed homology modeling using the program Modeller (*142*) to predict the tertiary structure of the protein based on similarity to a protein domain of known structure. AEBP1 has previously been characterized as containing three domains, a central carboxypeptidase (CP) domain with homology to regulatory CPs such as CPE and CPD, an N-terminal domain with homology to the discoidin-like domains (DLD) of coagulation factors V and VIII, and a C-terminal structurally uncharacterized domain. X-ray crystal structures have been solved for domains homologous to the CP and DLD domains of AEBP1, and I present here the homology models constructed for AEBP1. For domains of AEBP1, such as the C-terminus, where no homologous template structure was known, I relied on secondary structure predictions. Finally, I have identified homologues of AEBP1 from a variety of mammalian and non-mammalian species and have used this information as a starting point for identifying domains and residues unique to AEBP1 that may be critical for its function.

3.1.1. DLD domain modeling

The N-terminal domain of AEBP1 has been named the discoidin-like domain, due to homology with the Discoidin I protein from *D. discoideum*. X-ray crystal structures have recently been solved for two proteins with homology to the DLD domain of AEBP1 - the C2 domains of human coagulation factors V (PDB ID 1CZT) and VIII (PDB ID 1D7P). These C2 domains are composed of a β-barrel framework from which three spikes protrude. These

spikes are involved in membrane interaction through hydrophobic residues at the apices of the spikes which embed themselves in the hydrophobic membrane core (*138, 139*). Membrane phosphatidylserine head groups also make interactions with the grooves between the spikes of the C2 domains. The C2 domains of both coagulation factors V and VIII share considerable sequence similarity (~50%) with the DLD domain of AEBP1 and so were considered suitable candidates for homology modeling of AEBP1. A PSI-BLAST search of databases did not detect any other suitable templates.

The optimal alignment of the sequences of the DLD domain of AEBP1 and these two C2 domains, based on a BLAST search, is shown in Figure 7. Through multiple modeling experiments using both of these C2 domain structures as templates for the DLD domain of AEBP1, and subsequent quality assessment using ProsaII and Procheck, it was determined that the 1CZT structural template produced a higher quality model than 1D7P. However, modeling with 1CZT as a template continued to produce an energetically unfavorable conformation in a region around amino acids 25-40 of AEBP1 DLD, as determined from a ProsaII plot (Figure 8A). Through trial and error it was found that free modeling of amino acids 27-32 within this region using Modeller's *de novo* loop modeling ability removed this steric energy clash (Figure 8B), and this final model of the DLD domain of AEBP1 was named 1DLD.

Further analysis of the ProsaII plot produced for 1DLD indicates that the area around amino acids 30-50 is still somewhat strained, despite free modeling 6 amino acids in this region (Figure 8B). This area contains a 7-amino acid negatively charged loop not present in the template 1CZT. Modeller placed this loop so that it is largely occluded from the protein surface. This may be part of the reason for the strain in this area, as a charged loop would naturally tend to be on the hydrophilic surface of a protein. However, the template structure is itself generally strained in this region also, and so we can conclude that our model is no worse than the X-ray crystal structure against which it was modeled. In fact, for much of the sequence our model is energetically less strained than its template. The Ramachandran plots for 1DLD (Figure 8C) and its template (not shown) confirm the above observations. 1DLD contains one amino acid, leucine 28, in a disallowed region, while 1CZT contains two leucines

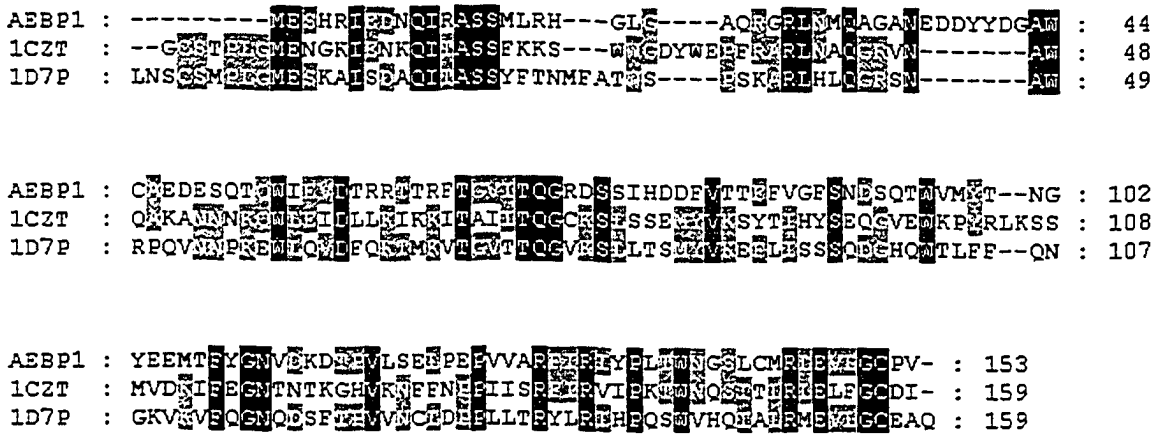
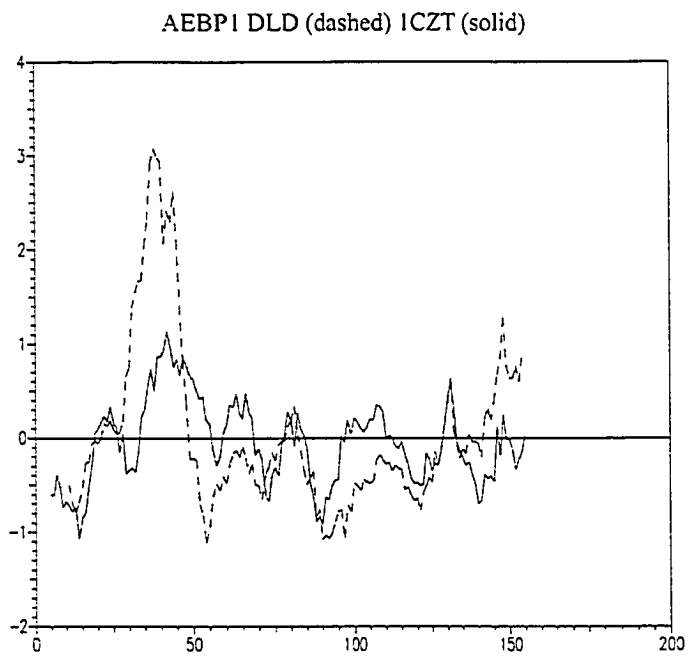


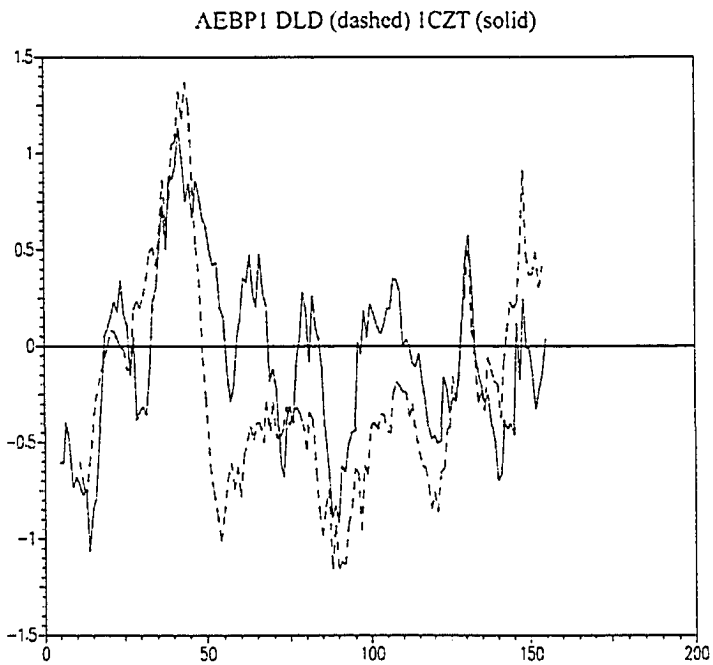
Figure 7. Alignment of the amino acid sequences of the DLD domain of AEBP1 and coagulation factor C2 domains. The amino acid sequence of the discoidin-like domain (DLD) of AEBP1 was aligned with the amino acid sequences of two homologous domains for which X-ray crystal structures have been solved - 1CZT (coagulation factor V C2 domain) and 1D7P (coagulation factor VIII C2 domain). The alignment was performed with ClustalW followed by shading with GeneDoc. A black background indicates identical residues and a grey background indicates chemically similar residues.

Figure 8. Assessment of 1DLD model quality. ProsaII was used to produce an energy diagram of the modeled AEBP1 DLD domain in comparison to its corresponding template, 1CZT. Both the 1DLD model based on the ClustalW alignment of AEBP1 and 1CZT (A) as well as the 1DLD model in which AEBP1 amino acids 27-32 were free-modeled (B) are shown. The x-axis indicates amino acids, while the y-axis is energy, represented in units of E/kT. A Ramachandran plot for this second 1DLD model (in which 6 amino acids were free-modeled) was produced by Procheck (C).

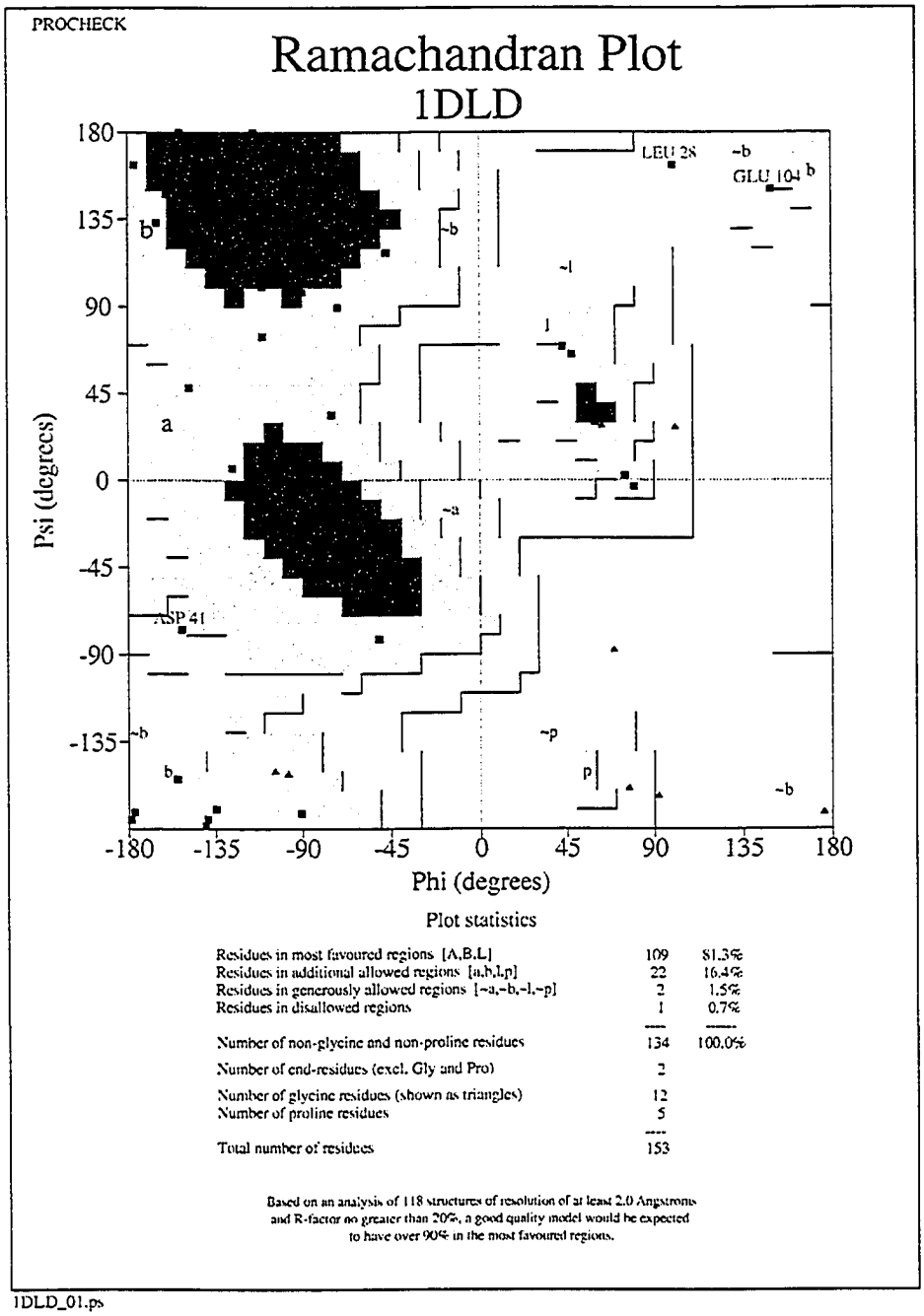
A.


AEBP1 DLD (dashed) 1CZT (solid)
4
3
2
1
0
-1
-2
0
50
100
150
200

B.


AEBP1 DLD (dashed) 1CZT (solid)
1.5
1
0.5
0
-0.5
-1
-1.5
0
50
100
150
200

C.


PROCHECK
Ramachandran Plot
1DLD
Psi (degrees)
Phi (degrees)
180
135
90
45
0
-45
-90
-135
-180
LEU 28
GLU 104
ASP 41
a
b
~a
~b
l
~l
p
~p
Plot statistics
Residues in most favoured regions [A,B,L] 109 81.3%
Residues in additional allowed regions [a,b,l,p] 22 16.4%
Residues in generously allowed regions [~a,~b,~l,~p] 2 1.5%
Residues in disallowed regions 1 0.7%
Number of non-glycine and non-proline residues 134 100.0%
Number of end-residues (excl. Gly and Pro) 2
Number of glycine residues (shown as triangles) 12
Number of proline residues 5
Total number of residues 153
Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

1DLD_01.ps

in disallowed regions. Based on these quality control measures the model was judged to be of a suitable quality for further study.

A closer look at the sequence and modeled structure of the DLD domain of AEBP1 reveals several interesting points. First, two cysteines, C45 and C143, are found in very close 3-dimensional proximity to each other. These cysteines are not found in the template molecules 1CZT or 1D7P but are conserved in all homologues of AEBP1 found to date, from mammals to fish and birds. The proximity of these cysteines to each other in the tertiary structure, as well as total conservation, suggests that they may form a disulfide bond contributing to the stability of the structure of the domain, and so I have constrained my model to include this disulfide bond (Figure 9). One additional conserved cysteine is present in this model at position 151. A homologous cysteine is also present in the discoidin domains of both Factors V and VIII and is disulfide bonded to a cysteine located at the extreme N-terminus of the domains (see Figure 7). This cysteine is lacking in AEBP1, which is missing 7 amino acids at the N-terminus of the DLD domain, but the sequence for ACLP indicates a conserved cysteine shifted just one amino acid C-terminal from the cysteine present in the discoidin domains of Factors V and VIII and is likely disulfide bonded with the cysteine at the C-terminus of the DLD domain.

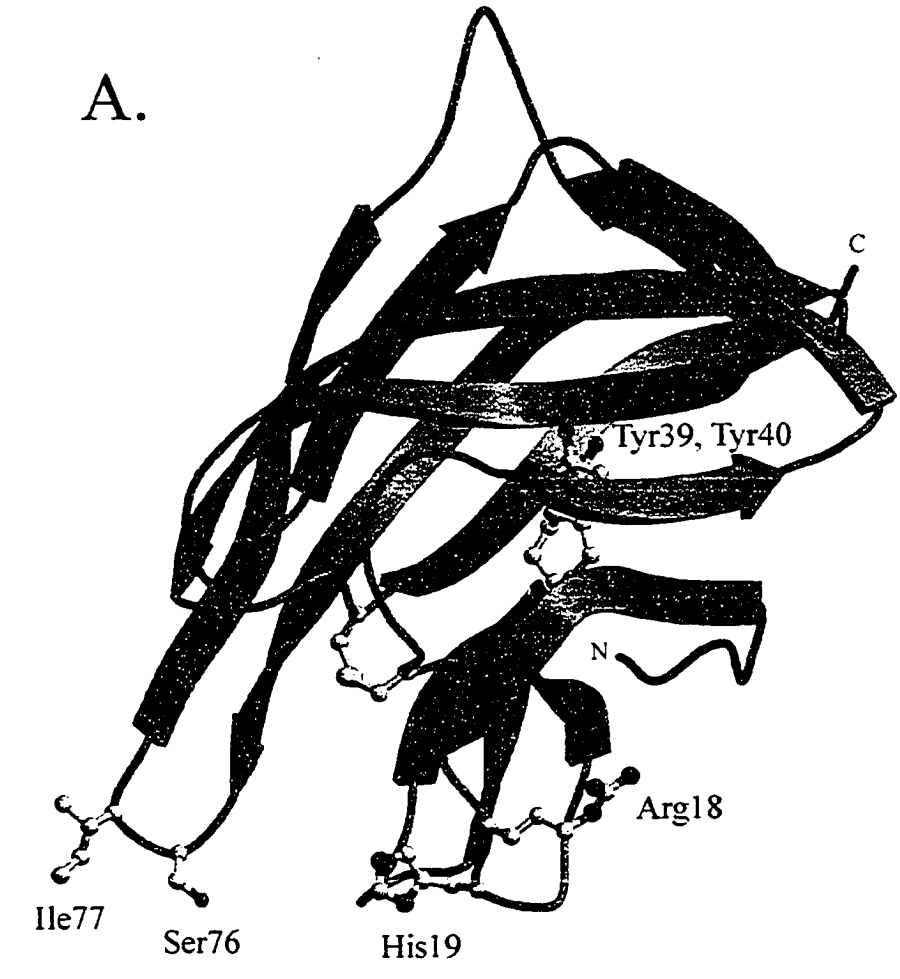
It is interesting to compare the structure and function of the Factor V/VIII C2 domains with the DLD domain of AEBP1/ACLP. The coagulation factor C2 domains are involved in membrane interactions by inserting hydrophobic residues located at the tips of three spikes into the lipid bilayer. These residues in the Factor V C2 domain (1CZT), used as a template for AEBP1, are W27 and W28 (spike 1), R44 (spike 2), and L80 (spike 3). The Factor VIII C2 domain also has three spikes, capped by M29 and F30 (spike 1), R45 (spike 2), and L81 and L82 (spike 3). It is proposed that the domain lies such that the hydrophobic spikes insert into the membrane, while the basic spikes and surrounding basic residues interact with the negatively charged surface of the membrane. The model of the DLD domain of AEBP1 also indicates several spikes (Figure 9). AEBP1 does not contain hydrophobic residues at the tip of spike 1, but instead contains an arginine and a histidine, the histidine being strictly conserved, similar to spike 2 of the coagulation factor C2 domains. A potential surface exposed loop is suggested by the sequence ^{33}GANEDDYYDG^{42} which contains a well-conserved 7-amino acid loop not present in the template structures. This loop contains four acidic amino acids and

two tyrosines, the tyrosines being modeled at the apex of the loop. In my model this loop is placed so that the tyrosines at the apex are inserted into the central cavity of the domain. If this loop were to face the opposite direction, these tyrosines would be on the surface of the protein, possibly functionally replacing the hydrophobic residues found at the tip of spike 1 of the C2 domain of coagulation factor V (*138*). AEBP1 Ser76 and Ile77 are structurally equivalent to spike 3 of the coagulation factor C2 domains. Ile77 is conserved in most AEBP1 orthologues identified except Fugu and Tetraodon (see Figure 14). Altogether it appears that the DLD domain of AEBP1/ACLP is structurally similar to coagulation factors V and VIII C2 domains and may also function in a similar manner in the interaction with membrane surfaces.

Figure 9 also shows the polarity of the domain. A smaller N-terminal subdomain of the DLD domain is composed of the N-terminal 45 amino acids. This subdomain is very highly charged, with 7 basic amino acids out of a total of 16 in the DLD domain (see figure 9B), while the opposite face of the DLD domain (Figure 9C) is largely hydrophobic and non-charged. In fact, it appears that the predicted disulfide bond may serve to hold the highly charged subdomain packed tightly against the rest of the DLD domain.

The DLD domain also contains a short glycine-rich sequence of interest within this 45-amino acid subdomain, ^{20}GLGAQRGR^{27}. This glycine-rich loop has some sequence-similarity to the P-loop motif involved in nucleotide binding (*161*) and the ATP-regulatory module (ARM) sequence motif found in receptor guanylate cyclases (*162, 163*). Many of the above-described aspects of the DLD domain of AEBP1 await investigation and were not pursued further in this thesis.

Figure 9. The DLD domain of AEBP1 is a β-barrel with two distinct faces. (A) One predicted disulfide bond is indicated by a ball and stick structure, connecting a highly charged N-terminal subdomain to the remainder of the domain. β-strands are indicated in green and α-helices in red. Amino acids 36-42, which could potentially be solvent exposed, forming a spike, are shown in red. All residues potentially forming the tips of spikes are shown as ball-and-stick structures and labelled. The N- and C-termini are labelled as such. (A) was created with Molscript/Raster3D. The N-terminal subdomain contains a large percentage of basic amino acids, thus resulting in a polarization of the DLD domain of AEBP1 with one face being highly positively charged (B) and the other face being uncharged/hydrophobic (C). Basic residues are indicated in blue, acidic residues in red, polar residues in yellow, and non-polar residues in white. (B) and (C) were created with Swiss PDB Viewer.


A.
C
Tyr39, Tyr40
N
Arg18
Ile77
Ser76
His19
B.
C.

3.1.2. CP domain modeling

The central half of the AEBP1 protein consists of a carboxypeptidase (CP)-like domain. This domain is a member of the regulatory carboxypeptidase family due to its sequence and structural similarities. In fact, AEBP1 has been found to exhibit low carboxypeptidase B-like enzymatic activity on the synthetic substrates hippuryl-arginine and hippuryl-lysine but not hippuryl-phenylalanine (*2, 164*). Recently, the first X-ray crystal structure of a member of the regulatory CP family has been solved (*113*). Carboxypeptidase D (CPD) contains three tandem repeats of a CP domain. The first domain is more efficient towards C-terminal arginine, the second domain is more efficient towards C-terminal lysine, and the third domain is inactive (*125*). F. X. Aviles and colleagues (*113*) determined the three-dimensional structure of the second domain of CPD, the more highly conserved of the two active domains. This domain has 40% sequence identity and 54% similarity to the CP domain of AEBP1, and so was considered to be a suitable template for modeling of the CP domain of AEBP1.

The crystal structure of duck CPD-II (PDB ID 1QMU) was used as a template for AEBP1, as well as a subsequently solved structure of duck CPD-II complexed with GEMSA (PDB ID 1H8L, Ref. (*127*)), a specific inhibitor of the enzyme. An optimal alignment was easily made because of the high sequence similarity (Figure 10). Two large gaps are present in the alignment due to two stretches of AEBP1 amino acid sequence, residues 316-342 and residues 386-405, that are not present in CPD or in most other CPs. The first stretch of amino acids is weakly conserved in a bacterial carboxypeptidase, CPT, for which a crystal structure has been solved (*165*). Some attempts were made to include this sequence in the model by using the structure for CPT to model this loop. However, the sequence similarity was not great enough to warrant including this loop in the model and therefore neither loop was included.

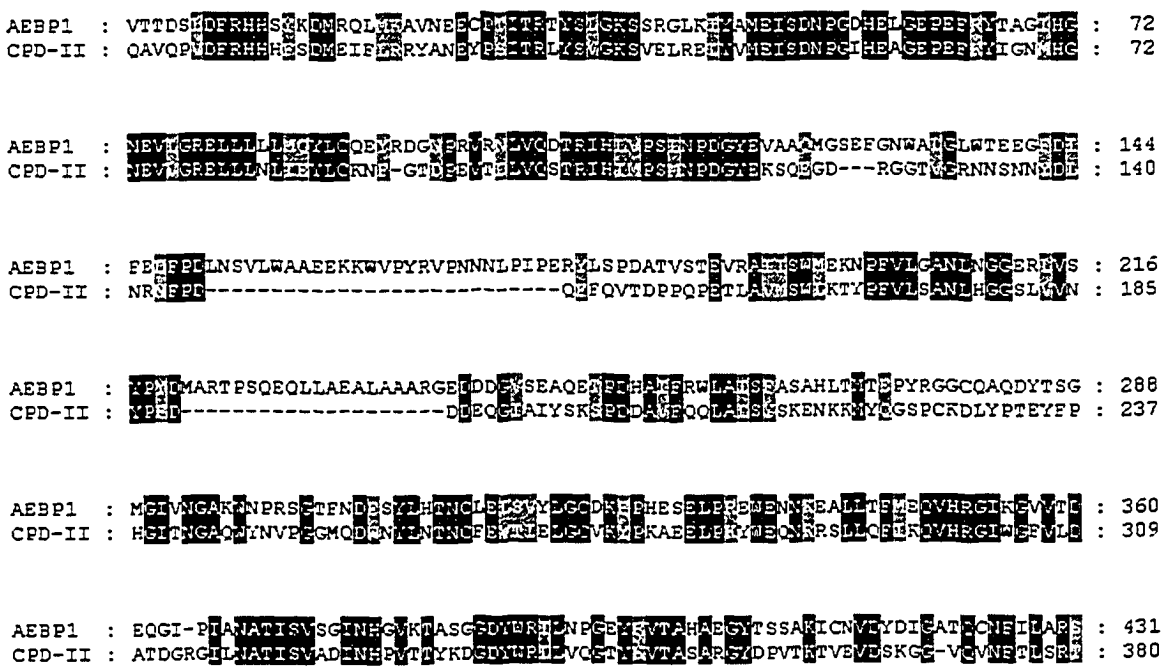
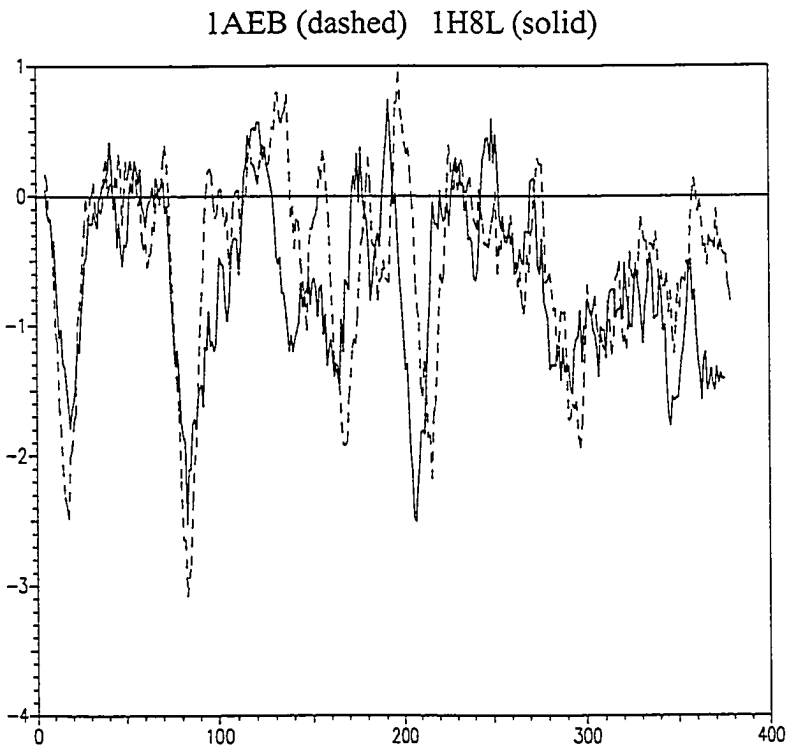


Figure 10. Alignment of the amino acid sequences of the CP domain of AEBP1 and duck CPD-II. Mouse AEBP1 was aligned using ClustalW with the protein sequence for the duck CPD-II domain as used for the determined X-ray crystal structure, 1QMU. A black background indicates identity while a grey background indicates similarity. Shading was produced using GeneDoc.

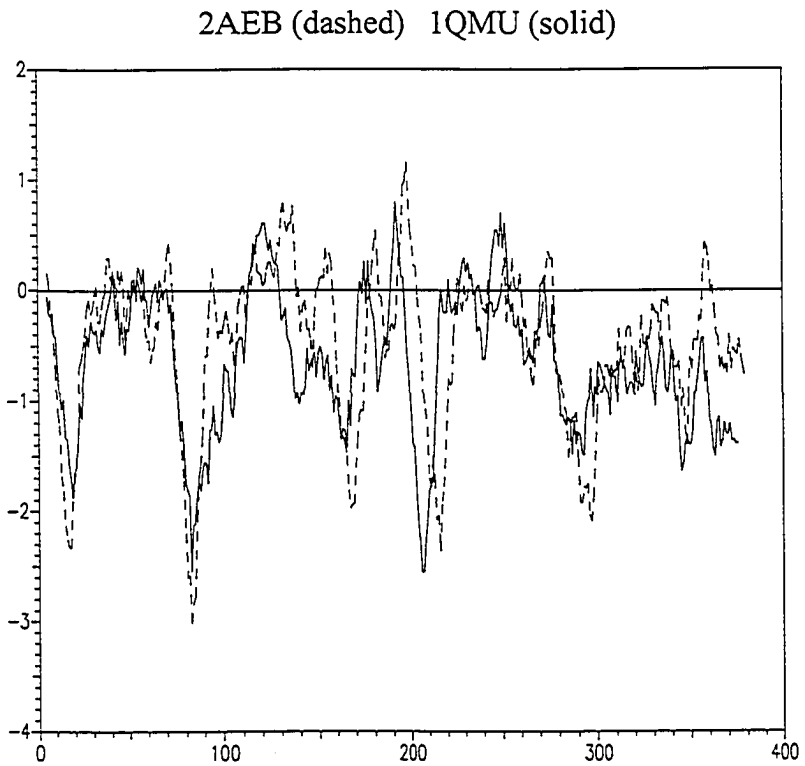
Quality control for both AEBP1 CP domain models (using 1H8L and 1QMU as templates) was performed as with 1DLD using ProsaII and Procheck (Figure 11). The ProsaII energy diagrams for the models and their respective templates are generally of equal quality, with the majority of the graph below zero. There are several locations in both models in which the model plot spikes above zero when the template plot is below zero. The first of these is around amino acids 130-140 of the CP domain of AEBP1. This is an area of low similarity between AEBP1 and CPD and follows a short 3 amino acid gap in the alignment (Figure 10), suggesting that the suboptimal energy plot at this point is due to a suboptimal three-dimensional model at this point also. The next spike is located at amino acid 150, the location at which a 27-amino acid loop was deleted from AEBP1 to create these CP domain models. The presence of this large loop would be expected to change the local structure in this area. Finally there is a small energy spike around amino acid 360 at the C-terminus of the CP domain. In the wild-type AEBP1 protein, the C-terminal domain likely affects the conformation of this area resulting in a more relaxed structure at this location. According to the Ramachandran plots for these models, neither model has any amino acids in stereochemically disallowed regions (Figure 11C and D). Based on the high similarity with the CPD template and the quality control measures taken, I felt that this model was of a suitable quality for further study.

Figure 11. Assessment of the quality of AEBP1 CP domain models. ProsaII was used to produce an energy diagram (A,B) of the AEBP1 CP domain (1AEB, 2AEB) modeled against two templates, 1H8L (A) and 1QMU (B). The x-axis indicates amino acids, while the y-axis is energy, represented in units of E/kT. Ramachandran plots of these models were produced by Procheck (C,D).

A.


1AEB (dashed) 1H8L (solid)
1
0
-1
-2
-3
-4
0
100
200
300
400

B.

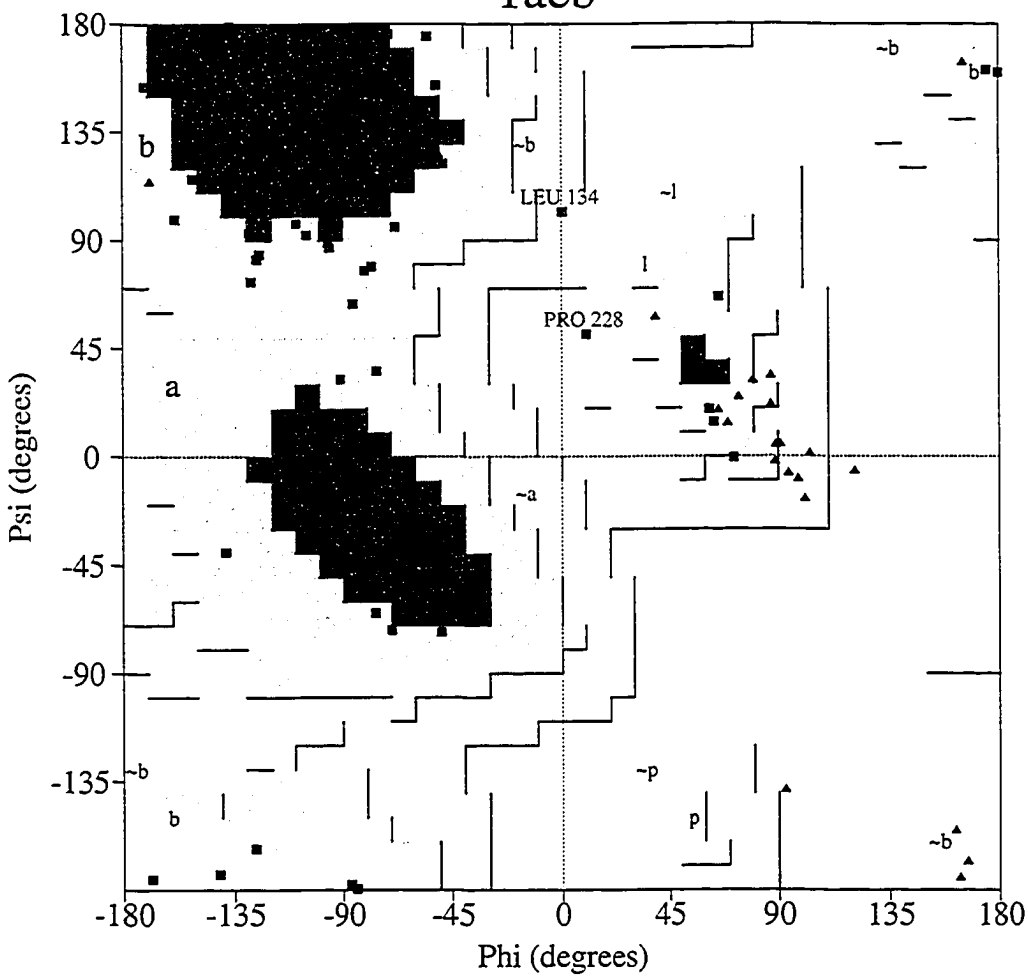

2AEB (dashed) 1QMU (solid)
2
1
0
-1
-2
-3
-4
0
100
200
300
400

C.

PROCHECK

Ramachandran Plot

1aeb



Plot statistics

Residues in most favoured regions [A,B,L]	298	90.0%
Residues in additional allowed regions [a,b,l,p]	32	9.7%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.3%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	331	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	34	
Number of proline residues	17	
Total number of residues	384	

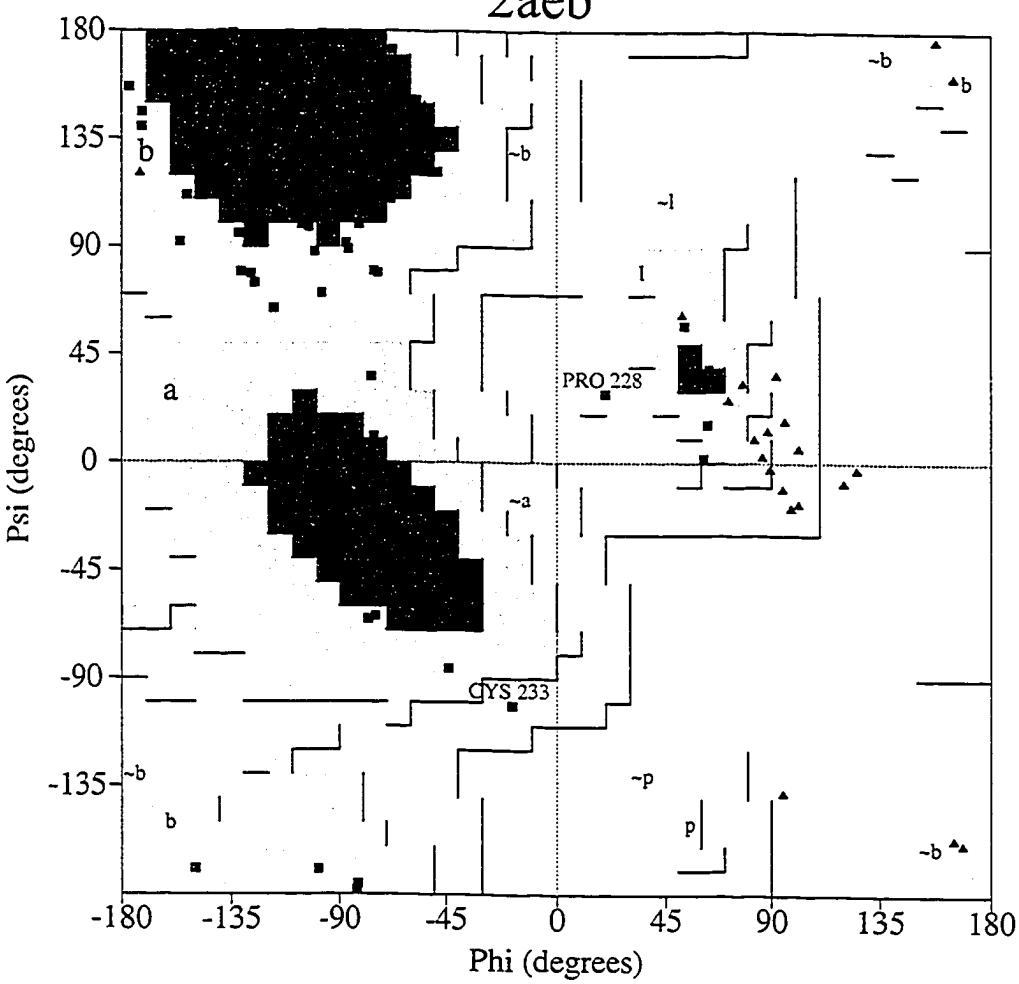
Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

D.

PROCHECK

Ramachandran Plot

2aeb



Plot statistics

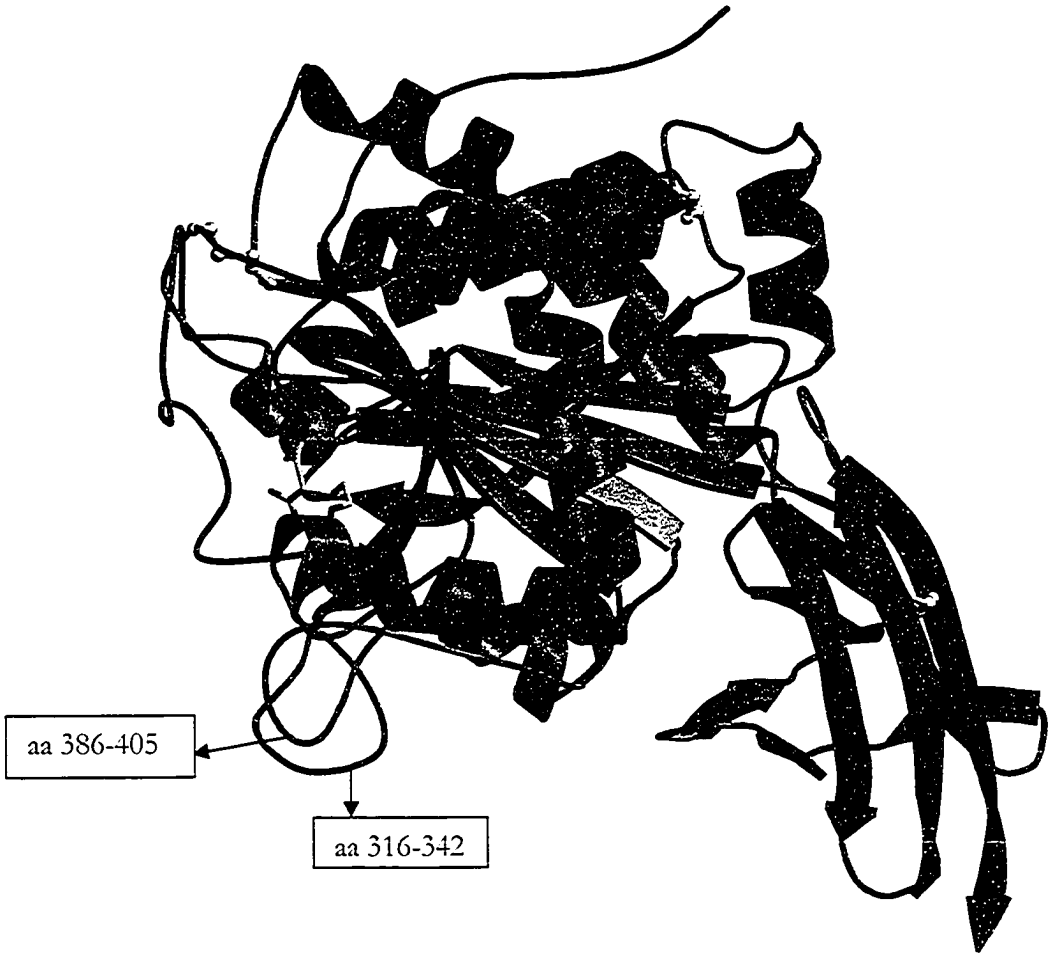
Residues in most favoured regions [A,B,L]	296	89.4%
Residues in additional allowed regions [a,b,l,p]	34	10.3%
Residues in generously allowed regions [~a,~b,~l,~p]	1	0.3%
Residues in disallowed regions	0	0.0%
Number of non-glycine and non-proline residues	331	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues (shown as triangles)	34	
Number of proline residues	17	
Total number of residues	384	

Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and R-factor no greater than 20%, a good quality model would be expected to have over 90% in the most favoured regions.

The construction of these CP domain models, both with and without a peptidomimetic inhibitor, allows us to make many predictions regarding the structure, function and regulation of AEBP1. Structurally speaking we can predict that the CP domain of AEBP1 contains three disulfide bonds based on proximity of cysteine residues to each other, only one of which is present in the CPD structure. These disulfide bonds are Cys192-Cys254, Cys445-Cys488, and Cys578-Cys589 (Figure 12). The first disulfide bond may link together two helices on the surface of the domain, and may serve to restrain the placement of the N-terminal DLD domain with respect to the CP domain. This disulfide bond is also present in the homologous CPE sequence, based on biochemical measurements (*127*). The second predicted disulfide bond, which is also present in the CPD-II structure, links a loop on the surface of the domain with the core of the domain, thus restraining the motion of this loop. The third predicted disulfide bond is located in the C-terminal subdomain. The structure of this subdomain in CPD is maintained by a buried cluster of hydrophobic amino acids running the length of the subdomain (*113*). Interestingly, the final two β-strands of this subdomain lack the majority of these hydrophobic residues in AEBP1. Instead, this region of the subdomain seems to be held together through this third putative disulfide bond functioning in lieu of the hydrophobic cluster.

The putative active site of AEBP1 has several distinctive features. The two loops which were deleted from AEBP1 in order to create the AEBP1 CP model most likely form part of the funnel leading to the putative active site, as indicated in Figure 12. This is also the case in CPE and CPB, which also have large loops in the same place as the first loop of AEBP1, although not as large. The active site itself is lacking many residues critical for CP enzymatic activity in most carboxypeptidases. While two zinc ligands, His236 and Glu239, are conserved in AEBP1, the remaining histidine zinc ligand and glutamate general base are not, but are modified to Asn374 and Tyr485, respectively. The insets at the bottom of Figure 12 illustrate how these residues may coordinate a catalytic zinc ion to produce enzymatic activity.

Figure 12. The CP domain of AEBP1 as modeled by homology against duck carboxypeptidase D. The CP domain of AEBP1 has two subdomains, a larger subdomain composed of a central 8-stranded β-sheet surrounded by 9 α-helices (top panel). Three predicted disulfide bonds are indicated by ball and stick structures, the modeled inhibitor, GEMSA, is indicated by a stick structure, and the putative catalytic zinc ion is indicated by a grey sphere. β-strands are shown in green and α-helices in red. The putative positions of two loops that were not modeled are indicated. The bottom left inset shows the putative active site of AEBP1 as modeled against 1QMU, with lines indicating bonds coordinating the catalytic zinc. The bottom right inset also shows the putative active site, as modeled against 1H8L and thus including the inhibitor, GEMSA. This figure was created with Molscript/Raster3D.

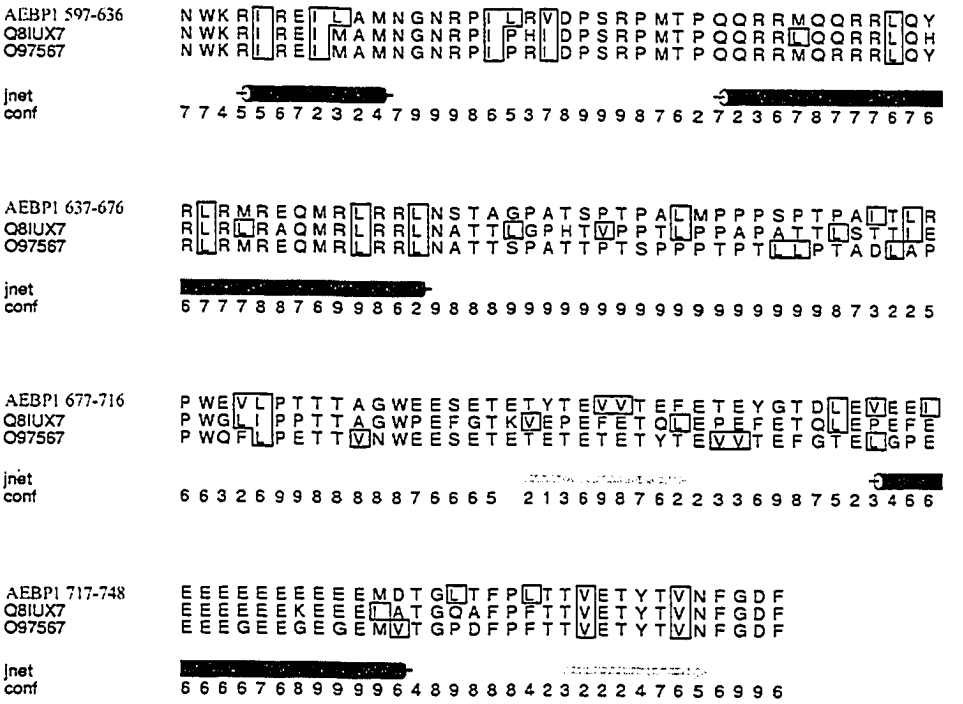

aa 386-405
aa 316-342

3.1.3. Secondary structure of the C-terminus and unmodelled loops within the CP domain of AEBP1.

The C-terminus of AEBP1 consists of a large percentage of repetitive and highly charged sequences. No templates suitable for modeling any of this domain were retrieved from PSI-BLAST searches. However, we can make use of secondary structure predictions to give us an idea as to the conformation of these regions (Figure 13). Based on sequence alone, we can see that there are four distinctive regions within the C-terminal domain. It begins with a highly basic region, rich in arginine, which has been proposed to be the DNA binding domain (*8, 164*). The secondary structure prediction indicates that the basic region is most likely an α-helix. Following this there is a long stretch of unstructured sequence rich in serine, threonine, and proline. Finally there is a long acidic stretch, characterized by a series of glutamate residues, and an uncharged 20-amino acid C-terminal tail. The secondary structure of this region is predicted to be a very acidic α-helix flanked on both ends by β-strands. Because of the highly charged nature of the C-terminus, followed by a relatively hydrophobic C-terminal tail, we might predict that the extreme C-terminus of the protein will not be solvent-exposed, but rather folded inwards. However, it is difficult to make further predictions based on this information, as the relative orientation of the secondary structures to each other is not known.

Two loops within the CP domain of AEBP1 were not modeled by homology due to the lack of a suitable template. Figure 13B shows the secondary structure prediction for this region of the CP domain, performed without the aid of PDB data. Several of the secondary structures predicted here were also predicted through homology modeling with the aid of PDB data (helices T353-E363 and D419-S432), thus adding confidence to this method of structure prediction. Within the first unmodelled loop (amino acids 316-342) a short α-helix is predicted, although with a relatively low confidence value. A larger α-helix is predicted with high confidence to make up the majority of the second unmodelled loop (amino acids 386-405). These loops probably form part of the funnel leading to the active site and thus may play a part in substrate specificity (*127*). They may also play a part in inter-domain interactions.

A.



B.

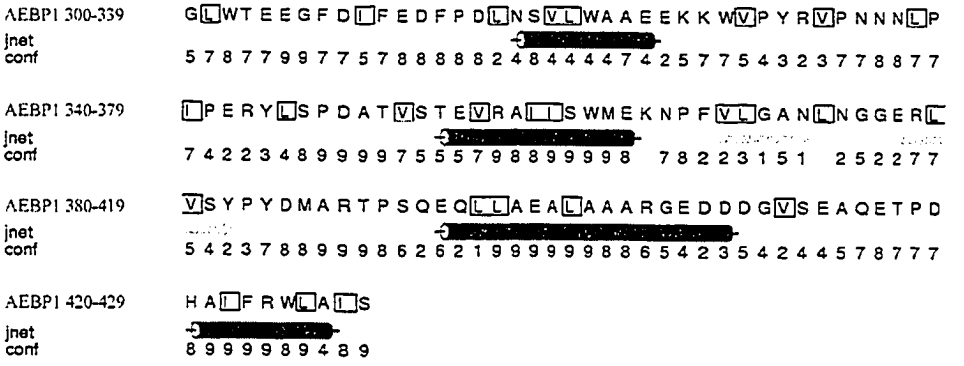
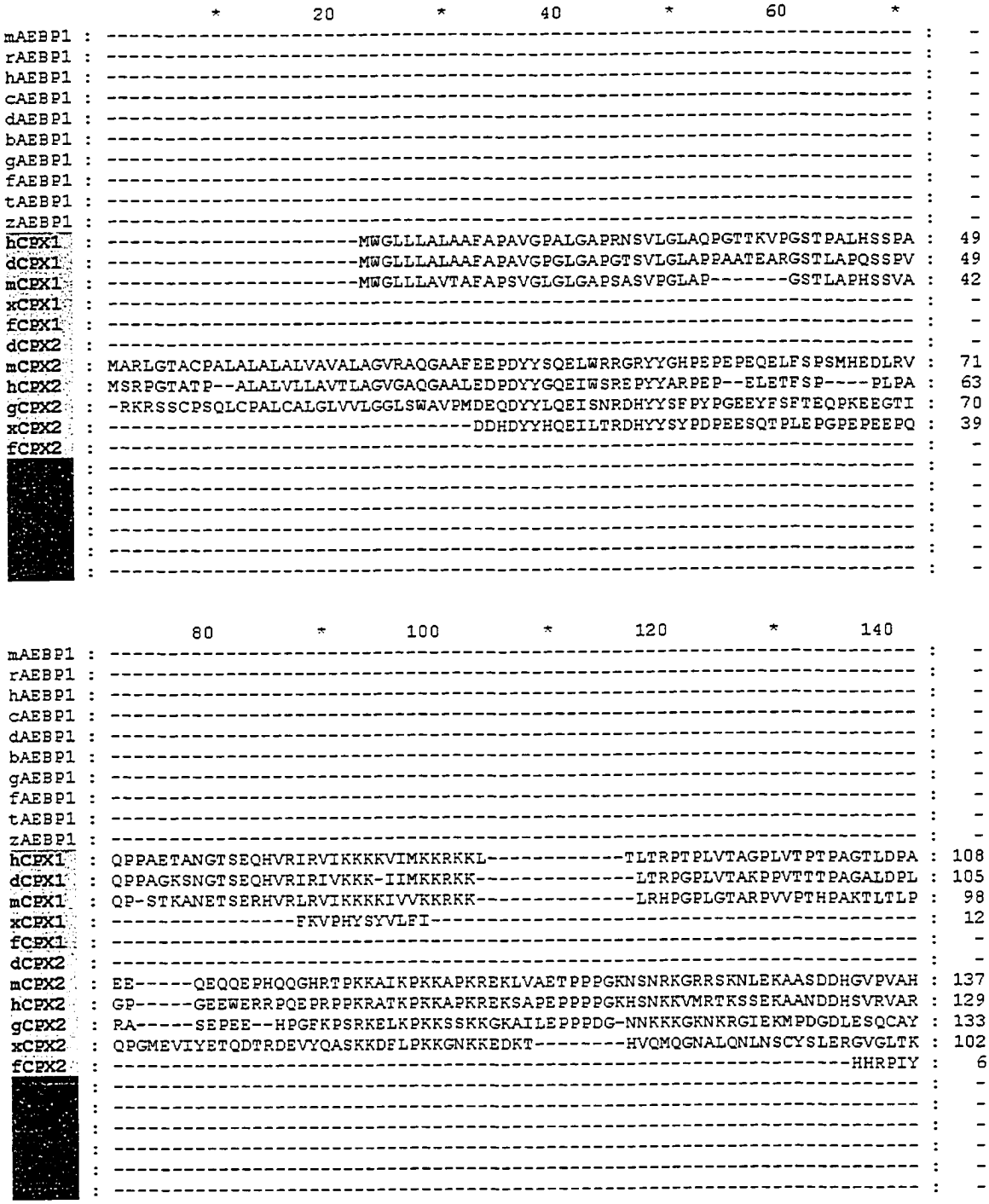


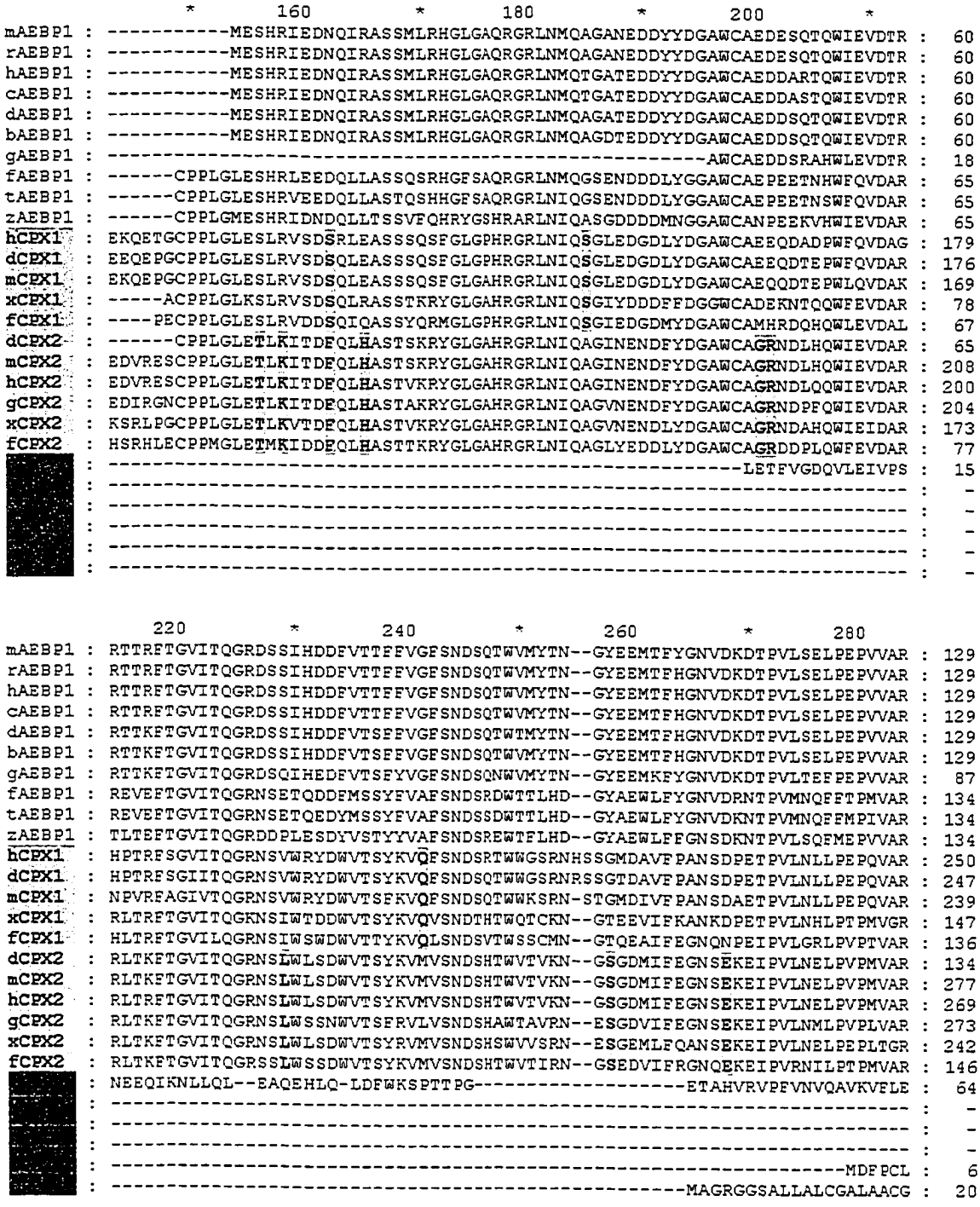
Figure 13. Secondary structure predictions for the C-terminal domain and the two loops within the CP domain of AEBP1. The secondary structure of the C-terminus (A) and a portion of the CP domain (B) of mouse AEBP1 was predicted using Jpred. Jpred is a web server which predicts secondary structure using a PSI-BLAST multiple alignment and a neural network called Jnet. The secondary structure is indicated by the cartoon helices and strands below the alignment, followed by numbering indicating the confidence in this prediction (9 being high confidence and 1 being low confidence).

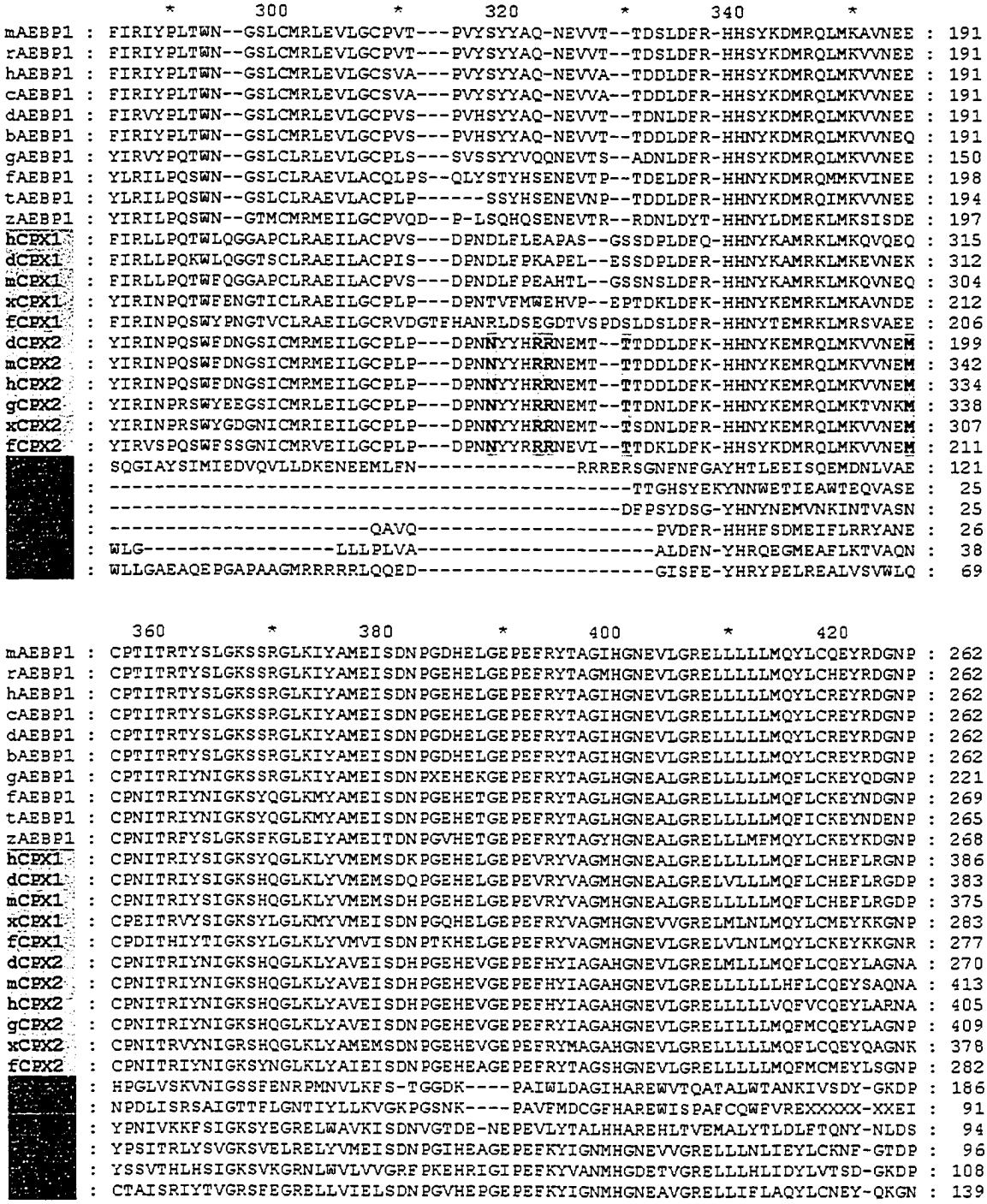
3.1.4. AEBP1 orthologue analysis

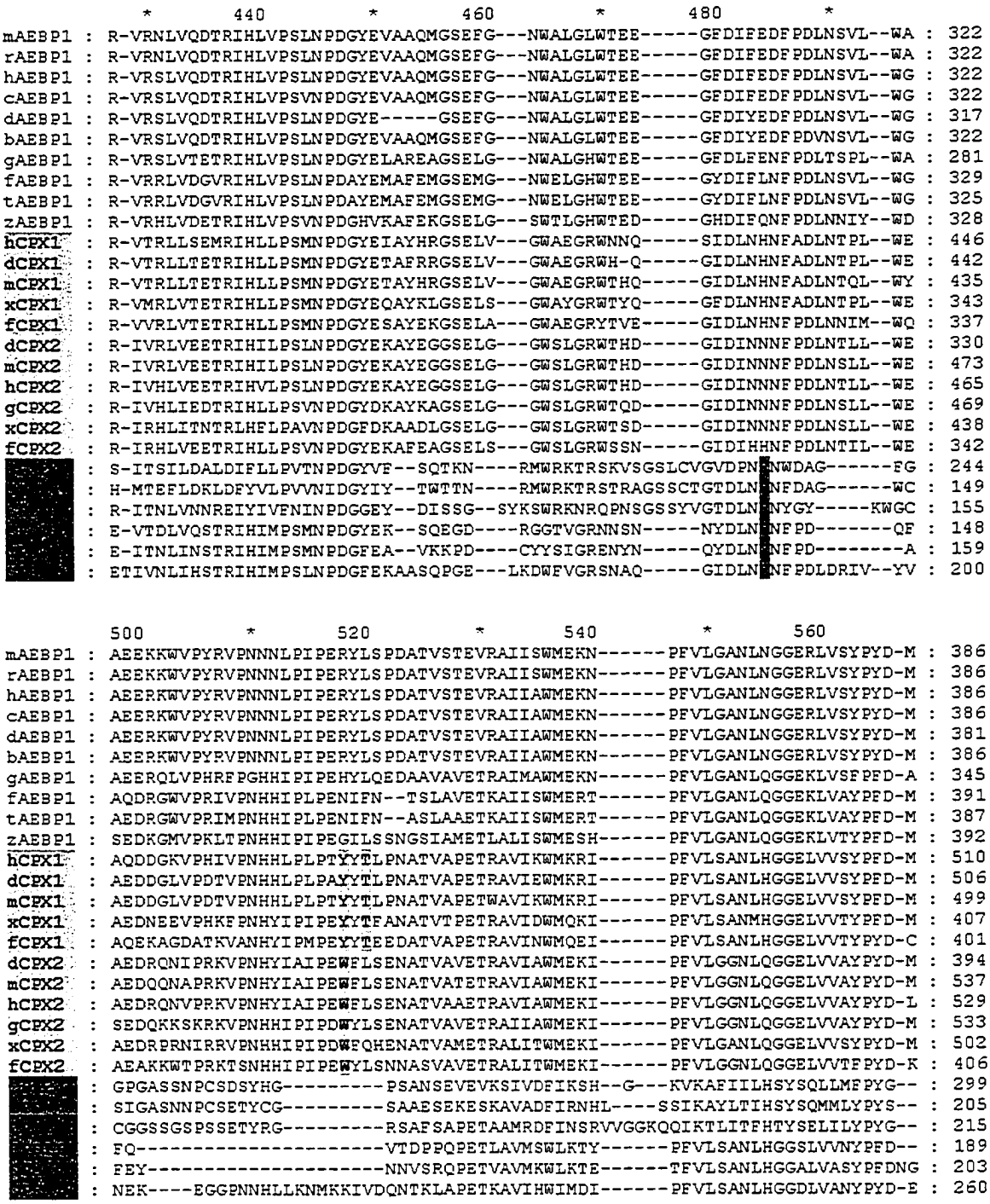
In recent years, many genomes have been sequenced, either completely or partially. This wealth of information has been used to make significant strides in many areas of research. In an attempt to understand what makes AEBP1 both structurally and functionally unique from other carboxypeptidase domain-containing proteins, I searched the available databases for putative orthologues of mouse AEBP1. These searches came up with a number of AEBP1 candidates from vertebrate species. No potential AEBP1 orthologues were identified in non-vertebrate organisms such as *C. elegans*, *S. cerevisiae*, or *D. melanogaster*. Because AEBP1 has two close paralogous relatives, CPX-1 and CPX-2, I wanted to confirm that the putative AEBP1 orthologues I identified were indeed AEBP1 through multiple sequence alignment and phylogenetic analysis. These multiple sequence alignments where then used as a tool to determine which residues of mouse AEBP1 are conserved and hence necessary to maintain the structure and function of AEBP1, while at the same time unique to AEBP1, suggesting a necessity for a unique function of AEBP1. Finally, I used the homology models described in the previous sections to visualize these critical residues in three-dimensional space, allowing predictions to be made as to the function of these residues based on their placement within the entire protein. All protein sequences discussed in this section are shown in a multiple sequence alignment in Figure 14 showing the differences between groups of related carboxypeptidases. A multiple alignment of only AEBP1 orthologues, including additional N-terminal ACLP sequence for mammalian orthologues, is also included in Appendix A illustrating sequence similarities between these orthologues.

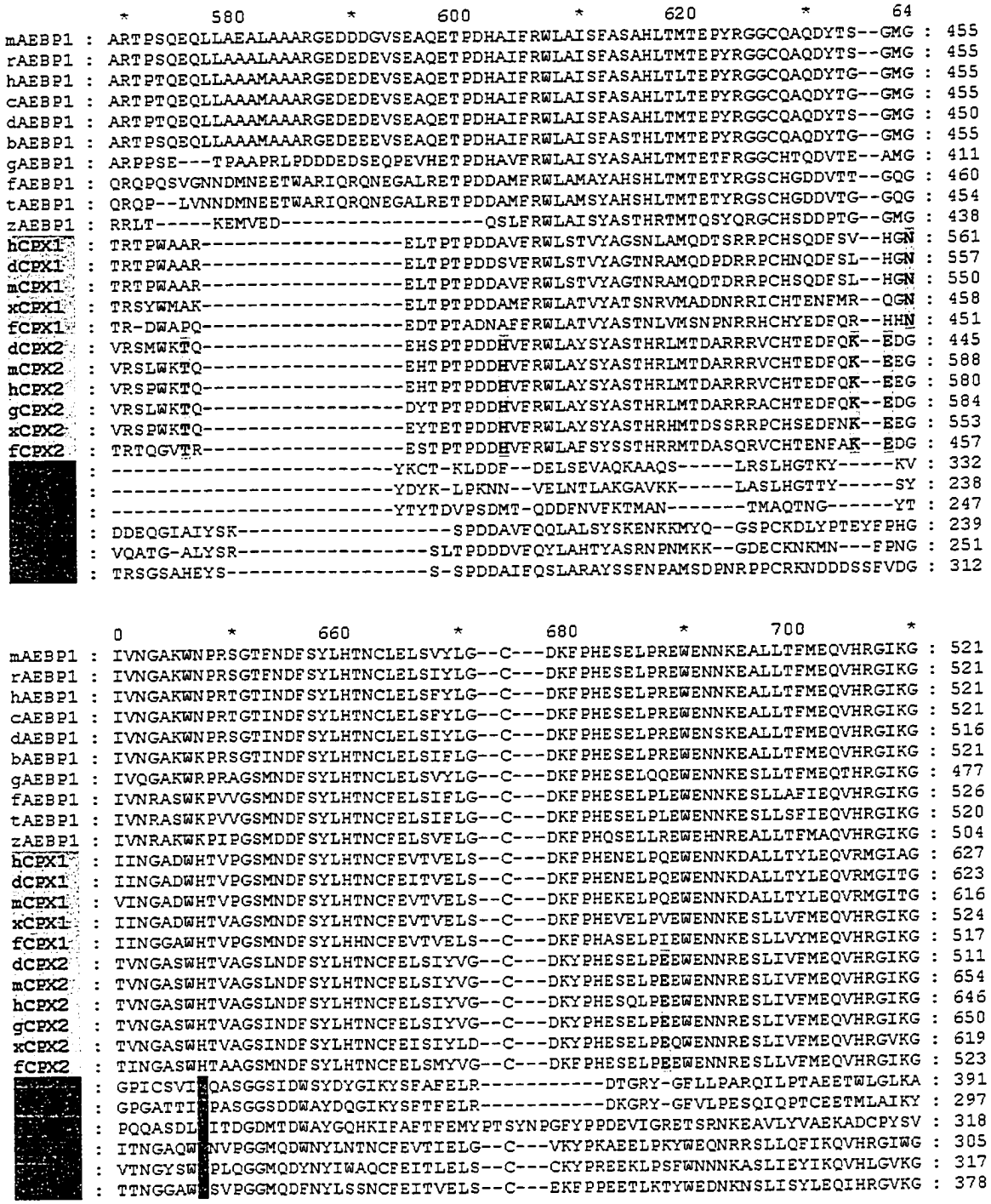
Figure 14. Multiple Alignment of AEBP1 homologues. Orthologues and paralogues of AEBP1 were retrieved from the NCBI and Ensembl databases, aligned using ClustalW, and analyzed using GeneDoc. Shading indicates residues that are conserved and unique to a particular CP group (AEBP1 - yellow, CPX1 - orange, CPX2 - pink, representative active CPs - blue) Abbreviations are as follows: m, mouse (Mus musculus); r, rat (Rattus norvegicus); h, human (Homo sapiens); c, chimp (Pan troglodytes); d, dog (Canis familiaris); b, cow (Bos taurus); g, chicken (Gallus gallus); f, fugu (Takifugu rubripes); t, tetraodon (Tetraodon nigroviridis); z, zebrafish (Danio rerio); x, Xenopus tropicalis. Sequences retrieved from the NCBI databases had the following accession numbers: mAEBP1, CAI25441; hAEBP1, NP_001120; mCPX1, NP_062670; hCPX1, NP_062555; mCPX2, NP_061355; hCPX2, NP_937791. Also found in the NCBI databases was an N-terminally truncated sequence for bAEBP1 (NP_777264), which was made complete with the nucleotide sequence NM_174839, and a C-terminally truncated sequence for cAEBP1 (XM_519550), which was made almost complete by the identification of the C-terminal sequence in the genomic sequence approximately 350 nt 3' from the last identified exon in Ensembl gene ENSPTRG00000019137. This lack of identification of the last exon was commonly found in many of the following predicted transcripts also. The following Ensembl genes were identified as putative orthologues based on reciprocal BLAST analysis: ENSG00000106624 (hAEBP1), ENSMUSG00000020473 (mAEBP1), ENSRNOG00000013720 (rAEBP1), ENSBTAG00000012237 (bAEBP1), ENSRNOG00000013720 (rAEBP1), ENSCAFG00000002917 (dAEBP1), SINFRUG00000151171 (fAEBP1), ENSDARG00000006901 (zAEBP1), GSTENP00025396001 (tAEBP1), ENSCAFG00000006608 (dCPX1), ENSXETG00000012756 (xCPX1), SINFRUG00000148855 (fCPX1), ENSCAFG00000012718 (dCPX2), ENSGALG00000009693 (gCPX2), ENSXETG00000009571 (xCPX2), SINFRUG00000151238 (fCPX2). The sequence for chicken AEBP1 was assembled from many overlapping EST clones obtained from www.chickest.udel.edu (pgf1n.pk006.h24, pgp1n.pk012.m18, pgf2n.pk001.c7, pgm2n.pk007.g13, pgf1n.pk006.o3, pgf1n.pk009.p24) and from www.chick.umist.ac.uk (040024.2). The remaining CPs shown in the alignment are human CPA2, cow CPB, CPT from Thermoactinomyces vulgaris, the second CP domain of duck CPD, human proCPM (CPM), and human proCPE (CPE).

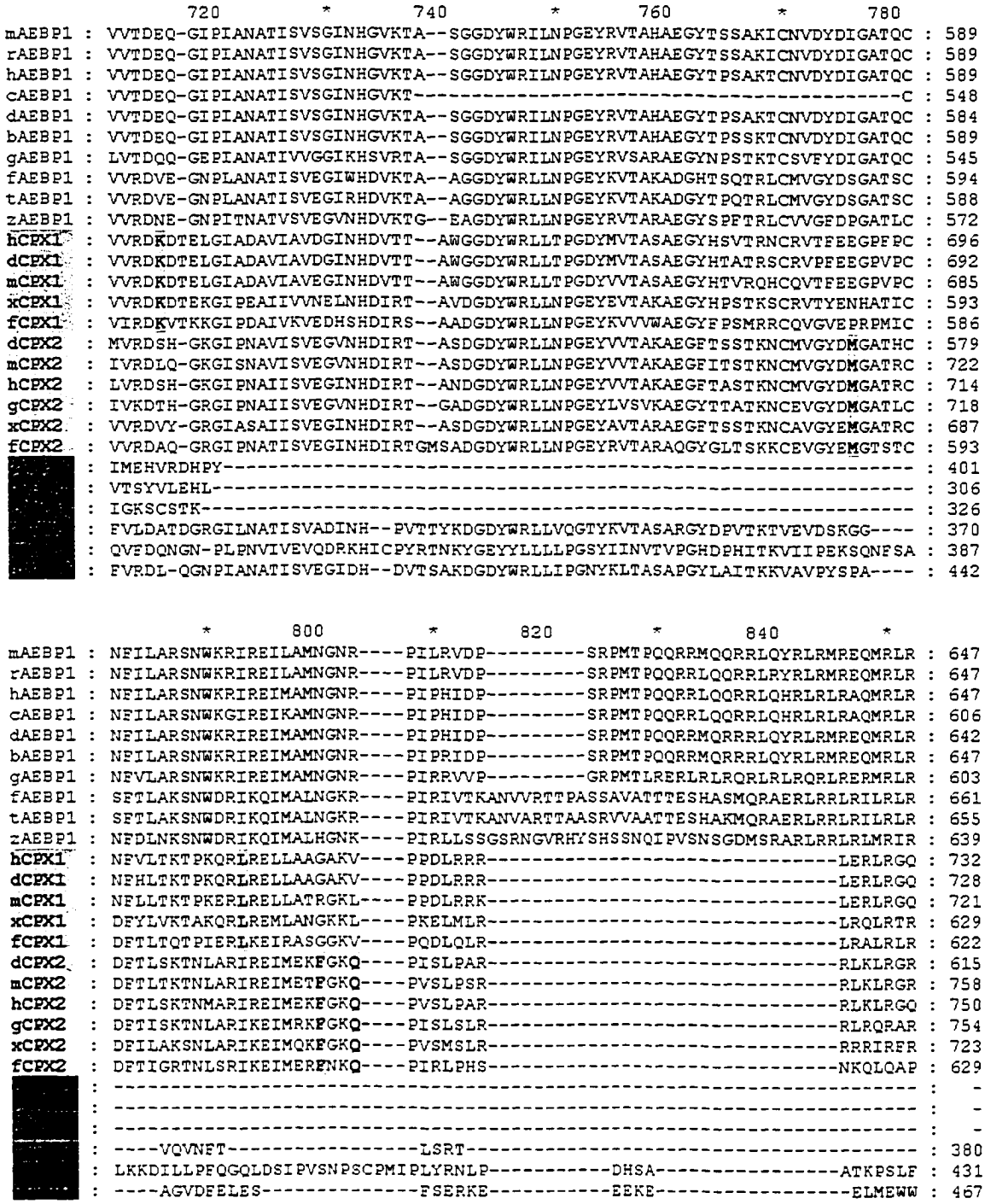

* 20 * 40 * 60 *
mAEBP1 : ---------- : -
rAEBP1 : ---------- : -
hAEBP1 : ---------- : -
cAEBP1 : ---------- : -
dAEBP1 : ---------- : -
bAEBP1 : ---------- : -
gAEBP1 : ---------- : -
fAEBP1 : ---------- : -
tAEBP1 : ---------- : -
zAEBP1 : ---------- : -
hCPX1 : ----------MWGLLLALAAFAPAVGPALGAPRNSVLGLAQPGTTKVPGSTPALHSSPA : 49
dCPX1 : ----------MWGLLLALAAFAPAVGPGLGAPGTSVLGLAPPAATEARGSTLAPQSSPV : 49
mCPX1 : ----------MWGLLLAVTAFAPSVGLGLGAPSASVPGLAP-------GSTLAPHSSVA : 42
xCPX1 : ---------- : -
fCPX1 : ---------- : -
dCPX2 : ---------- : -
mCPX2 : MARLGTACPALALALALVAVALAGVRAQGAAFEEPDYYSQELWRRGRYYGHPEPEPEQELFSPSMHEDLRV : 71
hCPX2 : MSRPGTATP--ALALVLLAVTLAGVGAQGAALEDPDYYGQEIWSREPYYARPEP--ELETFSP----PLPA : 63
gCPX2 : -RKRSSCPSQLCPALCALGLVVLGGLSWAVPMDEQDYYLQEISNRDHYYSFPYPGEEYFSFTEQPKEEGTI : 70
xCPX2 : ----------DDHDYYHQEILTRDHYYSYPDPEESQTPLEPGPEPEEPQ : 39
fCPX2 : ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -
80 * 100 * 120 * 140
mAEBP1 : ---------- : -
rAEBP1 : ---------- : -
hAEBP1 : ---------- : -
cAEBP1 : ---------- : -
dAEBP1 : ---------- : -
bAEBP1 : ---------- : -
gAEBP1 : ---------- : -
fAEBP1 : ---------- : -
tAEBP1 : ---------- : -
zAEBP1 : ---------- : -
hCPX1 : QPPAETANGTSEQHVRIRVIKKKKVIMKKRKKL----------TLTRPTPLVTAGPLVTPTPAGTLDPA : 108
dCPX1 : QPPAGKSNGTSEQHVRIRIVKKK-IIMKKRKK----------LTRPGPLVTAKPPVTTTPAGALDPL : 105
mCPX1 : QP-STKANETSERHVRLRVIKKKKIVVKKRKK----------LRHPGPLGTARPVVPTHPAKTLTLP : 98
xCPX1 : ----------FKVPHYSYVLFI---------- : 12
fCPX1 : ---------- : -
dCPX2 : ---------- : -
mCPX2 : EE-----QEQQEPHQQGHRTPKKAIKPKKAPKREKLVAETPPPGKNSNRKGRRSKNLEKAASDDHGVPVAH : 137
hCPX2 : GP-----GEEWERRPQEPRPPKRATKPKKAPKREKSAPEPPPPGKHSNKKVMRTKSSEKAANDDHSVRVAR : 129
gCPX2 : RA-----SEPEE--HPGFKPSRKELKPKKSSKKGKAILEPPPDG-NNKKKGKNKRGIEKMPDGDLESQCAY : 133
xCPX2 : QPGMEVIYETQDTRDEVYQASKKDFLPKKGNKKEDKT--------HVQMQGNALQNLNSCYSLERGVGLTK : 102
fCPX2 : ----------HHRPIY : 6
: ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -

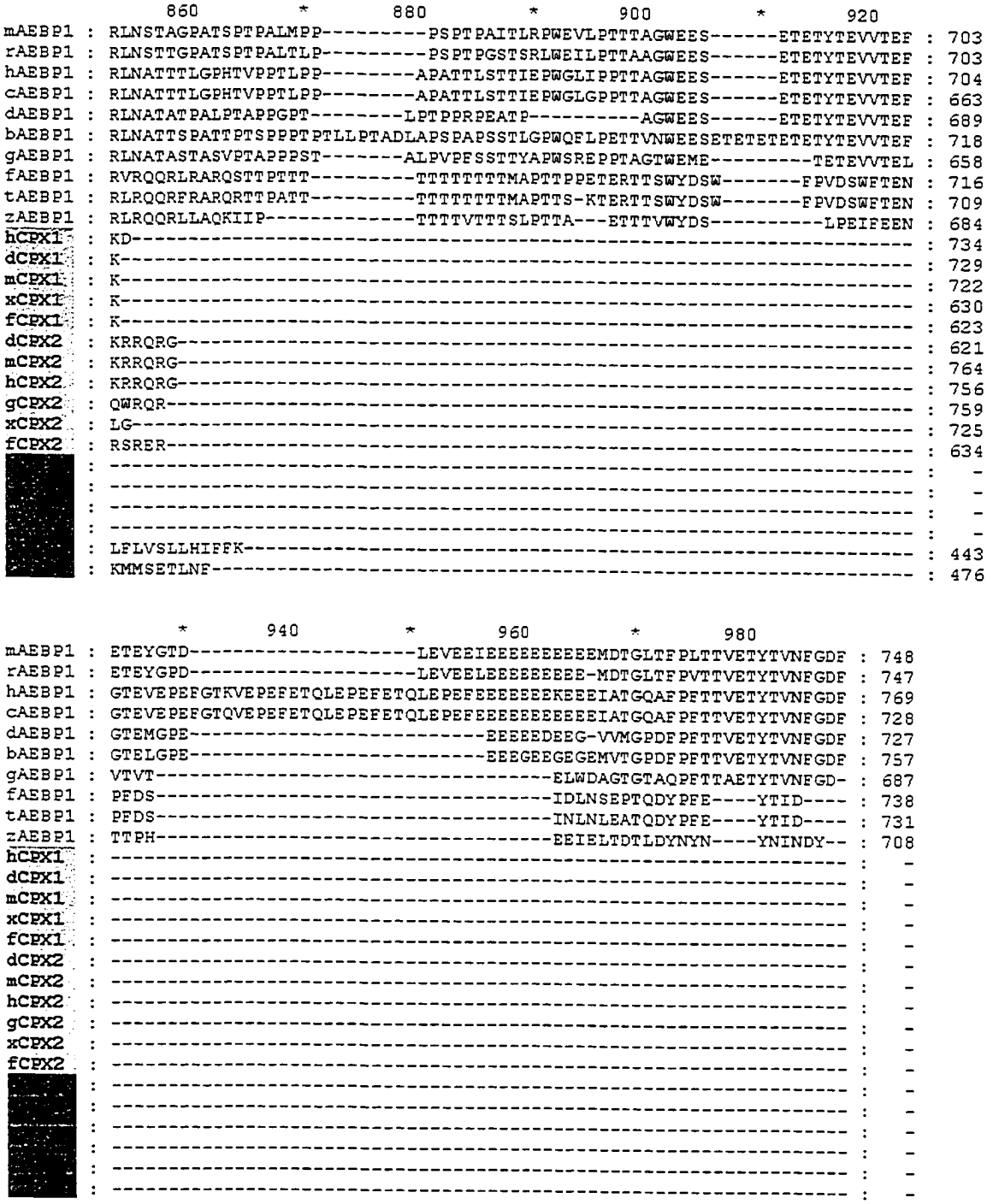

mAEBP1
rAEBP1
hAEBP1
cAEBP1
dAEBP1
bAEBP1
gAEBP1
fAEBP1
tAEBP1
zAEBP1
hCPX1
dCPX1
mCPX1
xCPX1
fCPX1
dCPX2
mCPX2
hCPX2
gCPX2
xCPX2
fCPX2


* 300 * 320 * 340 *
mAEBP1 : FIRIYPLTWN--GSLCMRLEVLGCPVT---PVYSYYAQ-NEVVT--TDSLDFR-HHSYKDMRQLMKAVNEE : 191
rAEBP1 : FIRIYPLTWN--GSLCMRLEVLGCPVT---PVYSYYAQ-NEVVT--TDSLDFR-HHSYKDMRQLMKVVNEE : 191
hAEBP1 : FIRIYPLTWN--GSLCMRLEVLGCSVA---PVYSYYAQ-NEVVA--TDDLDFR-HHSYKDMRQLMKVVNEE : 191
cAEBP1 : FIRIYPLTWN--GSLCMRLEVLGCSVA---PVYSYYAQ-NEVVA--TDDLDFR-HHSYKDMRQLMKVVNEE : 191
dAEBP1 : FIRVYPLTWN--GSLCMRLEVLGCPVS---PVHSYYAQ-NEVVT--TDNLDFR-HHSYKDMRQLMKVVNEE : 191
bAEBP1 : FIRIYPLTWN--GSLCMRLEVLGCPVS---PVHSYYAQ-NEVVT--TDDLDFR-HHNYKDMRQLMKVVNEQ : 191
gAEBP1 : YIRVYPQTWN--GSLCLRLEVLGCPLS---SVSSYYVQQNEVTS--ADNLDFR-HHSYKDMRQLMKVVNEE : 150
fAEBP1 : YLRILPQSWN--GSLCLRAEVLACQLPS--QLYSTYHSENEVTP--TDELDFR-HHNYKDMRQMMKVINEE : 198
tAEBP1 : YLRILPQSWN--GSLCLRAEVLACPLP------SSYHSENEVNP--TDDLDFR-HHNYKDMRQIMKVVNEE : 194
zAEBP1 : YIRILPQSWN--GTMCMRMEILGCPVQD--P-LSQHQSENEVTR--RDNLDYT-HHNYLDMEKLMKSISDE : 197
hCPX1 : FIRLLPQTWLQGGAPCLRAEILACPVS---DPNDLFLEAPAS--GSSDPLDFQ-HHNYKAMRKLMKQVQEQ : 315
dCPX1 : FIRLLPQKWLQGGTSCLRAEILACPIS---DPNDLFPKAPEL--ESSDPLDFR-HHNYKAMRKLMKEVNEK : 312
mCPX1 : FIRLLPQTWFQGGAPCLRAEILACPVS---DPNDLFPEAHTL--GSSNSLDFR-HHNYKAMRKLMKQVNEQ : 304
xCPX1 : YIRINPQTWFENGTICLRAEILGCPLP---DPNTVFMWEHVP--EPTDKLDFK-HHNYKEMRKLMKAVNDE : 212
fCPX1 : FIRINPQSWYPNGTVCLRAEILGCRVDGTFHANRLDSEGDTVSPDSLDSLDFR-HHNYTEMRKLMRSVAEE : 206
dCPX2 : YIRINPQSWFDNGSICMRMEILGCPLP---DPNNYYHRRNEMT--TTDDLDFK-HHNYKEMRQLMKVVNEM : 199
mCPX2 : YIRINPQSWFDNGSICMRMEILGCPLP---DPNNYYHRRNEMT--TTDDLDFK-HHNYKEMRQLMKVVNEM : 342
hCPX2 : YIRINPQSWFDNGSICMRMEILGCPLP---DPNNYYHRRNEMT--TTDDLDFK-HHNYKEMRQLMKVVNEM : 334
gCPX2 : YIRINPRSWYEEGSICMRLEILGCPLP---DPNNYYHRRNEMT--TTDNLDFK-HHNYKEMRQLMKTVNKM : 338
xCPX2 : YIRINPRSWYGDGNICMRIEILGCPLP---DPNNYYHRRNEMT--TSDNLDFR-HHNYKEMRQLMKVVNEM : 307
fCPX2 : YIRVSPQSWFSSGNICMRVEILGCPLP---DPNNYYRRRNEVI--TTDKLDFK-HHSYKDMRQLMKVVNEM : 211
[redacted] : SQGIAYSIMIEDVQVLLDKENEEMLFN--------------RRRERSGNFNFGAYHTLEEISQEMDNLVAE : 121
[redacted] : ----------TTGHSYEKYNNWETIEAWTEQVASE : 25
[redacted] : ----------DFPSYDSG-YHNYNEMVNKINTVASN : 25
[redacted] : ----------QAVQ----------PVDFR-HHHFSDMEIFLRRYANE : 26
[redacted] : WLG----------LLLPLVA----------ALDFN-YHRQEGMEAFLKTVAQN : 38
[redacted] : WLLGAEAQEPGAPAAGMRRRRRLQQED----------GISFE-YHRYPELREALVSVWLQ : 69
360 * 380 * 400 * 420
mAEBP1 : CPTITRTYSLGKSSRGLKIYAMEISDNPGDHELGEPEFRYTAGIHGNEVLGRELLLLLMQYLCQEYRDGNP : 262
rAEBP1 : CPTITRTYSLGKSSRGLKIYAMEISDNPGEHELGEPEFRYTAGMHGNEVLGRELLLLLMQYLCHEYRDGNP : 262
hAEBP1 : CPTITRTYSLGKSSRGLKIYAMEISDNPGEHELGEPEFRYTAGIHGNEVLGRELLLLLMQYLCREYRDGNP : 262
cAEBP1 : CPTITRTYSLGKSSRGLKIYAMEISDNPGEHELGEPEFRYTAGIHGNEVLGRELLLLLMQYLCREYRDGNP : 262
dAEBP1 : CPTITRTYSLGKSSRGLKIYAMEISDNPGDHELGEPEFRYTAGIHGNEVLGRELLLLLMQYLCREYRDGNP : 262
bAEBP1 : CPTITRTYSLGKSSRGLKIYAMEISDNPGDHELGEPEFRYTAGIHGNEVLGRELLLLLMQYLCREYRDGNP : 262
gAEBP1 : CPTITRIYNIGKSSRGLKIYAMEISDNPXEHEKGEPEFRYTAGLHGNEALGRELLLLLMQFLCKEYQDGNP : 221
fAEBP1 : CPNITRIYNIGKSYQGLKMYAMEISDNPGEHETGEPEFRYTAGLHGNEALGRELLLLLMQFLCKEYNDGNP : 269
tAEBP1 : CPNITRIYNIGKSYQGLKMYAMEISDNPGEHETGEPEFRYTAGLHGNEALGRELLLLLMQFICKEYNDENP : 265
zAEBP1 : CPNITRFYSLGKSFKGLEIYAMEITDNPGVHETGEPEFRYTAGYHGNEALGRELLLMFMQYLCKEYKDGNP : 268
hCPX1 : CPNITRIYSIGKSYQGLKLYVMEMSDKPGEHELGEPEVRYVAGMHGNEALGRELLLLLMQFLCHEFLRGNP : 386
dCPX1 : CPNITRIYSIGKSHQGLKLYVMEMSDQPGEHELGEPEVRYVAGMHGNEALGRELVLLLMQFLCHEFLRGDP : 383
mCPX1 : CPNITRIYSIGKSHQGLKLYVMEMSDHPGEHELGEPEVRYVAGMHGNEALGRELLLLLMQFLCHEFLRGDP : 375
xCPX1 : CPEITRVYSIGKSYLGLKMYVMEISDNPGQHELGEPEFRYVAGMHGNEVVGRELMLNLMQYLCMEYKKGNP : 283
fCPX1 : CPDITHIYTIGKSYLGLKLYVMVISDNPTKHELGEPEFRYVAGMHGNEVLGRELVLNLMQYLCKEYKKGNR : 277
dCPX2 : CPNITRIYNIGKSHQGLKLYAVEISDHPGEHEVGEPEFHYIAGAHGNEVLGRELMLLLMQFLCQEYLAGNA : 270
mCPX2 : CPNITRIYNIGKSHQGLKLYAVEISDHPGEHEVGEPEFHYIAGAHGNEVLGRELLLLLLHFLCQEYSAQNA : 413
hCPX2 : CPNITRIYNIGKSHQGLKLYAVEISDHPGEHEVGEPEFHYIAGAHGNEVLGRELLLLLVQFVCQEYLARNA : 405
gCPX2 : CPNITRIYNIGKSHQGLKLYAVEISDNPGEHEVGEPEFRYIAGAHGNEVLGRELILLLMQFMCQEYLAGNP : 409
xCPX2 : CPNITRVYNIGRSHQGLKLYAMEMSDNPGEHEVGEPEFRYMAGAHGNEVLGRELLLLLMQFLCQEYQAGNK : 378
fCPX2 : CPNITRIYNIGKSYNGLKLYAIEISDNPGEHEAGEPEFRYTAGSHGNEVLGRELLLLLMQFMCMEYLSGNP : 282
[redacted] : HPGLVSKVNIGSSFENRPMNVLKFS-TGGDK----PAIWLDAGIHAREWVTQATALWTANKIVSDY-GKDP : 186
[redacted] : NPDLISRSAIGTTFLGNTIYLLKVGKPGSNK----PAVFMDCGFHAREWISPAFCQWFVREXXXXX-XXEI : 91
[redacted] : YPNIVKKFSIGKSYEGRELWAVKISDNVGTDE-NEPEVLYTALHHAREHLTVEMALYTLDLFTQNY-NLDS : 94
[redacted] : YPSITRLYSVGKSVELRELYVMEISDNPGIHEAGEPEFKYIGNMHGNEVVGRELLLNLIEYLCKNF-GTDP : 96
[redacted] : YSSVTHLHSIGKSVKGRNLWVLVVGRFPKEHRIGIPEFKYVANMHGDETVGRELLLHLIDYLVTSD-GKDP : 108
[redacted] : CTAISRIYTVGRSFEGRELLVIELSDNPGVHEPGEPEFKYIGNMHGNEAVGRELLIFLAQYLCNEY-QKGN : 139


* 440 * 460 * 480 *
mAEBP1 : R-VRNLVQDTRIHLVPSLNPDGYEVAAQMGSEFG---NWALGLWTEE-----GFDIFEDFPDLNSVL--WA : 322
rAEBP1 : R-VRNLVQDTRIHLVPSLNPDGYEVAAQMGSEFG---NWALGLWTEE-----GFDIFEDFPDLNSVL--WA : 322
hAEBP1 : R-VRSLVQDTRIHLVPSLNPDGYEVAAQMGSEFG---NWALGLWTEE-----GFDIFEDFPDLNSVL--WG : 322
cAEBP1 : R-VRSLVQDTRIHLVPSVNPDGYEVAAQMGSEFG---NWALGLWTEE-----GFDIFEDFPDLNSVL--WG : 322
dAEBP1 : R-VRSLVQDTRIHLVPSLNPDGYE-----GSEFG---NWALGLWTEE-----GFDIYEDFPDLNSVL--WG : 317
bAEBP1 : R-VRSLVQDTRIHLVPSLNPDGYEVAAQMGSEFG---NWALGLWTEE-----GFDIYEDFPDVNSVL--WG : 322
gAEBP1 : R-VRSLVTETRIHLVPSLNPDGYELAREAGSELG---NWALGHWTEE-----GFDLFENFPDLTSPL--WA : 281
fAEBP1 : R-VRRLVDGVRIHLVPSLNPDAYEMAFEMGSEMG---NWELGHWTEE-----GYDIFLNFPDLNSVL--WG : 329
tAEBP1 : R-VRRLVDGVRIHLVPSLNPDAYEMAFEMGSEMG---NWELGHWTEE-----GYDIFLNFPDLNSVL--WG : 325
zAEBP1 : R-VRHLVDETRIHLVPSVNPDGHVKAFEKGSELG---SWTLGHWTED-----GHDIFQNFPDLNNIY--WD : 328
hCPX1 : R-VTRLLSEMRIHLLPSMNPDGYEIAYHRGSELV---GWAEGRWNNQ-----SIDLNHNFADLNTPL--WE : 446
dCPX1 : R-VTRLLTETRIHLLPSMNPDGYETAFRRGSELV---GWAEGRWH-Q-----GIDLNHNFADLNTPL--WE : 442
mCPX1 : R-VTRLLTETRIHLLPSMNPDGYETAYHRGSELV---GWAEGRWTHQ-----GIDLNHNFADLNTQL--WY : 435
xCPX1 : R-VMRLVTETRIHLLPSMNPDGYEQAYKLGSELS---GWAYGRWTYQ-----GFDLNHNFADLNTPL--WE : 343
fCPX1 : R-VVRLVTETRIHLLPSMNPDGYESAYEKGSELA---GWAEGRYTVE-----GIDLNHNFPDLNNIM--WQ : 337
dCPX2 : R-IVRLVEETRIHILPSLNPDGYEKAYEGGSELG---GWSLGRWTHD-----GIDINNNFPDLNTLL--WE : 330
mCPX2 : R-IVRLVEETRIHILPSLNPDGYEKAYEGGSELG---GWSLGRWTHD-----GIDINNNFPDLNSLL--WE : 473
hCPX2 : R-IVHLVEETRIHVLPSLNPDGYEKAYEGGSELG---GWSLGRWTHD-----GIDINNNFPDLNSLL--WE : 465
gCPX2 : R-IVHLIEDTRIHLLPSVNPDGYDKAYKAGSELG---GWSLGRWTQD-----GIDINNNFPDLNSLL--WE : 469
xCPX2 : R-IRHLITNTRLHFLPAVNPDGFDKAADLGSELG---GWSLGRWTSD-----GIDINNNFPDLNSLL--WE : 438
fCPX2 : R-IRHLVEETRIHLLPSVNPDGYEKAFEAGSELS---GWSLGRWSSN-----GIDIHHNFPDLNTIL--WE : 342
[redacted] : S-ITSILDALDIFLLPVTNPDGYVF--SQTKN----RMWRKTRSKVSGSLCVGVDPNRNWDAG------FG : 244
[redacted] : H-MTEFLDKLDFYVLPVVNIDGYIY--TWTTN----RMWRKTRSTRAGSSCTGTDLNRNFDAG------WC : 149
[redacted] : R-ITNLVNNREIYIVPNINPDGGEY--DISSG--SYKSWRKNRQPNSGSSYVGTDLNRNYGY-----KWGC : 155
[redacted] : E-VTDLVQSTRIHIMPSMNPDGYEK--SQEGD----RGGTVGRNNSN-----NYDLNRNFPD-------QF : 148
[redacted] : E-ITNLINSTRIHIMPSMNPDGFEA--VKKPD----CYYSIGRENYN-----QYDLNRNFPD--------A : 159
[redacted] : ETIVNLIHSTRIHIMPSLNPDGFEKAASQPGE---LKDWFVGRSNAQ-----GIDLNRNFPDLDRIV--YV : 200
500 * 520 * 540 * 560
mAEBP1 : AEEKKWVPYRVPNNNLPIPERYLSPDATVSTEVRAIISWMEKN------PFVLGANLNGGERLVSYPYD-M : 386
rAEBP1 : AEEKKWVPYRVPNNNLPIPERYLSPDATVSTEVRAIISWMEKN------PFVLGANLNGGERLVSYPYD-M : 386
hAEBP1 : AEERKWVPYRVPNNNLPIPERYLSPDATVSTEVRAIIAWMEKN------PFVLGANLNGGERLVSYPYD-M : 386
cAEBP1 : AEERKWVPYRVPNNNLPIPERYLSPDATVSTEVRAIIAWMEKN------PFVLGANLNGGERLVSYPYD-M : 386
dAEBP1 : AEERKWVPYRVPNNNLPIPERYLSPDATVSTEVRAIIAWMEKN------PFVLGANLNGGERLVSYPYD-M : 381
bAEBP1 : AEERKWVPYRVPNNNLPIPERYLSPDATVSTEVRAIIAWMEKN------PFVLGANLNGGERLVSYPYD-M : 386
gAEBP1 : AEERQLVPHRFPGHHIPIPEHYLQEDAAVAVETRAIMAWMEKN------PFVLGANLQGGEKLVSFPFD-A : 345
fAEBP1 : AQDRGWVPRIVPNHHIPLPENIFN--TSLAVETKAIISWMERT------PFVLGANLQGGEKLVAYPFD-M : 391
tAEBP1 : AEDRGWVPRIMPNHHIPLPENIFN--ASLAAETKAIISWMERT------PFVLGANLQGGEKLVAYPFD-M : 387
zAEBP1 : SEDKGMVPKLTPNHHIPIPEGILSSNGSIAMETLALISWMESH------PFVLGANLQGGEKLVTYPFD-M : 392
hCPX1 : AQDDGKVPHIVPNHHLPLPTYYTLPNATVAPETRAVIKWMKRI------PFVLSANLHGGELVVSYPFD-M : 510
dCPX1 : AEDDGLVPDTVPNHHLPLPAYYTLPNATVAPETRAVIEWMKRI------PFVLSANLHGGELVVSYPFD-M : 506
mCPX1 : AEDDGLVPDTVPNHHLPLPTYYTLPNATVAPETWAVIKWMKRI------PFVLSANLHGGELVVSYPFD-M : 499
xCPX1 : AEDNEEVPHKFPNHYIPIPEYYTFANATVTPETRAVIDWMQKI------PFVLSANMHGGELVVTYPFD-M : 407
fCPX1 : AQEKAGDATKVANHYIPMPEYYTEEDATVAPETRAVINWMQEI------PFVLSANLHGGELVVTYPYD-C : 401
dCPX2 : AEDRQNIPRKVPNHYIAIPEWFLSENATVAVETRAVIAWMEKI------PFVLGGNLQGGELVVAYPYD-M : 394
mCPX2 : AEDQQNAPRKVPNHYIAIPEWFLSENATVATETRAVIAWMEKI------PFVLGGNLQGGELVVAYPYD-M : 537
hCPX2 : AEDRQNVPRKVPNHYIAIPEWFLSENATVAAETRAVIAWMEKI------PFVLGGNLQGGELVVAYPYD-L : 529
gCPX2 : SEDQKKSKRKVPNHHIPIPDWYLSENATVAVETRAIIAWMEKI------PFVLGGNLQGGELVVAYPYD-M : 533
xCPX2 : AEDRPRNIRRVPNHHIPIPDWFQHENATVAMETRALITWMEKI------PFVLGANLQGGELVVSYPYD-M : 502
fCPX2 : AEAKKWTPRKTSNHHIPIPEWYLSNNASVAVETRALITWMEKI------PFVLGGNLQGGELVVTFPYD-K : 406
[redacted] : GPGASSNPCSDSYHG----------PSANSEVEVKSIVDFIKSH--G---KVKAFIILHSYSQLLMFPYG-- : 299
[redacted] : SIGASNNPCSETYCG----------SAAESEKESKAVADFIRNHL----SSIKAYLTIHSYSQMMLYPYS-- : 205
[redacted] : CGGSSGSPSSETYRG----------RSAFSAPETAAMRDFINSRVVGGKQQIKTLITFHTYSELILYPYG-- : 215
[redacted] : FQ----------VTDPPQPETLAVMSWLKTY------PFVLSANLHGGSLVVNYPFD-- : 189
[redacted] : FEY----------NNVSRQPETVAVMKWLKTE------TFVLSANLHGGALVASYPFDNG : 203
[redacted] : NEK----EGGPNNHLLKNMKKIVDQNTKLAPETKAVIHWIMDI------PFVLSANLHGGDLVANYPYD-E : 260


mAEBP1 : ARTPSQEQLLAEALAAARGEDDDGVSEAQETPDHAIFRWLAISFASAHLTMTEPYRGGCQAQDYTS--GMG : 455
rAEBP1 : ARTPSQEQLLAAALAAARGEDEDEVSEAQETPDHAIFRWLAISFASAHLTMTEPYRGGCQAQDYTS--GMG : 455
hAEBP1 : ARTPTQEQLLAAAMAAARGEDEDEVSEAQETPDHAIFRWLAISFASAHLTLTEPYRGGCQAQDYTG--GMG : 455
cAEBP1 : ARTPTQEQLLAAAMAAARGEDEDEVSEAQETPDHAIFRWLAISFASAHLTLTEPYRGGCQAQDYTG--GMG : 455
dAEBP1 : ARTPTQEQLLAAAMAAARGEDEDEVSEAQETPDHAIFRWLAISFASAHLTMTEPYRGGCQAQDYTS--GMG : 450
bAEBP1 : ARTPSQEQLLAAAMAAARGEDEEEVSEAQETPDHAIFRWLAISFASTHLTMTEPYRGGCQAQDYTG--GMG : 455
gAEBP1 : ARPPSE---TPAAPRLPDDDEDSEQPEVHETPDHAVFRWLAISYASAHLTMTETFRGGCHTQDVTE--AMG : 411
fAEBP1 : QRQPQSVGNNDMNEETWARIQRQNEGALRETPDDAMFRWLAMAYAHSHLTMTETYRGSCHGDDVTT--GQG : 460
tAEBP1 : QRQP--LVNNDMNEETWARIQRQNEGALRETPDDAMFRWLAMSYAHSHLTMTETYRGSCHGDDVTG--GQG : 454
zAEBP1 : RRLT-----KEMVED----------QSLFRWLAISYASTHRTMTQSYQRGCHSDDPTG--GMG : 438
hCPX1 : TRTPWAAR----------ELTPTPDDAVFRWLSTVYAGSNLAMQDTSRRPCHSQDFSV--HGN : 561
dCPX1 : TRTPWAAR----------ELTPTPDDSVFRWLSTVYAGTNRAMQDPDRRPCHNQDFSL--HGN : 557
mCPX1 : TRTPWAAR----------ELTPTPDDAVFRWLSTVYAGTNRAMQDTDRRPCHSQDFSL--HGN : 550
xCPX1 : TRSYWMAK----------ELTPTPDDAMFRWLATVYATSNRVMADDNRRICHTENFMR--QGN : 458
fCPX1 : TR-DWAPQ----------EDTPTADNAFFRWLATVYASTNLVMSNPNRRHCHYEDFQR--HHN : 451
dCPX2 : VRSMWKTQ----------EHSPTPDDHVFRWLAYSYASTHRLMTDARRRVCHTEDFQK--EDG : 445
mCPX2 : VRSLWKTQ----------EHTPTPDDHVFRWLAYSYASTHRLMTDARRRVCHTEDFQK--EEG : 588
hCPX2 : VRSPWKTQ----------EHTPTPDDHVFRWLAYSYASTHRLMTDARRRVCHTEDFQK--EEG : 580
gCPX2 : VRSLWKTQ----------DYTPTPDDHVFRWLAYSYASTHRLMTDARRRACHTEDFQK--EDG : 584
xCPX2 : VRSPWKTQ----------EYTETPDDHVFRWLAYSYASTHRHMTDSSRRPCHSEDFNK--EEG : 553
fCPX2 : TRTQGVTR----------ESTPTPDDHVFRWLAFSYSSTHRRMTDASQRVCHTENFAK--EDG : 457
[redacted] : ----------YKCT-KLDDF--DELSEVAQKAAQS-----LRSLHGTKY-----KV : 332
[redacted] : ----------YDYK-LPKNN--VELNTLAKGAVKK-----LASLHGTTY-----SY : 238
[redacted] : ----------YTYTDVPSDMT-QDDFNVFKTMAN-------TMAQTNG------YT : 247
[redacted] : DDEQGIAIYSK----------SPDDAVFQQLALSYSKENKKMYQ--GSPCKDLYPTEYFPHG : 239
[redacted] : VQATG-ALYSR----------SLTPDDDVFQYLAHTYASRNPNMKK--GDECKNKMN---FPNG : 251
[redacted] : TRSGSAHEYS----------S-SPDDAIFQSLARAYSSFNPAMSDPNRPPCRKNDDDSSFVDG : 312
mAEBP1 : IVNGAKWNPRSGTFNDFSYLHTNCLELSVYLG--C---DKFPHESELPREWENNKEALLTFMEQVHRGIKG : 521
rAEBP1 : IVNGAKWNPRSGTFNDFSYLHTNCLELSIYLG--C---DKFPHESELPREWENNKEALLTFMEQVHRGIKG : 521
hAEBP1 : IVNGAKWNPRTGTINDFSYLHTNCLELSFYLG--C---DKFPHESELPREWENNKEALLTFMEQVHRGIKG : 521
cAEBP1 : IVNGAKWNPRTGTINDFSYLHTNCLELSFYLG--C---DKFPHESELPREWENNKEALLTFMEQVHRGIKG : 521
dAEBP1 : IVNGAKWNPRSGTINDFSYLHTNCLELSIYLG--C---DKFPHESELPREWENSKEALLTFMEQVHRGIKG : 516
bAEBP1 : IVNGAKWKPRSGTINDFSYLHTNCLELSIFLG--C---DKFPHESELPREWENNKEALLTFMEQVHRGIKG : 521
gAEBP1 : IVQGAKWRPRAGSMNDFSYLHTNCLELSVYLG--C---DKFPHESELQQEWENNKESLLTFMEQTHRGIKG : 477
fAEBP1 : IVNRASWKPVVGSMNDFSYLHTNCFELSIFLG--C---DKFPHESELPLEWENNKESLLAFIEQVHRGIKG : 526
tAEBP1 : IVNRASWKPVVGSMNDFSYLHTNCFELSIFLG--C---DKFPHESELPLEWENNKESLLSFIEQVHRGIKG : 520
zAEBP1 : IVNRAKWKPIPGSMDDFSYLHTNCFELSVFLG--C---DKFPHQSELLREWEHNREALLTFMAQVHRGIKG : 504
hCPX1 : IINGADWHTVPGSMNDFSYLHTNCFEVTVELS--C---DKFPHENELPQEWENNKDALLTYLEQVRMGIAG : 627
dCPX1 : IINGADWHTVPGSMNDFSYLHTNCFEITVELS--C---DKFPHENELPQEWENNKDALLTYLEQVRMGITG : 623
mCPX1 : VINGADWHTVPGSMNDFSYLHTNCFEVTVELS--C---DKFPHEKELPQEWENNKDALLTYLEQVRMGITG : 616
xCPX1 : IINGADWHTVAGSMNDFSYLHTNCFEVTVELS--C---DKFPHEVELPVEWENNKESLLVFMEQVHRGIKG : 524
fCPX1 : IINGGAWHTVPGSMNDFSYLHHNCFEVTVELS--C---DKFPHASELPIEWENNKESLLVYMEQVHRGIKG : 517
dCPX2 : TVNGASWHTVAGSLNDFSYLHTNCFELSIYVG--C---DKYPHESELPEEWENNRESLIVFMEQVHRGIKG : 511
mCPX2 : TVNGASWHTVAGSLNDFSYLHTNCFELSIYVG--C---DKYPHESELPEEWENNRESLIVFMEQVHRGIKG : 654
hCPX2 : TVNGASWHTVAGSLNDFSYLHTNCFELSIYVG--C---DKYPHESQLPEEWENNRESLIVFMEQVHRGIKG : 646
gCPX2 : TVNGASWHTVAGSINDFSYLHTNCFELSIYVG--C---DKYPHESELPEEWENNRESLIVFMEQVHRGIKG : 650
xCPX2 : TVNGASWHTVAGSINDFSYLHTNCFEISIYLD--C---DKYPHESELPEQWENNRESLIVFMEQVHRGVKG : 619
fCPX2 : TINGASWHTAAGSMNDFSYLHTNCFELSMYVG--C---DKFPHESELPEEWENNRESLLVFMEQVHRGIKG : 523
[redacted] : GPICSVI[redacted]QASGGSIDWSYDYGIKYSFAFELR----------DTGRY-GFLLPARQILPTAEETWLGLKA : 391
[redacted] : GPGATTI[redacted]PASGGSDDWAYDQGIKYSFTFELR----------DKGRY-GFVLPESQIQPTCEETMLAIKY : 297
[redacted] : PQQASDL[redacted]ITDGDMTDWAYGQHKIFAFTFEMYPTSYNPGFYPPDEVIGRETSRNKEAVLYVAEKADCPYSV : 318
[redacted] : ITNGAQW[redacted]NVPGGMQDWNYLNTNCFEVTIELG--C---VKYPKAEELPKYWEQNRRSLLQFIKQVHRGIWG : 305
[redacted] : VTNGYSW[redacted]PLQGGMQDYNYIWAQCFEITLELS--C---CKYPREEKLPSFWNNNKASLIEYIKQVHLGVKG : 317
[redacted] : TTNGGAW[redacted]SVPGGMQDFNYLSSNCFEITVELS--C---EKFPPEETLKTYWEDNKNSLISYLEQIHRGVKG : 378


720 * 740 * 760 * 780
mAEBP1 : VVTDEQ-GIPIANATISVSGINHGVKTA--SGGDYWRILNPGEYRVTAHAEGYTSSAKICNVDYDIGATQC : 589
rAEBP1 : VVTDEQ-GIPIANATISVSGINHGVKTA--SGGDYWRILNPGEYRVTAHAEGYTSSAKICNVDYDIGATQC : 589
hAEBP1 : VVTDEQ-GIPIANATISVSGINHGVKTA--SGGDYWRILNPGEYRVTAHAEGYTPSAKTCNVDYDIGATQC : 589
cAEBP1 : VVTDEQ-GIPIANATISVSGINHGVKT----------C : 548
dAEBP1 : VVTDEQ-GIPIANATISVSGINHGVKTA--SGGDYWRILNPGEYRVTAHAEGYTPSAKTCNVDYDIGATQC : 584
bAEBP1 : VVTDEQ-GIPIANATISVSGINHGVKTA--SGGDYWRILNPGEYRVTAHAEGYTPSSKTCNVDYDIGATQC : 589
gAEBP1 : LVTDQQ-GEPIANATIVVGGIKHSVRTA--SGGDYWRILNPGEYRVSARAEGYNPSTKTCSVFYDIGATQC : 545
fAEBP1 : VVRDVE-GNPLANATISVEGIWHDVKTA--AGGDYWRLLNPGEYKVTAKADGHTSQTRLCMVGYDSGATSC : 594
tAEBP1 : VVRDVE-GNPLANATISVEGIRHDVKTA--AGGDYWRLLNPGEYKVTAKADGYTPQTRLCMVGYDSGATSC : 588
zAEBP1 : VVRDNE-GNPITNATVSVEGVNHDVKTG--EAGDYWRLLNPGEYRVTARAEGYSPFTRLCVVGFDPGATLC : 572
hCPX1 : VVRDKDTELGIADAVIAVDGINHDVTT--AWGGDYWRLLTPGDYMVTASAEGYHSVTRNCRVTFEEGPFPC : 696
dCPX1 : VVRDKDTELGIADAVIAVDGINHDVTT--AWGGDYWRLLTPGDYMVTASAEGYHTATRSCRVPFEEGPVPC : 692
mCPX1 : VVRDKDTELGIADAVIAVEGINHDVTT--AWGGDYWRLLTPGDYVVTASAEGYHTVRQHCQVTFEEGPVPC : 685
xCPX1 : VVRDKDTEKGIPEAIIVVNELNHDIRT--AVDGDYWRLLNPGEYEVTAKAEGYHPSTKSCRVTYENHATIC : 593
fCPX1 : VIRDKVTKKGIPDAIVKVEDHSHDIRS--AADGDYWRLLNPGEYKVVVWAEGYFPSMRRCQVGVEPRPMIC : 586
dCPX2 : MVRDSH-GKGIPNAVISVEGVNHDIRT--ASDGDYWRLLNPGEYVVTAKAEGFTSSTKNCMVGYDMGATHC : 579
mCPX2 : IVRDLQ-GKGISNAVISVEGVNHDIRT--ASDGDYWRLLNPGEYVVTAKAEGFITSTKNCMVGYDMGATRC : 722
hCPX2 : LVRDSH-GKGIPNAIISVEGINHDIRT--ANDGDYWRLLNPGEYVVTAKAEGFTASTKNCMVGYDMGATRC : 714
gCPX2 : IVKDTH-GRGIPNAIISVEGVNHDIRT--GADGDYWRLLNPGEYLVSVKAEGYTTATKNCEVGYDMGATLC : 718
xCPX2 : VVRDVY-GRGIASAIISVEGINHDIRT--ASDGDYWRLLNPGEYAVTARAEGFTSSTKNCAVGYEMGATRC : 687
fCPX2 : VVRDAQ-GRGIPNATISVEGINHDIRTGMSADGDYWRLLNPGEYRVTARAQGYGLTSKKCEVGYEMGTSTC : 593
[redacted] : IMEHVRDHPY---------- : 401
[redacted] : VTSYVLEHL---------- : 306
[redacted] : IGKSCSTK---------- : 326
[redacted] : FVLDATDGRGILNATISVADINH--PVTTYKDGDYWRLLVQGTYKVTASARGYDPVTKTVEVDSKGG---- : 370
[redacted] : QVFDQNGN-PLPNVIVEVQDRKHICPYRTNKYGEYYLLLLPGSYIINVTVPGHDPHITKVIIPEKSQNFSA : 387
[redacted] : FVRDL-QGNPIANATISVEGIDH--DVTSAKDGDYWRLLIPGNYKLTASAPGYLAITKKVAVPYSPA---- : 442
* 800 * 820 * 840 *
mAEBP1 : NFILARSNWKRIREILAMNGNR----PILRVDP---------SRPMTPQQRRMQQRRLQYRLRMREQMRLR : 647
rAEBP1 : NFILARSNWKRIREILAMNGNR----PILRVDP---------SRPMTPQQRRLQQRRLRYRLRMREQMRLR : 647
hAEBP1 : NFILARSNWKRIREIMAMNGNR----PIPHIDP---------SRPMTPQQRRLQQRRLQHRLRLRAQMRLR : 647
cAEBP1 : NFILARSNWKGIREIKAMNGNR----PIPHIDP---------SRPMTPQQRRLQQRRLQHRLRLRAQMRLR : 606
dAEBP1 : NFILARSNWKRIREIMAMNGNR----PIPHIDP---------SRPMTPQQRRMQRRRLQYRLRMREQMRLR : 642
bAEBP1 : NFILARSNWKRIREIMAMNGNR----PIPRIDP---------SRPMTPQQRRMQRRRLQYRLRMREQMRLR : 647
gAEBP1 : NFVLARSNWKRIREIMAMNGNR----PIRRVVP---------GRPMTLRERLRLRQRLRLRQRLRERMRLR : 603
fAEBP1 : SFTLAKSNWDRIKQIMALNGKR----PIRIVTKANVVRTTPASSAVATTTESHASMQRAERLRRLRILRLR : 661
tAEBP1 : SFTLAKSNWDRIKQIMALNGKR----PIRIVTKANVARTTAASRVVAATTESHAKMQRAERLRRLRILRLR : 655
zAEBP1 : NFDLNKSNWDRIKQIMALHGNK----PIRLLSSGSRNGVRHYSHSSNQIPVSNSGDMSRARLRRLRLMRIR : 639
hCPX1 : NFVLTKTPKQRLRELLAAGAKV----PPDLRRR----------LERLRGQ : 732
dCPX1 : NFHLTKTPKQRLRELLAAGAKV----PPDLRRR----------LERLRGQ : 728
mCPX1 : NFLLTKTPKERLRELLATRGKL----PPDLRRK----------LERLRGQ : 721
xCPX1 : DFYLVKTAKQRLREMLANGKKL----PKELMLR----------LRQLRTR : 629
fCPX1 : DFTLTQTPIERLKEIRASGGKV----PQDLQLR----------LRALRLR : 622
dCPX2 : DFTLSKTNLARIREIMEKFGKQ----PISLPAR----------RLKLRGR : 615
mCPX2 : DFTLTKTNLARIREIMETFGKQ----PVSLPSR----------RLKLRGR : 758
hCPX2 : DFTLSKTNMARIREIMEKFGKQ----PVSLPAR----------RLKLRGQ : 750
gCPX2 : DFTISKTNLARIKEIMRKFGKQ----PISLSLR----------RLRQRAR : 754
xCPX2 : DFILAKSNLARIKEIMQKFGKQ----PVSMSLR----------RRRIRFR : 723
fCPX2 : DFTIGRTNLSRIKEIMERFNKQ----PIRLPHS----------NKQLQAP : 629
[redacted] : ---------- : -
[redacted] : ---------- : -
[redacted] : ---------- : -
[redacted] : ----VQVNFT----------LSRT---------- : 380
[redacted] : LKKDILLPFQGQLDSIPVSNPSCPMIPLYRNLP----------DHSA----------ATKPSLF : 431
[redacted] : ----AGVDFELES----------FSERKE----------EEKE----------ELMEWW : 467


860 * 880 * 900 * 920
mAEBP1 : RLNSTAGPATSPTPALMPP---------PSPTPAITLRPWEVLPTTTAGWEES------ETETYTEVVTEF : 703
rAEBP1 : RLNSTTGPATSPTPALTLP---------PSPTPGSTSRLWEILPTTAAGWEES------ETETYTEVVTEF : 703
hAEBP1 : RLNATTTLGPHTVPPTLPP---------APATTLSTTIEPWGLIPPTTAGWEES------ETETYTEVVTEF : 704
cAEBP1 : RLNATTTLGPHTVPPTLPP---------APATTLSTTIEPWGLGPPTTAGWEES------ETETYTEVVTEF : 663
dAEBP1 : RLNATATPALPTAPPGPT---------LPTPPRPEATP----------AGWEES------ETETYTEVVTEF : 689
bAEBP1 : RLNATTSPATTPTSPPPTPTLLPTADLAPSPAPSSTLGPWQFLPETTVNWEESETETETETETYTEVVTEF : 718
gAEBP1 : RLNATASTASVPTAPPPST--------ALPVPFSSTTYAPWSREPPTAGTWEME---------TETEVVTEL : 658
fAEBP1 : RVRQQRLRARQSTTPTTT---------TTTTTTTTMAPTTPPETERTTSWYDSW-------FPVDSWFTEN : 716
tAEBP1 : RLRQQRFRARQRTTPATT---------TTTTTTTTMAPTTS-KTERTTSWYDSW-------FPVDSWFTEN : 709
zAEBP1 : RLRQQRLLAQKIIP-------------TTTTVTTTSLPTTA---ETTTVWYDS----------LPEIFEEN : 684
hCPX1 : KD---------- : 734
dCPX1 : K---------- : 729
mCPX1 : K---------- : 722
xCPX1 : K---------- : 630
fCPX1 : K---------- : 623
dCPX2 : KRRQRG---------- : 621
mCPX2 : KRRQRG---------- : 764
hCPX2 : KRRQRG---------- : 756
gCPX2 : QWRQR---------- : 759
xCPX2 : LG---------- : 725
fCPX2 : RSRER---------- : 634
: ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -
: LFLVSLLHIFFK---------- : 443
: KMMSETLNF---------- : 476
* 940 * 960 * 980
mAEBP1 : ETEYGTD----------LEVEEIEEEEEEEEEEMDTGLTFPLTTVETYTVNFGDF : 748
rAEBP1 : ETEYGPD----------LEVEELEEEEEEEEE-MDTGLTFPVTTVETYTVNFGDF : 747
hAEBP1 : GTEVEPEFGTKVEPEFETQLEPEFETQLEPEFEEEEEEEKEEEIATGQAFPFTTVETYTVNFGDF : 769
cAEBP1 : GTEVEPEFGTQVEPEFETQLEPEFETQLEPEFEEEEEEEEEEIATGQAFPFTTVETYTVNFGDF : 728
dAEBP1 : GTEMGPE----------EEEEEDEEG-VVMGPDFPFTTVETYTVNFGDF : 727
bAEBP1 : GTELGPE----------EEEGEEGEGEMVTGPDFPFTTVETYTVNFGDF : 757
gAEBP1 : VTVT----------ELWDAGTGTAQPFTTAETYTVNFGD- : 687
fAEBP1 : PFDS----------IDLNSEPTQDYPFE----YTID---- : 738
tAEBP1 : PFDS----------INLNLEATQDYPFE----YTID---- : 731
zAEBP1 : TTPH----------EEIELTDTLDYNYN----YNINDY-- : 708
hCPX1 : ---------- : -
dCPX1 : ---------- : -
mCPX1 : ---------- : -
xCPX1 : ---------- : -
fCPX1 : ---------- : -
dCPX2 : ---------- : -
mCPX2 : ---------- : -
hCPX2 : ---------- : -
gCPX2 : ---------- : -
xCPX2 : ---------- : -
fCPX2 : ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -
: ---------- : -

When complete protein sequences are available for comparison, there are several large-scale features that distinguish AEBP1 from CPX1/2. First, both CPX-1 and CPX-2 have large N-terminal extensions. These extensions are distinct from that present in ACLP, an alternatively spliced isoform of AEBP1 also with an N-terminal extension. Second, CPX-1 and -2 lack the large C-terminal domain of AEBP1, but instead have a short 10-15 residue basic stretch at their C-termini. Third, mammalian AEBP1 contains two internal loops not present in most other carboxypeptidases. This is a distinguishing feature for mammalian AEBP1, but not for fish AEBP1, which may or may not retain the second loop. The presence of this loop is debatable due to the fact that most AEBP1 orthologues have been identified by the Ensembl database by reciprocal BLAST analysis. These cDNA/protein sequences have been predicted through comparison with known AEBP1 genes. The Ensembl database makes two predictions for Fugu AEBP1. One prediction includes a loop (VGNNDMNEETWARIQRQ NEGA) of a similar size to loop 2 of mammalian AEBP1 but with no sequence homology. Another prediction classifies this loop sequence as an intron and therefore eliminates it from the protein sequence. This sequence is also found in Tetraodon, but because both are pufferfish and highly related this could still be a well-conserved intron. It is not found in zebrafish AEBP1. Without experimental evidence we cannot be confident of this part of the fish AEBP1 sequence. One interesting note, however, is that secondary structure predictions do predict this possible Fugu loop sequence to form an α-helix, similar to the secondary structure predicted for this loop in other AEBP1 orthologues, although with no primary sequence homology. While CPX-1/2 do not contain this second loop, they do conserve the first loop which is absent in most other CPs. In summary, AEBP1 can be distinguished by three predominant large-scale features – (1 and 2) two loops within the CP domain which distinguish it from many CPs, although not all, and (3) the C-terminal domain.

In cases where the entire sequence of a protein is not available, another method is useful to ensure that a designation of AEBP1 is correct. Even with incomplete sequences, a phylogenetic tree clearly shows relationships between proteins in order to classify them into families. Therefore I searched the databases for putative orthologues of CPX-1 and CPX-2. These were aligned using ClustalW along with all identified orthologues of AEBP1 (Figure 14). A phylogenetic tree was produced showing three distinct branches, one each for AEBP1, CPX-1, and CPX-2 (Figure 15), suggesting that these proteins were correctly designated.

Because all gaps in the alignment were ignored when this phylogenetic tree was constructed we can be confident that like was compared with like. From this alignment we can also determine the residues of AEBP1 that define it as AEBP1, and not CPX, even in the absence of complete sequence showing the larger domain differences.

A close analysis of the multiple alignment indicates that there are few residues that are conserved in all AEBP1 orthologues and also unique to AEBP1 when comparing AEBP1 to CPX-1 and CPX-2. Only 17 AEBP1 residues, 21 CPX-1 residues, and 26 CPX-2 residues fit this classification of being conserved and unique when comparing the three groups. These residues distinguish an AEBP1 protein from a CPX-1 protein or a CPX-2 protein. Analysis of the 17 residues in AEBP1 that make it unique from CPX-1/2 shows that most are likely involved in maintaining the structural integrity of the protein. Within the CP domain of AEBP1, V276 and L379 are both hydrophobic residues in the central core of the protein. Prolines are often involved in placing kinks where needed in a structure, and so P464 and P530 likely fulfill this structural function. On the surface of the CP domain are T436 and T451. It is difficult to predict the function of these uncharged, polar residues. Also on the surface are two unique acidic residues, D259 and E304. It seems likely that D259 may form a hydrogen bond with nearby D217, while E304 may be able to form a salt bridge with R206, thus holding surface loops and helices together in the correct position. R206, with the exception of Fugu and Tetraodon, is also unique and conserved in AEBP1 (with the sole exception being K212 at the homologous site in zebrafish) and is likely necessary to hold the loop near E304 in place. Within the C-terminus of AEBP1 there are two conserved and unique tryptophans, W598 and W688, two arginines, R637 and R639, a glutamate E702, and a tyrosine Y741. The bulky hydrophobic structures of the tryptophans and tyrosine may help to maintain interdomain structure through hydrophobic interactions. Within the DLD domain there are two residues conserved and unique to AEBP1 (Y103 and N139), along with one residue, S158, in the linker region following the DLD domain. Once again, one might propose the surface-located tyrosine to be involved in some interdomain interaction. If one removes chicken AEBP1 from the alignment, because we are lacking the N-terminal 43 amino acids, two unique and conserved histidines become evident in the DLD domain, H4 and H19, although a function for these is not obvious. Interestingly, when one includes other carboxypeptidases (CPA, CPB,

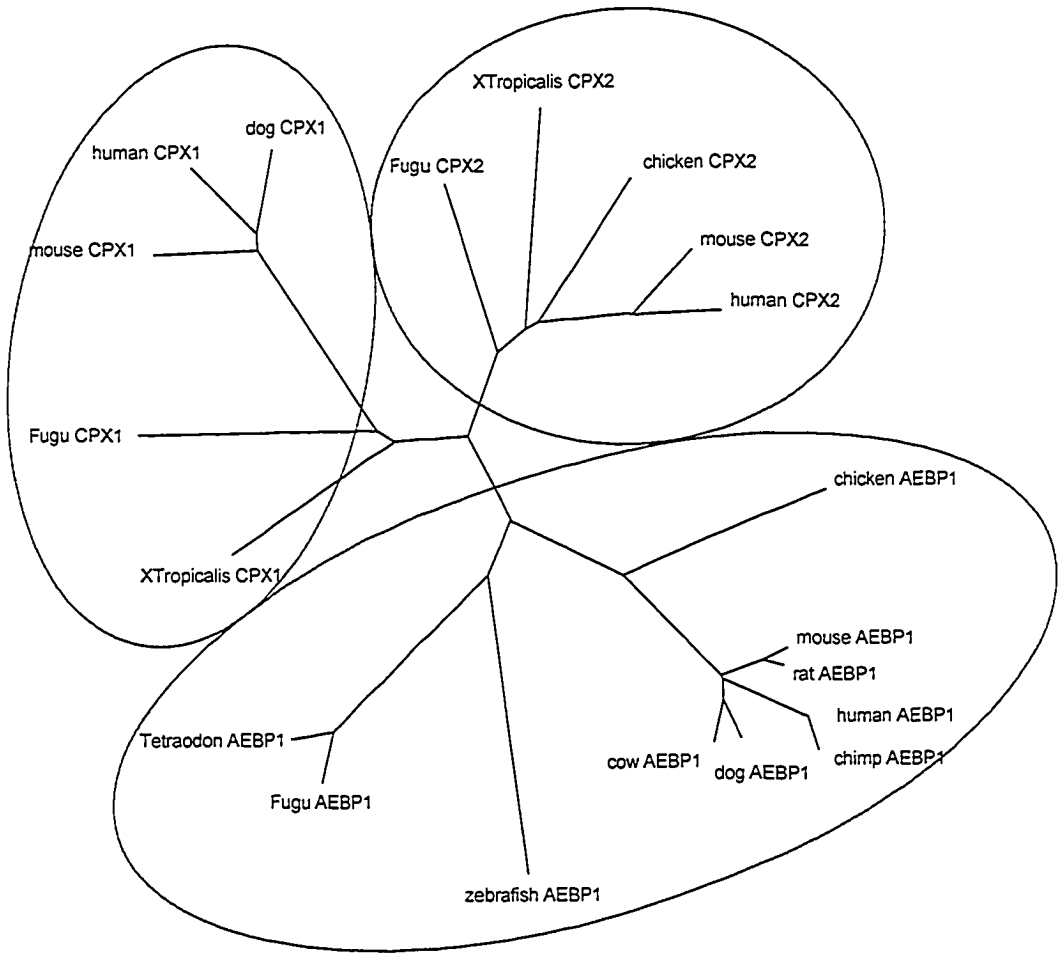


Figure 15. Phylogenetic comparison of AEBP1, CPX-1, and CPX-2. The Ensembl and NCBI databases were searched for potential orthologues of AEBP1, CPX-1, and CPX-2. These sequences were aligned using ClustalW and an unrooted phylogenetic tree produced from the dnd file with Treeview. Circled are the three distinct groups of orthologues. Branch lengths indicate number of substitutions per site according to the scale at the bottom left.

CPT, CPD, CPE, and CPM) in the alignment, most so-called unique residues in the CP domain of AEBP1 disappear, and we are left with only D259, E304, and T436. We might conclude that the only major differences between AEBP1 and other CPs are the large-scale differences in the C-terminal domains and the two CP domain loops.

The extensive similarities between AEBP1 and CPX-1/2 suggested that AEBP1 and CPX1/2 might have similar functions. In this case, comparison of AEBP1 and CPX1/2 to other CPs might allow us to see those residues associated with the putative function for this group. Residues L208, G291, P341, M362, E377, R424, W425, H476, D489, H493, C578, and C589 were found to be conserved and unique to the CP domains of AEBP1/CPX1/CPX2. Of these, residues likely involved in structural integrity include L208, G291, P341, C578, and C589. H476 may help to hold two surface helices together by forming a salt bridge with E363, which is conserved in all AEBP1 and CPX-2 but not in CPX-1, while neighboring M362 may also help in this interaction of two helices through hydrophobic interactions. R424, W425, D489, and H493 are all located on the surface and may be involved in some presently unknown interaction. E377 is located in proximity to the putative active site of AEBP1, and therefore is an interesting candidate for a residue involved in substrate interaction.

Finally, the broadest comparison we could make was to exclude CPX-1/2 and compare only AEBP1 orthologues with other more distantly related CPs. Residues N261, R265, W302, and H434, as well as several small and uncharged residues, were found to be conserved and unique to AEBP1 in addition to many residues already mentioned above. These residues are all located on the surface of the CP domain and no obvious prediction can be made regarding function. Additional analysis suggests a function for several other residues. N374, which is either an asparagine or a glutamine in all AEBP1 orthologues, replaces the zinc-coordinating histidine present in most active CPs, and may functionally replace this histidine as well. R388 is the only residue within the second loop conserved in all AEBP1 orthologues, and therefore may be necessary for the function of this loop. N470 is conserved in all AEBP1/CPX1/2 proteins except zebrafish AEBP1. Mutation of the paralogous residue in TAFI, a carboxypeptidase described earlier, from an aspartate to a glutamine resulted in a decrease in activity towards hippuryl-arginine and a slight increase in activity towards hippuryl-histidine-

leucine (*166*), suggesting that the asparagine in this postion in AEBP1 may allow AEBP1 to also have activity toward hippuryl-histidine-leucine.

When all the residues that make AEBP1 unique are mapped onto the modeled 3-D structure of the CP domain of AEBP1, nothing stands out. There is no area in which these unique residues are in high density. They are fairly evenly distributed throughout the protein, although the majority of these residues are located in loops or at the ends of β-strand structures. Additionally, since we know little about the interdomain interactions of AEBP1, little can be predicted regarding surface residues. Residues located near the putative active site of AEBP1 might have some role in the cleavage of, or interaction with, carboxypeptidase substrates. When one analyzes only AEBP1 orthologues for conservation, no particular domains stand out. The majority of the protein is very well conserved, except for the second loop and the C-terminal domain, which are not well-conserved between mammalian and fish and chicken orthologues. Altogether this analysis of AEBP1 and its homologues leads us to focus on large-scale domain and loop deletions, as well as on several amino acids in proximity to the active site, specifically N374, E377, R388, and N470. Other residues unique to AEBP1 may have a crucial function that we are not able to predict at this point.

NOTE: Following the completion of this thesis, some discussion has focused on the existence of the AEBP1 arising from proteolytic truncation or, as proposed by Ro et al. (*32*), through alternative splicing in which intron 9 is retained in the mature transcript. This alternative splicing mechanism relies on three things, which would be expected to be conserved: 1) a termination codon within intron 9 to stop translation of ACLP, 2) an initation codon within exon 10 to start translation of AEBP1, and 3) an optimal Kozak sequence surrounding the AEBP1 initiation site. The exact boundaries of fish and chicken intron 9 are not known. However, it has been observed that both Fugu and Tetraodon AEBP1 lack the methionine initiation codon in the proposed translation start site, suggesting that they may only translate an ACLP protein. All predicted mammalian AEBP1 genes have a conserved Kozak sequence at the translation initiation site, in which a purine (in this case, guanine) is located at position -3 and a guanine is located at position +4. All intron 9 sequences contain a conserved in-frame stop codon (`intron 9 5'-gtgagtag`...), except human and chimp sequences, which do not conserve this stop codon only 6 nucleotides from exon 9, but rather contain an

alternative in-frame stop codon approximately 250 nucleotides from exon 9 within intron 9 (see Ensembl database for intron sequences). Therefore the termination and initiation codons required for AEBP1 translation are not completely conserved across identified orthologues.

3.2 Expression of Full-Length and Truncated AEBP1

3.2.1 Expression and purification of recombinant AEBP1 in a bacterial system

Many biochemical techniques rely on the availability of well-purified protein components. Previously, AEBP1 was expressed in *E. coli* as a fusion protein with a polyhistidine (His6) tag and purified through metal-affinity chromatography of denatured *E. coli* extracts followed by stepwise protein renaturation. In my hands, this purification protocol has consistently given less than satisfactory results, with difficulty in getting a homogenous product and rapid degradation of the product. Thus efforts were made to optimize the expression plasmid and purification method for recombinant AEBP1. In the process, much was learned regarding the structure and characteristics of AEBP1.

The *E. coli* expression plasmid used for earlier studies on AEBP1, pET16b-AEBP1, was constructed by digesting the AEBP1 cDNA with NaeI and StuI restriction enzymes followed by blunt-end ligation into the NdeI site (blunted with Klenow) of the pET16b expression vector in frame with the amino terminal polyhistidine tag (*8*). Twenty-six amino acids from the N-terminus and 18 amino acids from the C-terminus of AEBP1 were not contained in this fusion protein (Figure 16). As the first step in my optimization of the expression of AEBP1 in *E. coli*, I constructed two plasmids encoding full-length AEBP1. These plasmids were pET21d-AEBP1 and pET22b-AEBP1, which both encoded a His6 tag at the C-terminus of AEBP1 (Figure 16). A C-terminal tag was chosen, because I considered that the C-terminal basic and STP regions might pose codon usage problems in *E. coli*, which has low levels of many mammalian codons for arginine and proline. A C-terminal tag would prevent premature translation termination products from being purified as they would not have the His6 tag. In pET22b-AEBP1, the protein expressed had a pelB leader sequence fused to the N-terminus to target it to the periplasm, a more favourable environment for protein folding (*167*).

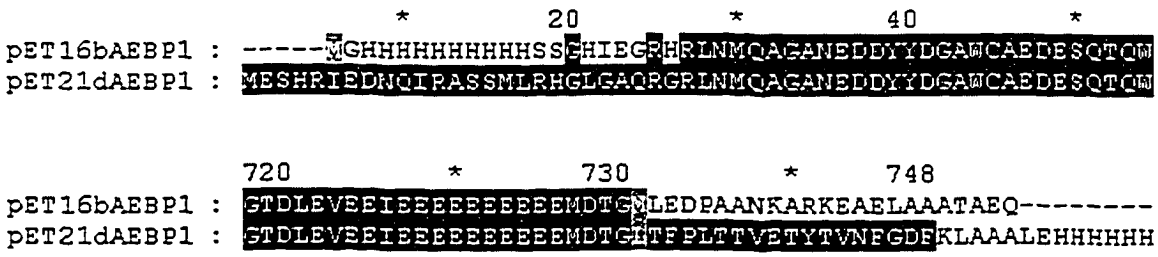


Figure 16. Comparison of proteins encoded by pET16b-AEBP1 and pET21d-AEBP1. The proteins encoded by pET16b- and pET21d-AEBP1 were aligned using GeneDoc. Shown here are are the N-terminal residues (top) and C-terminal residues (bottom). All AEBP1 protein sequence is shaded black and any chemically similar sequences are shaded grey. Amino acid numbering, based on wild-type mouse AEBP1, is shown above the alignment.

It is generally desirable to purify a protein in its soluble, native form, avoiding the process of renaturation from insoluble inclusion bodies. Inclusion bodies are thought to be formed when partially folded intermediates expose hydrophobic "sticky" surfaces, which self-associate and lead to protein aggregation (*168, 169*). Inclusion body formation in *E. coli* cells can be correlated with many parameters of protein composition, such as charge, the number of turn-forming residues, the number of cysteine and proline residues, hydrophilicity, and total number of residues (*170*). While these parameters cannot be changed for a given protein, many parameters of the protein expression system can be optimized for protein solubility. These parameters include the rate of protein expression, subcellular localization, and the formation of disulfide bonds. The pET22b-AEBP1 expression plasmid, which should have exported AEBP1 to the bacterial periplasm, was initially transformed into Tuner DE3 *E. coli* (Novagen). This host has a mutation in the lacYI permease which would allow for more uniform entry of the inducer into all cells in the population so that expression levels could be finely regulated. It is also deficient in two proteases which have been reported to allow some proteins to be more stable in this strain (*167, 171*). Unfortunately, AEBP1 expressed from this plasmid was not efficiently exported to the periplasm and therefore this was not used in further studies (data not shown).

The wild-type bacterial cytoplasm is a highly reducing environment because of the presence of reductases such as gor and trxB and therefore is not conducive to the disulfide bond formation necessary for the proper folding of many eukaryotic proteins (*172*). Some proteins that require disulfide bond formation have been reported to be expressed in soluble form in *E. coli* with mutations in the gor and trxB genes (*173*). Because AEBP1 is predicted by homology modeling to form four disulfide bonds, a considerable barrier to solubility in *E. coli*, Origami B (DE3) *E. coli* was used as a host to promote its expression in a soluble form. In addition to the lacYI permease mutation and protease deficiency described for Tuner DE3 cells, Origami B (DE3) *E. coli* also contain mutations within the glutathione reductase (gor) and thioredoxin reductase (trxB) genes (*172*). Expression of AEBP1 from pET21d-AEBP1 and from pET16b-AEBP1 in Origami B (DE3) *E. coli* was induced and the solubility of the protein expressed from each plasmid compared (Figure 17). This induction was performed with 0.1 mM IPTG and growth overnight at room temperature to allow a lower rate of protein expression, generally considered optimal for proper protein folding. A significant difference

can be seen in the relative amount of soluble AEBP1. Expression from pET16b-AEBP1 results in AEBP1 with about 40% solubility, whereas expression from pET21d-AEBP1 results in near total solubility. Therefore expression of AEBP1 from pET21d-AEBP1 in Origami B *E. coli* was further optimized (Figure 18). Induction for 3.5 hours at 37°C resulted in higher yields (approximately 5-10 fold greater) of soluble AEBP1 protein than when induction was performed overnight at room temperature, although the amount of insoluble AEBP1 was also increased. The 3.5 hour induction at 37°C also resulted in the production of fewer truncation products than seen in the overnight induction. The conventional induction protocol, in which protein expression was induced when cells reached logarithmic growth phase, gave more soluble AEBP1 than a less-common protocol in which cells were grown for a standard 4 hours before induction. Induction with either 0.1 or 1.0 mM IPTG gave equivalent expression of AEBP1. Although soluble AEBP1 was preferred for subsequent biochemical analysis, total protein staining of the insoluble fraction indicated that this fraction was largely made up of AEBP1 , as the prominent band seen at 85 kDa (Figure 18, bottom panel) was absent when expression of non-AEBP1 proteins was induced (results not shown). Therefore a significant purification of AEBP1 has already been completed by isolating the insoluble fraction.

AEBP1, expressed using the optimized conditions, was subjected to Ni-NTA column affinity purification from the soluble fraction. However, based on western blot analysis of column input, flowthrough, and eluate, only 2-5% of total AEBP1 was able to bind to the Ni-NTA resin (results not shown). Batch incubation of extracts with Ni-NTA resin versus direct loading onto the column were compared. Additionally, TCEP was added to extracts to partially reduce disulfide bonds. However, neither of these approaches improved column binding. Despite being soluble, the His6 tag on the AEBP1 fusion protein was not interacting with the affinity column under the conditions used.

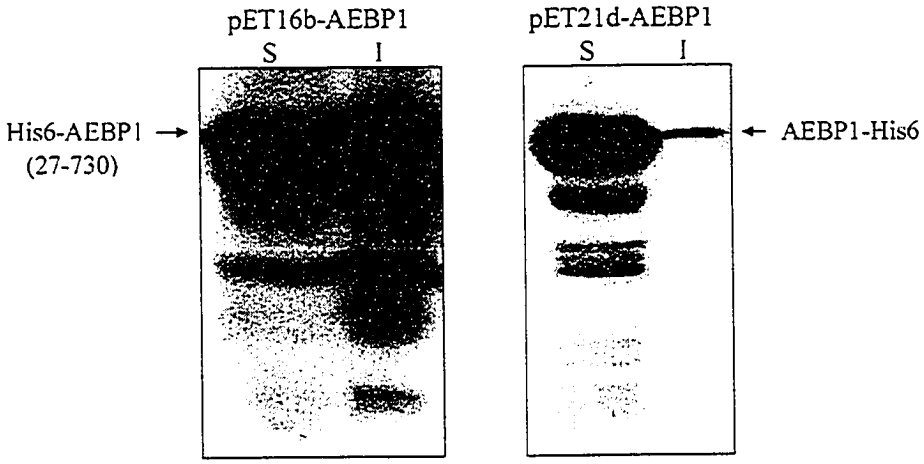


Figure 17. Comparison of AEBP1 expressed from pET16b-AEBP1 and pET21d-AEBP1 in Origami B *E. coli*. AEBP1 expression was induced overnight at room temperature using 0.1 mM IPTG. Equal amounts of protein from soluble (S) and insoluble (I) fractions were resolved by SDS-PAGE and immunoblotted with an anti-His6 antibody.

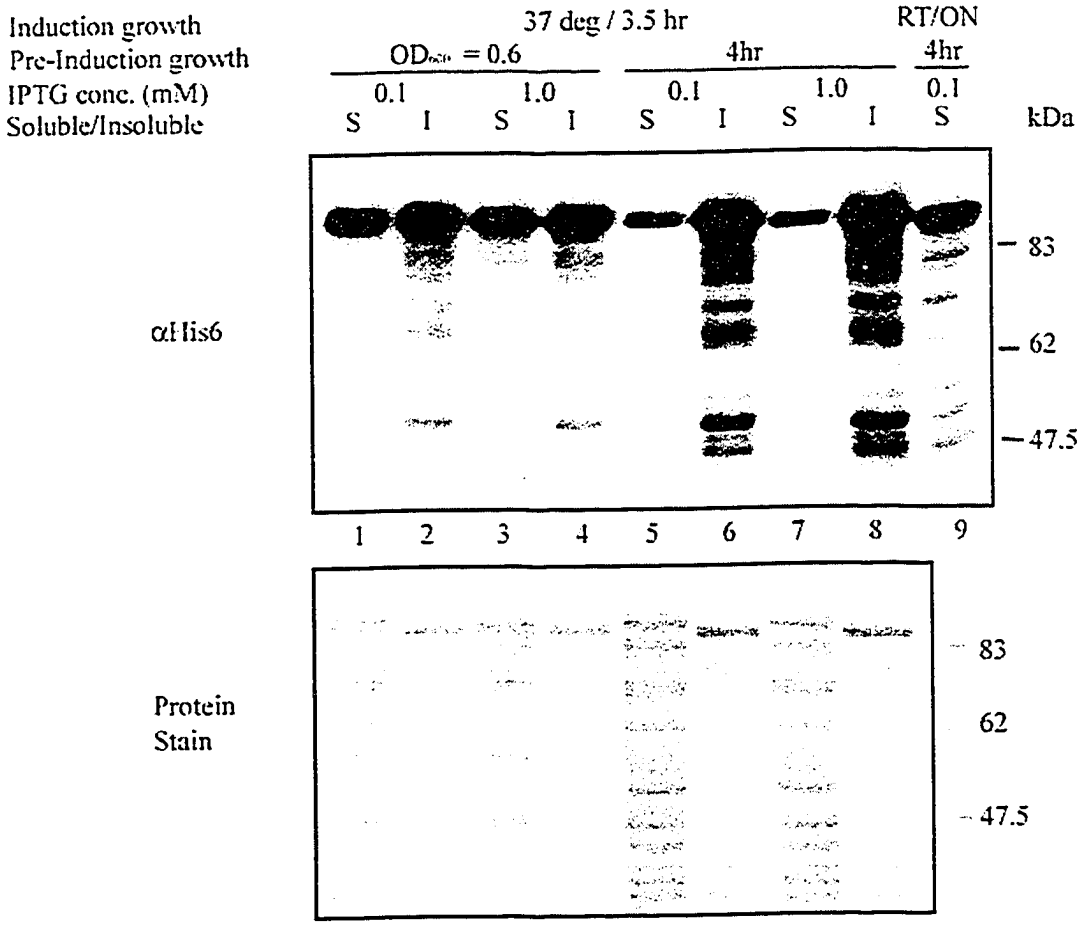


Figure 18. Optimization of AEBP1 expression in *E. coli*. Origami B (DE3) *E. coli* transformed with pET21d-AEBP1 were induced under various conditions as indicated. Prior to induction, cultures were grown at 37°C for 4 hrs or until the OD600 was approximately 0.6. Cells were induced with either 0.1 or 1.0 mM IPTG for 3.5 hrs at 37°C or overnight (ON) at room temperature (RT). Protein was extracted by French press and equivalent amounts of soluble (S) and insoluble (I) fractions were analyzed by SDS-PAGE followed by Western blotting using an anti-His6 antibody (top) or total protein staining (bottom). The gel shown in the bottom panel, as well as lane 9 of the top panel, was loaded with approximately 7X more protein than that shown in lanes 1-8 of the top panel.

Since the soluble AEBP1 could not be affinity-purified, expression of AEBP1 was induced in BL21 (DE3) *E. coli* cells. These are healthier and faster-growing *E. coli* than the Origami B *E. coli*, and although not optimal for expression of soluble AEBP1 due to the reducing environment of the bacterial cytoplasm, are optimized for high-level protein expression. The IPTG concentrations were optimized for greatest expression of AEBP1 in these cells (Figure 19). In contrast to induction of AEBP1 expression in Origami B *E. coli*, in which low concentrations of IPTG induced as much AEBP1 as high concentrations of IPTG, very little AEBP1 was induced in BL21 (DE3) *E. coli* cells at low levels of IPTG. The optimal IPTG concentration in BL21 cells was approximately 1.0 mM. In most of my subsequent purifications 2 mM IPTG was used to ensure high level expression as well as the partitioning of greater than 95% of induced AEBP1 into the insoluble inclusion body fraction (Figure 19).

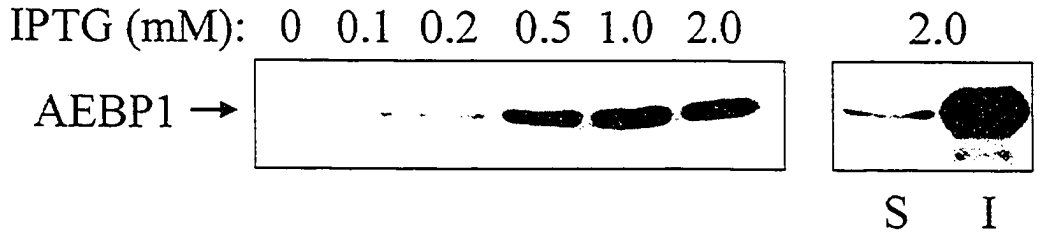
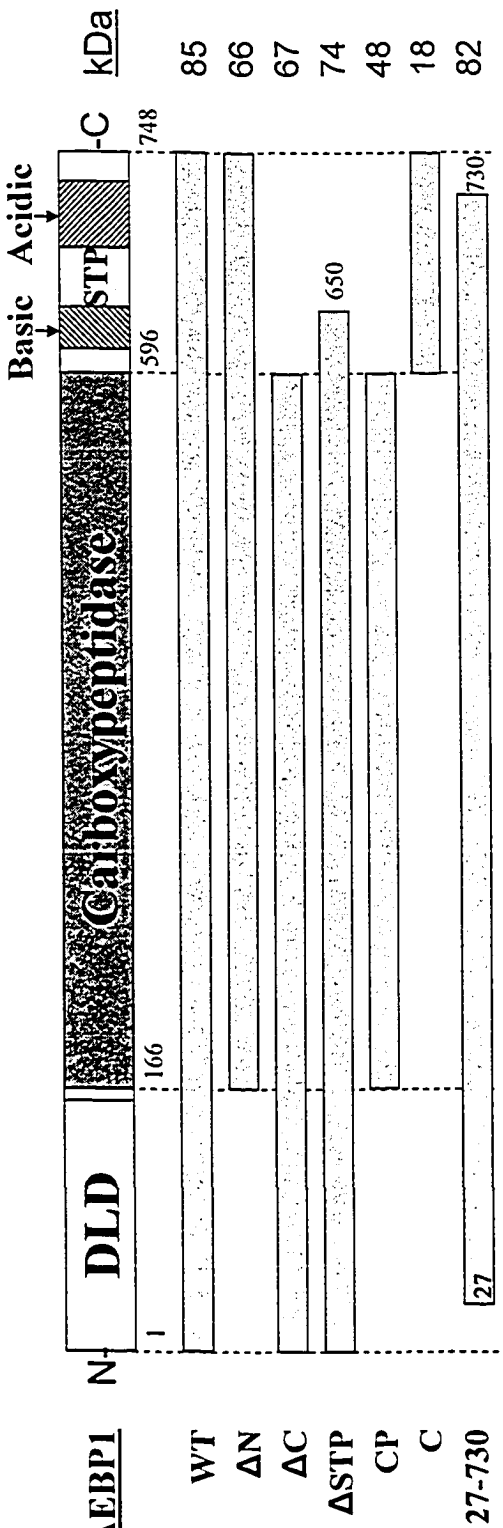


Figure 19. Optimal concentrations of IPTG result in high level expression of AEBP1 in inclusion bodies. BL21 *E. coli* cells transformed with pET21d-AEBP1 were grown to OD600=0.6. AEBP1 expression was induced for 3.5 hours at 37°C with the indicated concentrations of IPTG. Total protein was extracted from each culture and for the 2 mM IPTG induction only, separated into soluble (S) and insoluble (I) fractions. Proteins were resolved by SDS-PAGE and AEBP1 was detected by western blotting with an anti-His6 antibody.

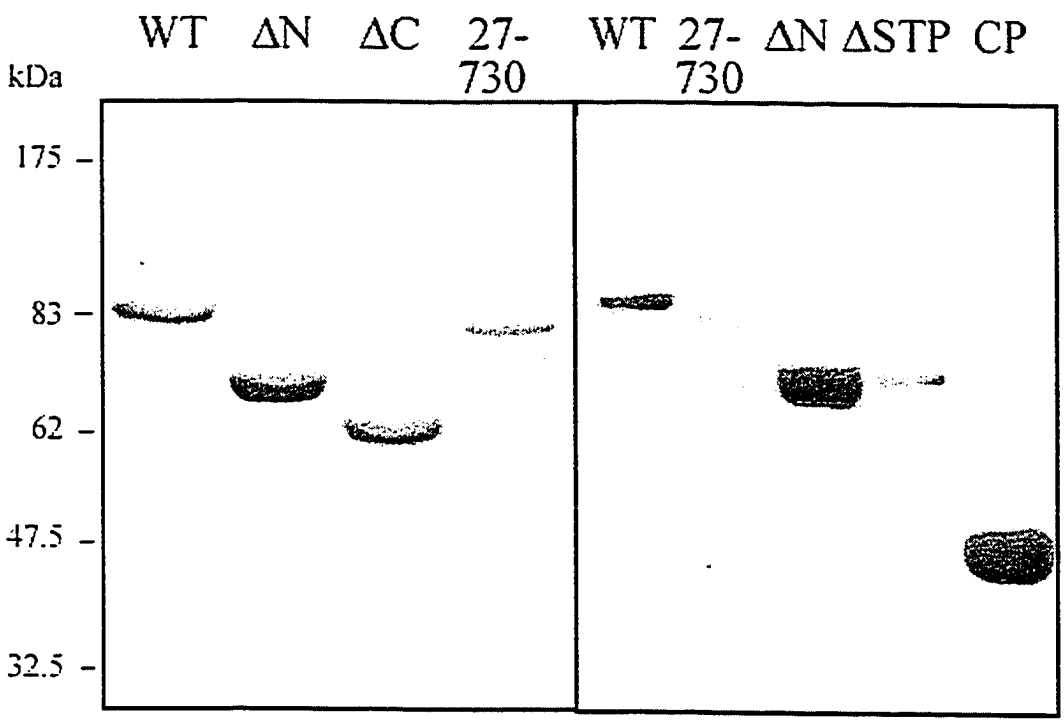
In addition to full-length AEBP1, expression of C-terminally His6-tagged AEBP1 deletion derivatives was induced and the truncated proteins purified under the same conditions as those found to be optimal for full-length AEBP1 (Figure 20). The yields of AEBP1-ΔN and AEBP1-CP were generally higher than AEBP1-WT or other deletions, while significantly less AEBP1-ΔSTP was obtained. In fact, ΔSTP tended to precipitate out of solution during the step-wise dialysis renaturation procedure unless it was diluted to concentrations of 80 ng/μl or less. The protein expressed from pET21d-AEBP1 was also compared with that expressed from pET16b-AEBP1 with both being induced and purified under the same conditions. While the full-length AEBP1 protein purified as a single major species with an observed molecular weight of 85 kDa, the protein expressed from pET16b-AEBP1 (AEBP1 27-730) appeared as three major bands of apparent molecular weights of 82, 76 and 70 kDa and numerous lower molecular weight species. There are two major differences between pET16b- and pET21d-expressed AEBP1 proteins that may be cause of these multiple protein species. First, the AEBP1 protein expressed from pET16b contains an N-terminal His6 tag whereas pET21d-AEBP1 expresses a C-terminally tagged AEBP1. An N-terminal tag will enable products of premature translation termination to be purified along with full-length products. Secondly, the protein expressed from pET16b is lacking residues from both the N- and the C-termini of AEBP1. Deletion of these residues may prohibit the folding of AEBP1 into a compact structure and thus make it more available for proteolytic attack.

Figure 20. AEBP1 and deletion derivatives expressed in and purified from *E. coli*. (A) A diagram is shown illustrating the various versions of AEBP1 expressed as fusion proteins in this study and their calculated molecular weights. (B) His6-tagged AEBP1 (WT) and some of the truncated derivatives shown in (A) were expressed in *E. coli*, purified on Talon metal affinity columns, and equal volumes of eluate analyzed by SDS-PAGE followed by protein staining with GelCode Blue. AEBP1 27-730 was expressed from pET16b-AEBP1. All others were expressed from pET21d-AEBP1 and derivative plasmids. The two panels shown in B show proteins purified separately and resolved on separate gels.

A.


AEBP1
N
DLD
Carboxypeptidase
Basic
Acidic
STP
-C
kDa
1
166
596
748
650
27
730
WT
ΔN
ΔC
ΔSTP
CP
C
27-730
85
66
67
74
48
18
82

B.



3.2.2 Expression of AEBP1 in mammalian cells.

The coding regions for AEBP1 as well as all deletion and point mutants were inserted into the pcDNA3.1-Myc/His mammalian expression vector resulting in the expression of proteins lacking the vector-encoded tags due to a prior in-frame stop codon. These plasmids were transfected into CHO cells and resulted in overexpression of full-length and all mutant versions of AEBP1 (Figure 21A). A polyclonal anti-AEBP1 antibody was used to detect the overexpressed protein. Because the mutant versions of AEBP1 may lack some of the epitopes recognized by this antibody we cannot be confident that the band intensities truly correlate with protein abundance. However, in the event that the deletions and mutants are recognized by the anti-AEBP1 antibody to a similar extent as wild-type AEBP1, we might make several conclusions. The AEBP1ΔN mutant protein consistently exhibited lower protein levels than the other versions of AEBP1. Wild-type, ΔC, and ΔSTP AEBP1 protein levels were typically similar, while AEBP1 CP levels were often higher than the levels of the other proteins detected by the anti-AEBP1 antibody. When full-length AEBP1 was overexpressed from pcDNA-AEBP1, a single major species with an apparent molecular weight of 85 kDa was observed along with a minor band of about 60 kDa. When an N-terminally HA-tagged version of AEBP1, AEBP1Δ1-31, was expressed from pJ3H-AEBP1 in CHO cells, full-length AEBP1 and 60 kDa bands were also observed when probed with an anti-HA antibody (data not shown). The N-terminal position of the tag in the fusion protein suggested that this truncation was at the C-terminus in both the WT AEBP1 and HA tagged fusion proteins. However, another plasmid expressing AEBP1 with a C-terminal myc/His tag was also made, resulting in a larger protein of 90 kDa apparent MW, consistent with the 88 kDa MW expected for this fusion protein. As can be seen in Figure 21B, lanes 12 and 13, there is a truncated, approximately 62 kDa band in this sample also. This suggests that the truncation in this case is at the N-terminus, thus retaining the C-terminal tag in the 60 kDa truncated band. Because a truncated form of AEBP1 is consistently observed, it raises the question as to whether a truncated form might also be present *in vivo*, reflecting proteolytic loss of either the N-terminal DLD or C-terminal domains and possible functional regulation of AEBP1. The potential for such cleavage has not been studied further.

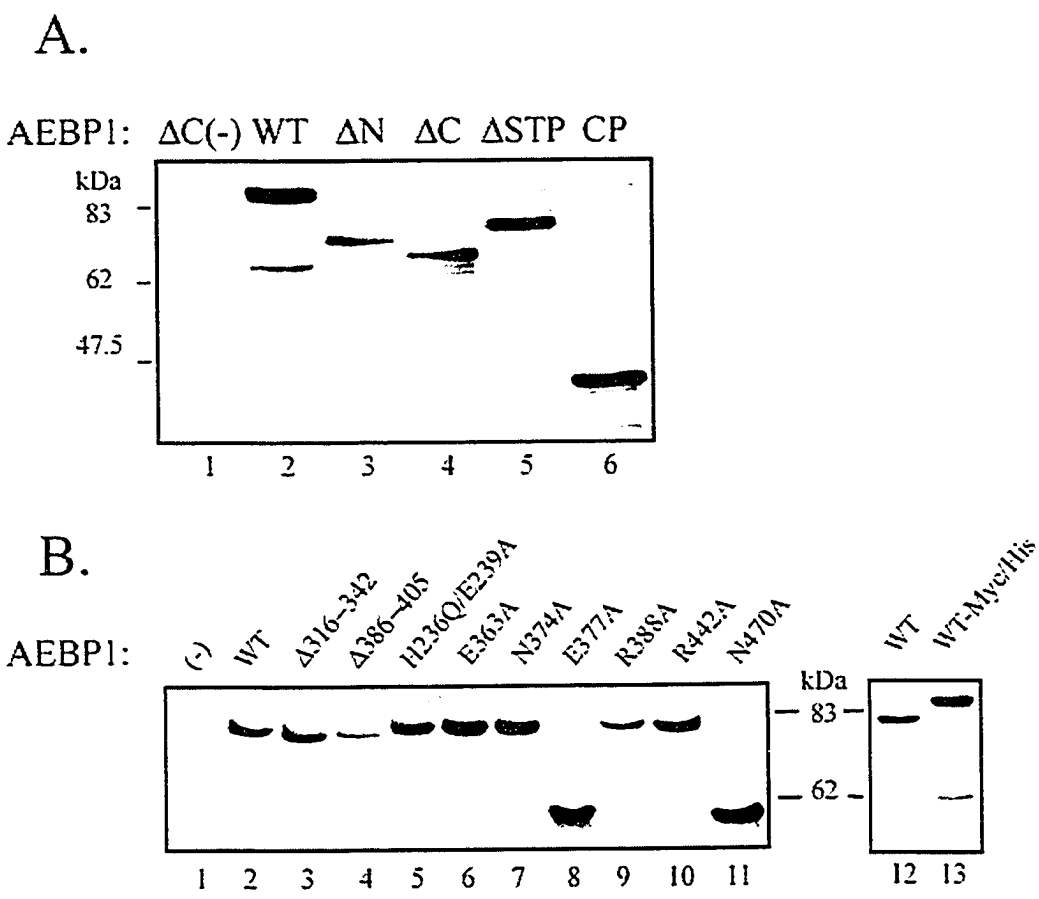


Figure 21. Expression of wild-type and mutant versions of AEBP1 in CHO cells. AEBP1 and derivative deletions and mutants were expressed from a pcDNA based vector in CHO cells. AEBP1 shown in lane 13 of (B) was C-terminally fused with a Myc/His tag, while all other proteins were untagged. As negative controls, cells were transfected with a pcDNA3.1 plasmid in which the AEBP1ΔC ORF was inserted in the opposite orientation (ΔC(-)) or empty vector (-). Total protein was extracted from the transfected cells and equal amounts were separated by SDS-PAGE and immunoblotted with an anti-AEBP1 polyclonal antibody.

Several site-directed mutants and smaller deletions of AEBP1 were also made based on the homology modeling and phylogenetic analysis described in the previous section. Two large loops within the CP domain of AEBP1, that are not present in many related CPs, were deleted in the AEBP1 Δ316-342 and Δ386-405 deletions. Three residues, H236, E239, and N374, equivalent to known zinc coordinating residues in related CPs, were mutated to glutamine or alanine in a double mutant, H236Q/E239A and a single mutant N374A . These proteins were not expected to exhibit any CP activity. E363 was predicted to form a salt bridge with H476 that would be disrupted upon mutation of E363 to an alanine. E377 was considered to be a possible functional replacement for the glutamate that performs the function of a general base in the enzymatic mechanism of most CPs but is replaced by Y485 in AEBP1. Therefore mutation of E377 to an alanine was also predicted to be enzymatically inactive. R388 is the only absolutely conserved residue in the 386-405 loops and predicted through analysis of the homology model to possibly function in substrate binding. R442 was also thought to be a substrate binding candidate residue, while N470 was considered possibly involved in substrate specificity, as mutation of the equivalent residue in TAFI resulted in an apparent alteration in substrate specificity from Hippuryl-Arg to Hippuryl-His-Leu (*166*). These were expressed in CHO cells and analyzed by immunoblotting. All but two of the deleted or mutant proteins exhibited levels similar to wild-type AEBP1. Δ386-405 AEBP1 and R388A both had much lower steady state levels than WT and other mutant forms of AEBP1. R388 is the only residue in the amino acids 386-405 loop that is conserved in all AEBP1 orthologues analyzed to date, and therefore the lack of this one residue may be responsible for the decreased levels of Δ386-405 AEBP1 also. It may be that R388 is important for maintaining the tertiary structure of AEBP1 or for some other aspect of protein stability. AEBP1 mutants E377A and N470A had much higher mobility on this gel due to a plasmid construction error. In the case of E377A, two nucleotides were inappropriately lost from the SmaI site during construction of the pcDNA-AEBP1 E377A plasmid resulting in a frameshift mutation and the expression of a C-terminally truncated protein (amino acids 1-559 of AEBP1 E377A). Similarly, one nucleotide was lost from the SmaI site during construction of the pcDNA-AEBP1 N470A expression plasmid thus creating a frameshift mutation expressing amino acids 1-560 of AEBP1 along with an additional 4 unrelated amino acids (VSTV) at the C-terminus. Both of these expressed proteins did contain the desired point mutation within the CP domain. Comparison of

AEBP1 E377A and N470A with AEBP1 ΔC (not shown) indicates these mutations did not affect steady-state levels of the protein.

3.3 Function of the CP domain of AEBP1

The presence of a CP domain is the most striking characteristic of AEBP1. This is a large domain which shares high sequence identity with other CPs. Much controversy has swirled around whether this domain has enzymatic activity or not. While it lacks many residues considered necessary for CP activity, it has been documented to still retain low enzymatic activity (*2, 164*). Whether or not this enzymatic activity is significant, the CP domain has been found to be necessary for the function of AEBP1 as a transcriptional repressor (*2*). To date, the exact substrate or interacting partner for the CP domain of AEBP1 leading to its transcriptional repression ability has not been determined.

Predictions of possible interacting partners and functions for AEBP1 have been made based on its structure and on experimental observations. One observation was that the CP activity of AEBP1 can be inhibited by captopril (*164*), an inhibitor of dipeptidases and most notably used as an angiotensin-converting enzyme (ACE) inhibitor in the treatment of hypertension. Although captopril was used under the false assumption that it was a CP inhibitor rather than a dipeptidase inhibitor, inhibition of the CP activity was observed (*164*). This inhibition suggested that the active site of AEBP1 may interact with captopril, a dipeptide peptidomimetic, and that AEBP1 may have a dipeptidase function. Two recent reports illustrate how a conserved domain might exhibit different substrate specificities in different family members. ACE2, a recently discovered enzyme structurally related to ACE, cleaves a single C-terminal amino acid rather than having the dipeptidase activity of its close relative, ACE (*108, 174*). As well, TAFI, a member of the metallocarboxypeptidase family, is able to cleave Hippuryl-His-Leu, the traditionally used synthetic substrate for dipeptidases such as ACE (*166*). Additional support for a dipeptidase function for AEBP1 can be found in the sequence of AEBP1. Two arginines (equivalent to Arg135 and Arg145 in CPDII, see Figure 6) generally thought to be necessary for interaction of CPs with the C-terminal carboxyl group of a substrate protein are absent in AEBP1. Interestingly, these arginine residues are also absent in a distantly related bacterial carboxypeptidase which is also able to function as a dipeptidase (*175-177*).

While AEBP1 may function as a dipeptidase, an equally likely possibility is that this domain of AEBP1 interacts with other proteins. This possibility has been previously proposed

based on an observation from another related CP, carboxypeptidase X2 (CPE), which lacks enzymatic activity, but retains the ability to bind C-terminal arginine (*116*). Additionally, mutation of the CPE residue, E300, proposed to function as the general base in the catalytic mechanism, to a glutamine, eliminated enzymatic activity but did not abolish the ability of the enzyme to interact with substrates (*112, 122*). The residue equivalent to CPE E300 in AEBP1, as well as CPX-2 and CPD-III is a tyrosine, suggesting that this may be sufficient for substrate binding in these proteins. Another study investigating the role of the equivalent residue in CPM (E264) also found that mutation of this residue eliminated enzymatic activity but retained the ability to bind C-terminal arginine (*117*). Although low CP activity was previously found for AEBP1 (*2, 164*), deletions which eliminated the zinc-coordinating residues H236 and E239 or possible catalytic general base residues E481 or Y485 did not eliminate enzymatic activity (*2*). It would be expected that mutation of catalytic residues would eliminate the CP activity of AEBP1 as it did in the related carboxypeptidases CPE and CPM mentioned above. The low activity of AEBP1 and the inability to eliminate this activity upon mutation of active site residues suggests that this activity may not be physiologically relevant, and the CP domain of AEBP1 may, instead, mediate interaction with C-terminal dipeptide motifs.

Two proteins known to interact with AEBP1, MAP kinase and PTEN, both contain a penultimate basic amino acid, raising the possibility that the CP domain of AEBP1/ACLP may mediate these interactions by binding the C-terminal dipeptides. My homology model for AEBP1 provides some support for this hypothesis. A comparison of the active site cavities of AEBP1 and CPDII indicates that AEBP1 may have a larger cavity with which to accommodate a C-terminal dipeptide substrate (Figure 22).

Figure 22. The active site cavity of AEBP1 is larger than that of CPDII. The active sites of the AEBP1 model (1AEB) and the X-ray crystal structure of CPD-II (1HSL) were visualized using the surface representation mode of VMD. The coloring (blue to green to yellow to red) indicates progressively shorter distance to the center of the protein. The GEMSA inhibitor molecule is modeled as a stick structure, while the catalytic zinc ion is modeled as a grey sphere.

3.3.1 The CP domain of AEBP1 interacts with C-terminal motifs.

In order to test the C-terminal interaction ability of the CP domain of AEBP1, the interaction of AEBP1 with PTEN was examined. The N-terminus of AEBP1 has previously been shown to interact with the C-terminus of PTEN in a yeast two-hybrid genetic assay (*178*). In this study, the two fusion proteins shown to interact were a truncated version of AEBP1 containing approximately amino acids 35-235, and the C-terminal half of PTEN. Amino acids 35-235 of AEBP1 include the majority of the N-terminal DLD domain, as well as about 80 amino acids of the CP domain. This suggests that the DLD domain and/or the N-terminal 80 amino acids of the CP domain of AEBP1 are sufficient for interaction with the C-terminal region of PTEN. A GST pull-down assay was performed in which GST-tagged PTEN was used to pull down His6-tagged full-length and truncated versions of AEBP1 (Figure 23). The GST-PTEN fusion protein interacted strongly with all versions of AEBP1, except that in which only the C-terminus of AEBP1 remained (C). A small amount of both AEBP1ΔN and C were pulled down non-specifically with GST (Figure 23A, lanes 2 and 5). However, significantly more AEBP1ΔN was brought down with GST-PTEN, indicating that it did interact specifically with the PTEN portion of the fusion protein. In contrast, the amount of AEBP1-C brought down with GST-PTEN was not significantly different than that brought down with GST alone. Because AEBP1ΔN which lacks the DLD domain of AEBP1, and AEBP1ΔC which lacks the C-terminus of AEBP1, are able to interact with PTEN while AEBP1-C, which is only the C-terminus of AEBP1, is not sufficient to interact with PTEN, this suggests that AEBP1 interacts with PTEN through its CP domain.

To test whether AEBP1 interacts with the C-terminal dipeptide motif of PTEN, a GST-PTEN protein C-terminally truncated by two amino acids was used in a GST-pulldown assay with both AEBP1-WT and AEBP1-CP (Figure 23B and C). Both WT and CP AEBP1 were pulled down with the GST-PTEN even when it lacked the extreme C-terminal motif (Figure 23B and C, lane 3). This suggests that PTEN does not require this extreme C-terminal region for interaction with AEBP1. The CP domain of AEBP1 was not brought down by GST but was as efficiently brought down as full-length AEBP1 by both GST-PTEN fusion proteins. The interaction of the isolated CP domain with PTEN supports the hypothesis that AEBP1

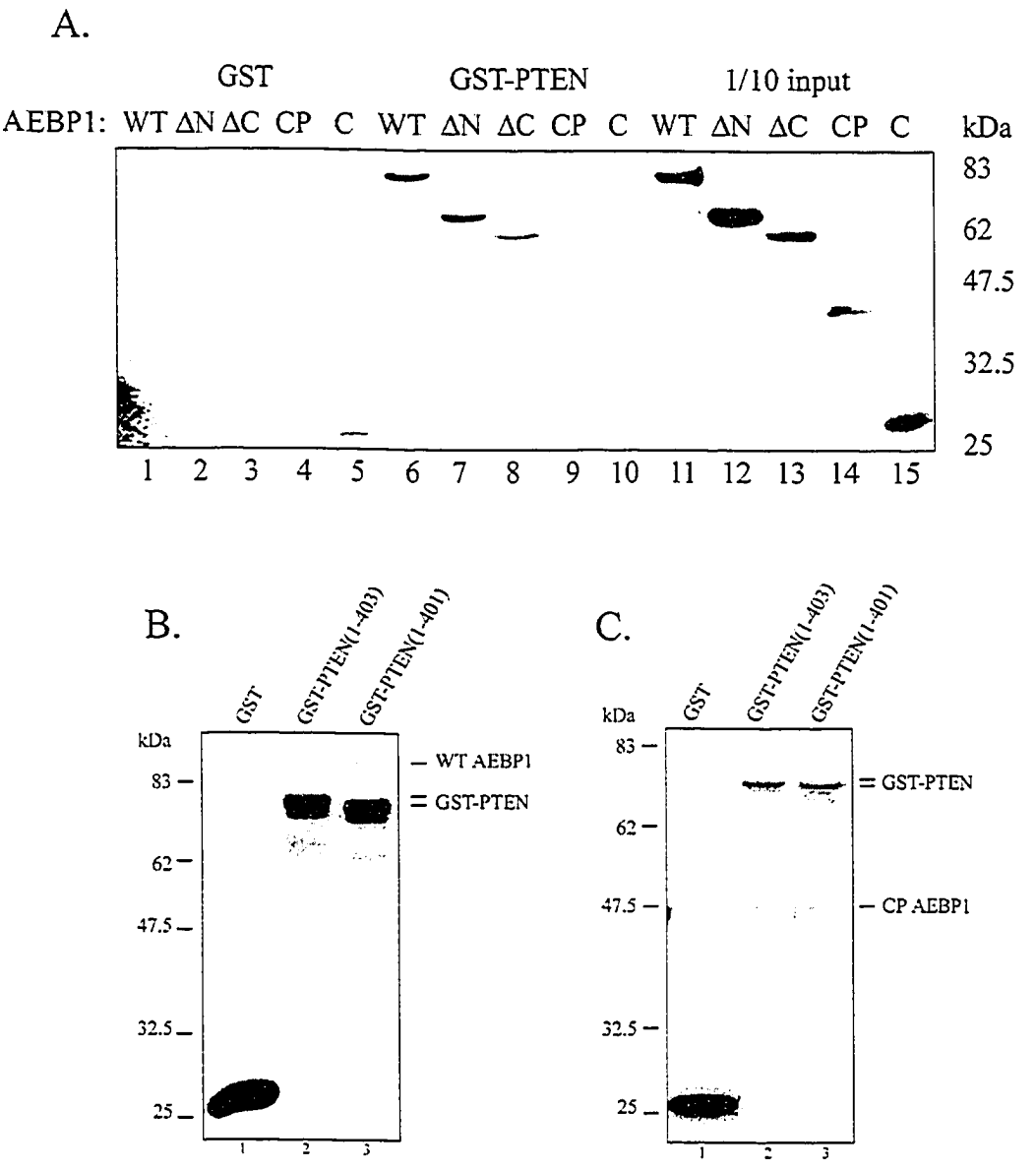


Figure 23. AEBP1 interacts with PTEN through its CP domain. (A) GST (lanes 1-5) or GST-PTEN (lanes 6-10) were incubated with 5 μg wild-type or various deletions of AEBP1 (lanes 11-15) and affinity purified using glutathione-Sepharose beads. Proteins were resolved by SDS-PAGE followed by immunoblotting with an anti-AEBP1 antibody. GST (lane 1), GST-PTEN (lane 2), or GST-PTEN 1-401 (lane 3) were incubated with (B) AEBP1-WT or (C) AEBP1-CP and affinity purified as in (A). Proteins were resolved by SDS-PAGE followed by protein staining with GelCode Blue.

interaction with PTEN is mediated by the CP domain, although it does not require the C-terminal dipeptide motif of PTEN.

To further investigate the possibility of a C-terminal dipeptide interaction function of the CP domain of AEBP1, the 12 C-terminal amino acids of different proteins were fused to the C-terminus of GST and expressed and purified from *E. coli.* To assess possible interactions of AEBP1 with specific C-terminal residues, the C-terminal residues of three proteins were chosen for analysis: 1) ERK1, which is known to interact with AEBP1 and contains a penultimate basic residue within its C-terminus (EETARFQPGY**R**S), 2) histone H3, because this histone also has a penultimate basic residue (IQLARRIRGE**R**A), and 3) NHE3, because this protein also has a cytoplasmic C-terminal basic penultimate residue (RPPAALPEST**H**M). Histone H3 was selected based on the observation that AEBP1 is a transcriptional repressor and interaction with DNA results in an increase in the detected CP activity (*164*). This suggests that the target for the CP domain may be a chromatin component, such as one of the histone proteins. Also, a protein named p85, which is part of a histone H3 arginine kinase complex has recently been identified by mass spectrometry to be AEBP1 (B. Wakim, unpublished, and Ref. (*101*), further suggesting that AEBP1 may interact with histone H3. NHE3 was chosen because reduction in NHE3 expression resulted in a phenotype similar to that seen in the reproductive tract of the male AEBP1 knockout mice, suggesting that AEBP1 might be able to interact with NHE3 and regulate its activity (*44*).

GST or GST terminating in the 12 C-terminal amino acids of histone H3, ERK1, or NHE3 were tested for their ability to interact with AEBP1 or its CP domain *in vitro* (Figure 24). In comparison to the amount of input AEBP1, only a small proportion of AEBP1 was brought down by all GST fusion proteins and by GST alone which suggested this was a non-specific interaction in all cases. Surprisingly, the GST-NHE3 fusion protein brought down even less AEBP1 than GST alone, which raises the possibility that the AEBP1 interaction with GST is specific. A similar experiment using WT AEBP1 in which additional wash steps were added eliminated the GST interaction (data not shown). Wild-type AEBP1 as well as the CP domain of AEBP1 only were consistently brought down by the GST-H3 fusion proteins, while GST-ERK1 showed a weaker interaction and GST-NHE3 consistently showed no specific interaction (Figure 24, data not shown). These *in vitro* interactions taken together show a clear

preference for specific C-terminal peptides by AEBP1, with the strongest interaction being with the C-terminus of histone H3. Additional experiments in which the penultimate arginine of the GST-H3 protein was mutated to an alanine did not reduce the level of interaction with AEBP1 (results not shown), suggesting that this penultimate arginine is not required for the interaction. While this suggests that the CP domain may not mediate C-terminal dipeptide interaction function, a specific interaction with the C-terminus of histone H3 was displayed. Other residues within this 12-amino acid stretch may be involved in an AEBP1-histone H3 interaction.

3.3.2 AEBP1 interacts with Ca^{2+}/calmodulin in a Ca^{2+}/calmodulin-dependent histone H3 arginine kinase complex.

AEBP1 was identified as the p85 component of a Ca^{2+}/calmodulin-dependent histone H3 arginine kinase complex through tryptic peptide fingerprinting of p85 by MALDI-TOF mass spectrometry (B. Wakim, unpublished results). p85 is the major component of this arginine kinase complex as well as the only component of this complex capable of interacting with Ca^{2+}/calmodulin (*101*). To confirm that AEBP1 is actually p85, Dr. Wakim performed a western blot using his p85 antibody on two versions of N-terminally His6-tagged AEBP1 that I purified, AEBP1 (27-730) and AEBP1 (123-748). Both versions of AEBP1 were recognized by the p85 antibody (Figure 25). Several contaminating proteins which were copurified with AEBP1 and present in the AEBP1 samples were not recognized by the p85 antibody (results not shown). This, together with the mass spectrometry analysis, suggests that AEBP1 and p85 are one and the same.

p85 has been shown to interact with Ca^{2+}/calmodulin (*101*). To confirm the identification of AEBP1 as p85, the interaction of AEBP1 and its deletion derivatives with Ca^{2+}/calmodulin was characterized by calmodulin-agarose pulldown assays (Figure 26). Full-length AEBP1 as well as the AEBP1ΔN and AEBP1ΔSTP derivatives showed strong binding to Ca^{2+}/calmodulin that could be abolished by inclusion of EGTA. AEBP1 CP and ΔC both lack the entire C-terminus and both showed only weak association with Ca^{2+}/calmodulin that was not sensitive to the presence of EGTA indicating that it was a non-specific association. However, AEBP1 ΔSTP, which contains the basic region of the C-terminus, was able to

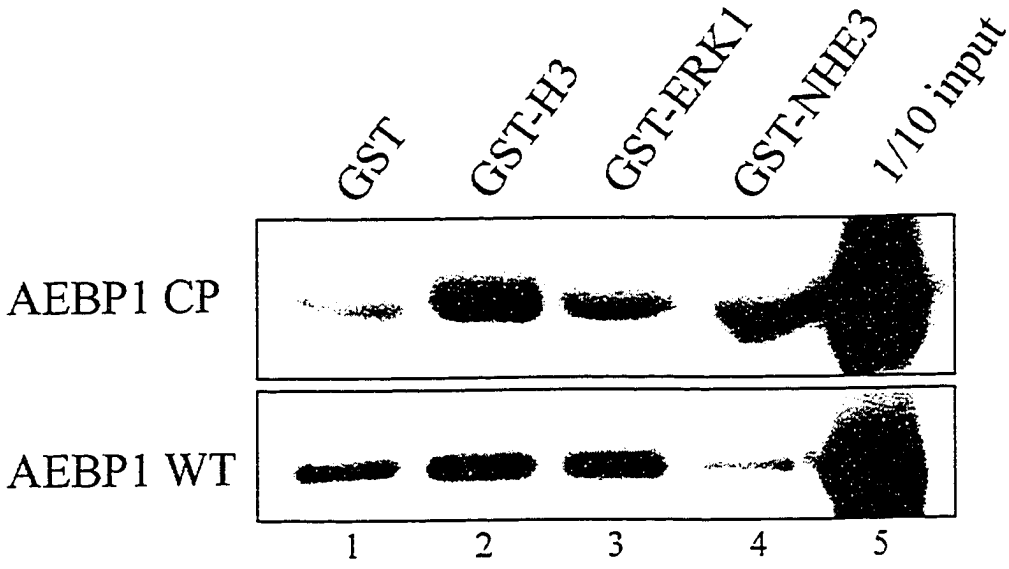


Figure 24. AEBP1 interacts with the C-terminus of histone H3. GST pulldown assays were performed with His-tagged AEBP1 CP and WT protein and GST fusion proteins in which the 12 C-terminal amino acids of a protein (histone H3 (lane 2), extracellular-regulated kinase 1 (ERK1, lane 3), and sodium-hydrogen exchanger 3 (NHE3, lane 4)) were fused to the C-terminus of GST. One-tenth of the input AEBP1 was run in lane 5. Proteins were resolved by SDS-PAGE and immunoblotted with an anti-AEBP1 antibody.

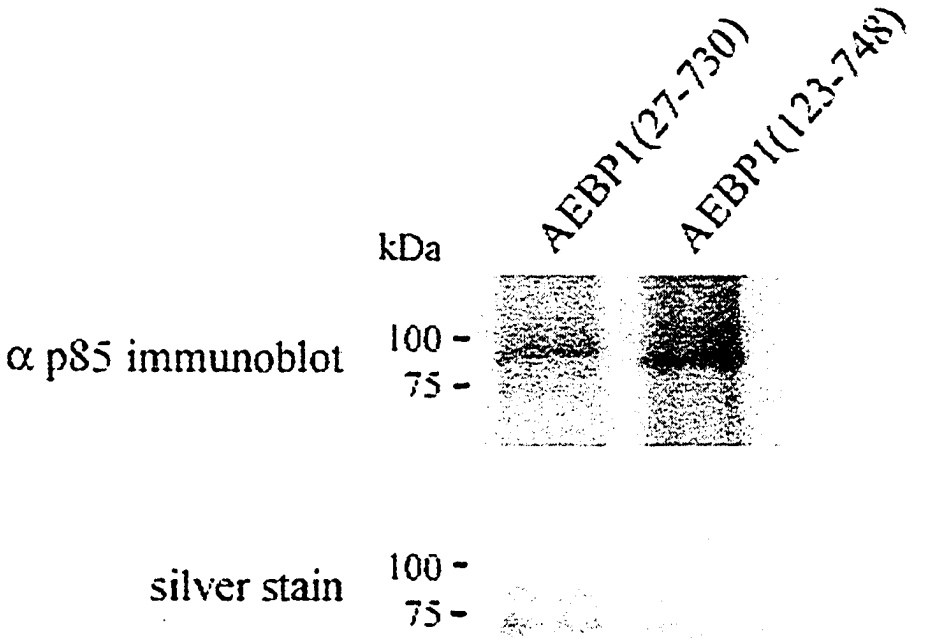


Figure 25. AEBP1 is the recognized by an anti-p85 antibody. N-terminally His6-tagged versions of AEBP1, AEBP1(27-730) and AEBP1(123-748), were resolved by SDS-PAGE and analyzed by silver staining and immunoblotting with an anti-p85 antibody followed by detection by alkaline phosphatase.

interact with Ca^{2+}/calmodulin very strongly, considering that the amount of AEBP1 ΔSTP input into this pulldown was much less than that used in Ca^{2+}/calmodulin pulldowns of AEBP1 WT and other AEBP1 truncations. Multiple lower molecular weight bands were also observed in all pulldown lanes following protein staining. These proteins appear to arise from the calmodulin-agarose beads, either calmodulin itself or other co-purified proteins. The ability of AEBP1ΔN and AEBP1ΔSTP to interact with Ca^{2+}/calmodulin suggests that neither the DLD domain nor the STP and acidic regions within the C-terminus of AEBP1 are necessary for interaction with Ca^{2+}/calmodulin. The lack of a strong association of Ca^{2+}/calmodulin with AEBP1ΔC suggests that the C-terminus of AEBP1 is necessary for interaction. The only part of the C-terminus that is present in AEBP1ΔSTP and lacking in AEBP1ΔC is the basic region, suggesting that this region is necessary for interaction of AEBP1 with Ca^{2+}/calmodulin.

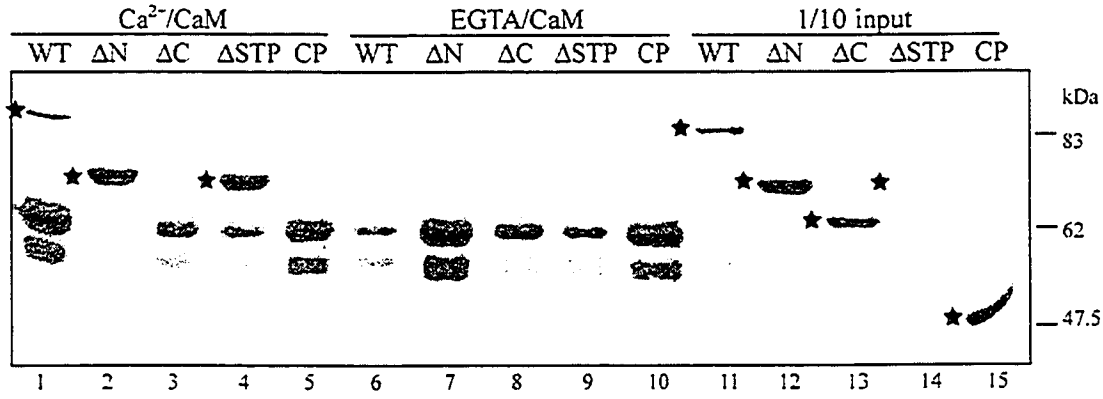


Figure 26. AEBP1 interacts with Ca^{2+}/calmodulin through its basic region. Calmodulin-sepharose was used to pull down AEBP1 and indicated truncation forms of AEBP1 in the presence of Ca^{2+} (lane 1-5) or EGTA (lanes 6-10). Proteins pulled down were separated by SDS-PAGE and stained with GelCode Blue. A star indicates strongly bound AEBP1 as well as input AEBP1 (lanes 11-15). Protein markers were loaded in lane 14 along with AEBP1ΔSTP protein.

3.4 Phosphorylation of AEBP1

ERK1/2 MAP kinase and AEBP1 have a "symbiotic" relationship. Not only does interaction of MAP kinase with AEBP1 protect MAP kinase from dephosphorylation, it also enhances the DNA binding ability of AEBP1 and leads to phosphorylation of AEBP1 *in vitro* (*7, 8*). This phosphorylation does not occur when the C-terminus of AEBP1 is truncated (*7*), indicating that either the phosphorylation site is in the C-terminus or that this truncation causes a conformational change in AEBP1 that prevents phosphorylation from occuring. Previous studies have shown a slight decrease in mobility of AEBP1 in nuclear extracts from 3T3-L1 cells, suggesting that phosphorylation of AEBP1 might lead to its nuclear translocation (*3*). Therefore I wanted to examine the nature and function of this phosphorylation both *in vivo* and *in vitro*.

3.4.1 AEBP1 is phosphorylated by MAP kinase in CHO cells upon EGF stimulation.

The MAP kinase pathway is activated by many upstream effectors (*179*). Because these cytokines and growth factors can be found in undefined serum, serum stimulation is often used to activate MAP kinase. I chose to use a more defined method of MAP kinase activation by stimulating cells with commercially available epidermal growth factor (EGF). Chinese Hamster Ovary (CHO) cells do not express EGF receptor (EGFR) and therefore it was necessary to co-transfect EGFR- along with AEBP1-expressing plasmids into these cells. Following transfection and serum starvation, cells were stimulated with EGF and harvested at various time points to determine the optimal stimulation time (Figure 27A). When extracts from EGF-stimulated cells were separated on a low percentage SDS-PAGE gel, a shift in the mobility of AEBP1 was observed. Approximately half of the total AEBP1 was detected as a higher molecular weight species consistent with this being a phosphorylated form of AEBP1. The percentage of AEBP1 in the upper band was greatest between 5 and 15 minutes of stimulation, although this level did not change to any great extent across the range of time

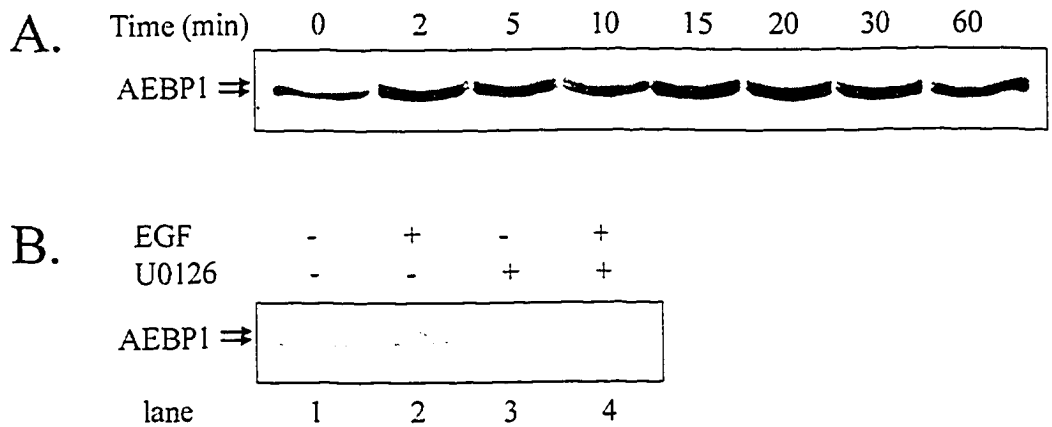


Figure 27. AEBP1 is phosphorylated by ERK1/2 MAP kinase. CHO cells were co-transfected with plasmids expressing AEBP1 (pcDNA-AEBP1) and EGFR (Rc/CMV-EGFR). Two days later cells were starved in serum-free media for 3 hours followed by stimulation for the indicated times with 10 ng/ml EGF (A). In panel B, cells were stimulated with EGF for 10 minutes and 10 μM U0126 was added to the cells during the starvation/stimulation periods for the indicated samples. Extracts were separated by SDS-PAGE and western blotted with an anti-AEBP1 antibody.

points tested, from 2 minutes to one hour EGF stimulation. This timing is similar to that reported for maximal activation of ERK1/2 by serum stimulation (S.W. Kim, unpublished), and suggests that this phosphorylation of AEBP1 is relatively stable or that AEBP1 is repeatedly phosphorylated and dephosphorylated over the time period of one hour.

To confirm that the mobility shift observed for AEBP1 upon EGF stimulation was indeed caused by phosphorylation by ERK1/2 MAP kinases, cells were treated with a MAP kinase kinase (MEK) inhibitor, U0126 (Figure 27B). Inhibiting the phosphorylation and activation of MEK would be expected to eliminate activation of its downstream kinase substrates, the ERK1/2 MAP kinases, and hence the phosphorylation of MAP kinase substrates. As in the previous experiment, stimulating cells with EGF in the absence of U0126 produced the expected shift in mobility of AEBP1 (lane 2). However, when U0126 was added, no shift in mobility was observed upon EGF stimulation (lane 4). The ability of a MEK inhibitor to abolish the mobility shift observed in AEBP1 upon EGF stimulation suggests that the higher molecular weight species arises by ERK1/2 phosphorylation of AEBP1.

3.4.2 AEBP1 S668 mutation eliminates phosphorylation of AEBP1 by MAP kinase *in vivo*.

Sequence analysis of mouse AEBP1 revealed several potential MAP kinase phosphorylation sites based on the consensus motif PXS/TP (Figure 28). Sites identified with highest confidence were T623, S658 and S668 in mouse AEBP1. The T623 MAP kinase consensus phosphorylation motif is the only putative phosphorylation site conserved in all mammalian AEBP1 orthologues analyzed. In fact, T623 is the only putative MAP kinase phosphorylation site present in the entire human AEBP1 sequence. Of the six orthologues shown in Figure 28, the mouse S658 and S668 phosphorylation motifs are conserved only in rat and cow AEBP1 and not in the human sequence.

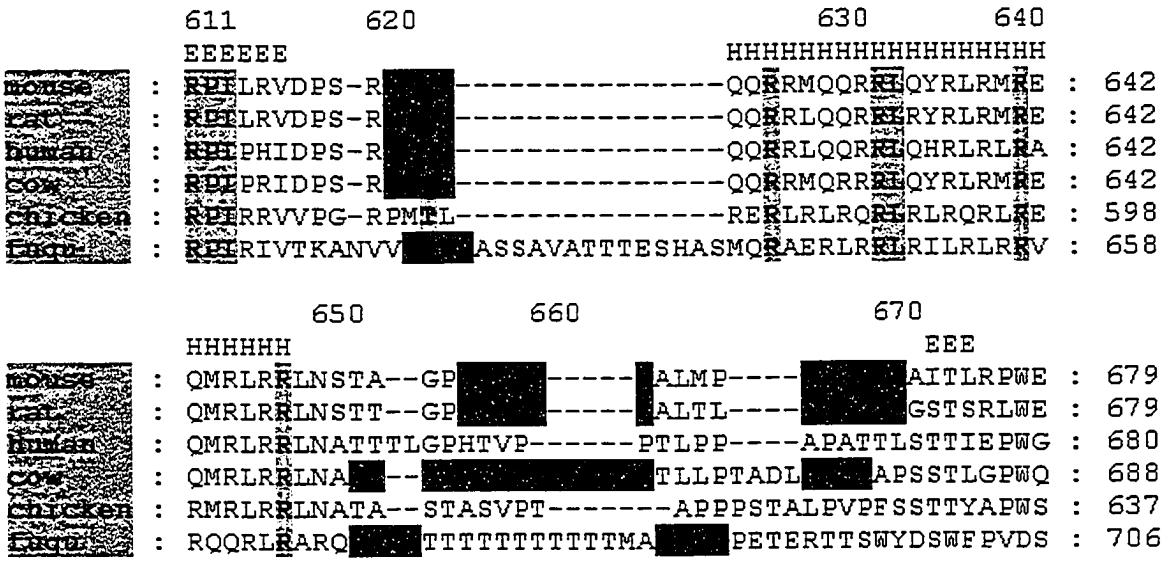


Figure 28. MAP kinase consensus phosphorylation sites within the C-terminus of AEBP1. A multiple alignment of a portion of the C-terminus of AEBP1 is shown. The numbering at the top indicates amino acid numbering of mouse AEBP1. The second row contains a secondary structure prediction obtained from the program PhDsec, in which E = β-strand and H = α-helical secondary structure. Conserved residues are coloured orange, while potential MAP kinase phosphorylation sites are coloured red.

Site-directed mutagenesis was performed on these potential phosphorylation sites to mutate each to both an alanine and an aspartate, mimicking the dephosphorylated and phosphorylated states, respectively. These phosphorylation site mutants of AEBP1 were expressed from a mammalian expression vector in CHO cells. Upon stimulation of the cells with EGF, wild-type AEBP1 along with T623A and S658A AEBP1 mutants all exhibited a mobility shift when analyzed by SDS-PAGE and immunoblotting (Figure 29A). Interestingly, S668A AEBP1 did not show a mobility shift, demonstrating that S668 in mouse AEBP1 is critical for phosphorylation of AEBP1 by ERK1/2. The loss of the mobility shift suggests that S668 may itself be the residue phosphorylated by ERK1/2 upon EGF stimulation, as mutation of the other two potential phosphorylation sites did not affect this shift in mobility of AEBP1. If S668 is itself phosphorylated, the phosphorylation mimetic mutation, S668D, might be expected to migrate with a mobility similar to the higher molecular weight form of AEBP1 observed when cells were stimulated with EGF. If S668 is not phosphorylated, but only necessary for the phosphorylation of another residue, we would expect mutation to an aspartate to abolish the mobility shift. The S668D mutant form of AEBP1 was tested to determine which of these hypotheses was correct (Figure 29B). The S668D AEBP1 exhibited a constitutive mobility shift. Although not quite as large a shift as that caused by phosphorylation, the altered mobility further supports the hypothesis that S668 in mouse AEBP1 is phosphorylated by ERK1/2.

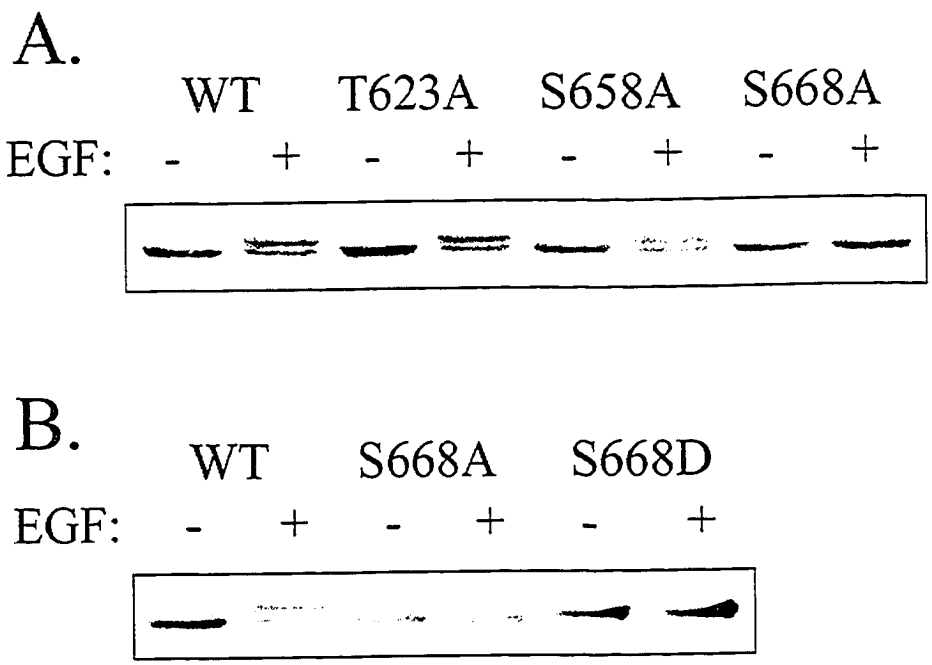


Figure 29. S668 is necessary for phosphorylation of AEBP1. CHO cells were co-transfected with plasmids expressing EGFR (Rc/CMV-EGFR) and either wild-type (WT) or the indicated mutants of AEBP1. For expression of AEBP1, either pJ3H-AEBP1 (panel A) or pcDNA-AEBP1-Myc/His (panel B) and mutant derivative expression plasmids were used. Two days later cells were starved in serum-free media for 3 hours followed by stimulation for the indicated times with 10 ng/ml EGF. Extracts were separated by SDS-PAGE and immunoblotted with an anti-AEBP1 antibody.

3.4.3 AEBP1 S668 mutation eliminates phosphorylation by MAP kinase *in vitro*.

MEK inhibitor studies shown above suggest that ERK1/2 is the kinase responsible for phosphorylation of AEBP1. Stimulation of cells with EGF resulted in a mobility shift of AEBP1 consistent with phosphorylation, which was eliminated upon mutation of S668 to alanine. To further characterize this phosphorylation of AEBP1, an *in vitro* kinase assay was used to test whether recombinant ERK2 could phosphorylate WT AEBP1 and the phosphorylation site mutants of AEBP1 (Figure 30). No phosphorylation was detected when a kinase assay was performed using S668A AEBP1. Phosphorylation of the T623 and S658 AEBP1 mutants was slightly reduced relative to wild type. Kinase assays using equivalent amounts of AEBP1 27-730 (Figure 30) also showed efficient incorporation of radiolabel. As seen for full-length AEBP1, mutation of T623 to alanine within AEBP1 27-730 also reduced phosphorylation, by approximately half that of AEBP1 27-730 (compare lanes 5 and 6). Kinase assays performed with AEBP1ΔDLD exhibited even less phosphorylation (lane 7), possibly due to a conformational change which prevents interaction with MAP kinase. Altogether, kinase assays suggest that S668 of AEBP1 is the major MAP kinase phosphorylation site, as this mutation eliminates all phosphorylation. These *in vitro* kinase assays also suggest that T623 and S658 may be phosphorylated by MAP kinase. However, mutation of T623 or S658 did not cause any noticeable change in the mobility shift of AEBP1 observed upon EGF stimulation of cells. It could be that phosphorylation of T623 and/or S658 does not cause a mobility shift or that phosphorylation of these residues may be dependent on phosphorylation of S668 as the master regulator. However, kinase assays performed with AEBP1 S668D do not show any increase in phosphorylation when compared with AEBP1 S668A (data not shown). This does not support a role for S668 phosphorylation as a master regulator of phosphorylation, unless the aspartate mutation is not able to mimic the physiological role of S668 phosphorylation, but supports the hypothesis that S668 is the major phosphorylation site within AEBP1.

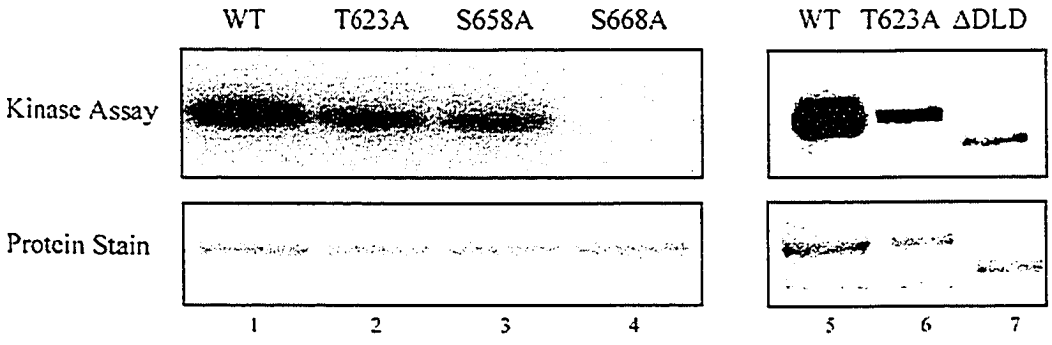


Figure 30. S668A mutation of AEBP1 eliminates all phosphorylation by ERK2 *in vitro*. Kinase assays were performed by incubating 0.5 μg (lanes 1-4) or 1.0 μg (lanes 5-7) of *E. coli* produced AEBP1 and indicated mutant AEBP1 proteins with ERK2 (NEB) and [γ-^{32}P]ATP. The samples were then separated by SDS-PAGE and visualized by autoradiography (top) and gelcode blue (Pierce) staining (bottom). AEBP1 was either full-length (lanes 1-4), amino acids 27-730 (lanes 5 and 6), or amino acids 123-748 (lane 7).

A comparison was made of the ability of MAP kinase to phosphorylate different truncation versions of AEBP1. Equal amounts of WT AEBP1 and ΔN, ΔSTP, and C truncated versions of AEBP1 were compared for their ability to be phosphorylated *in vitro* (Figure 31). No phosphorylation of ΔSTP AEBP1 was observed, suggesting that the phosphorylation site was no longer contained in this deleted version of AEBP1. The ΔSTP truncation still contains T623, suggesting that this is not the site of phosphorylation. Alternatively, deletion of the STP and acidic regions of AEBP1 may eliminate the ability of AEBP1 to interact with MAP kinase. Phosphorylation of ΔN AEBP1 by MAP kinase was approximately 1.8 fold greater than phosphorylation of WT AEBP1. Interestingly, AEBP1 C, which consists of only the C-terminal domain of AEBP1, is able to be phosphorylated by MAP kinase to an extent equivalent to that of ΔN AEBP1. These results suggest that the C-terminus of AEBP1 is sufficient for phosphorylation by MAP kinase *in vitro*, and that deletion of the N-terminal domain of AEBP1 may make the C-terminal domain more accessible to the kinase.

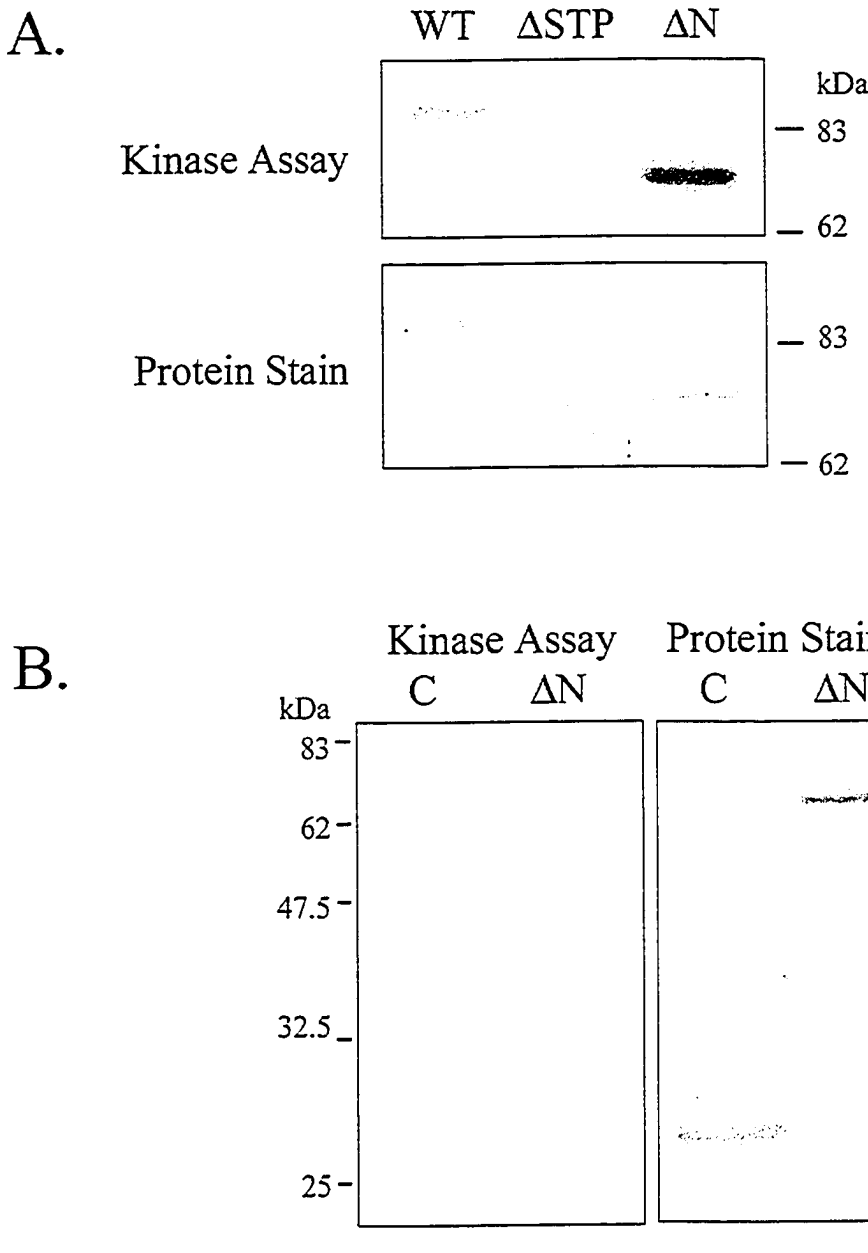


Figure 31. Truncation of AEBP1 affects *in vitro* phosphorylation by MAP kinase. Kinase assays were performed by incubating 0.5 μg of *E. coli* produced AEBP1 and indicated mutant AEBP1 proteins with ERK2 (NEB) and [γ-^{32}P]ATP. The samples were then separated by SDS-PAGE and visualized by autoradiography and Gelcode Blue staining.

3.4.4 Function of AEBP1 S668 phosphorylation by MAP kinase.

The most well characterized function of AEBP1 is that of a transcriptional repressor. AEBP1 translocates to the nucleus where it binds a specific DNA sequence and mediates its transcriptional repression through its CP domain (*2*). The details regarding how AEBP1 translocates to the nucleus, how it binds DNA, and what proteins it interacts with in order to repress transcription have not yet been fully elucidated. Phosphorylation of AEBP1 may regulate one or more of these functions of AEBP1. In fact, a shift in mobility has been reported in nuclear AEBP1 from 3T3-L1 cells (*3*). The same study also reported approximately equal amounts of endogenous AEBP1 in nuclear and cytoplasmic compartments of 3T3-L1 cells (*3*). To determine if phosphorylation of AEBP1 controls the nuclear/cytoplasmic shuttling of AEBP1, CHO cells were transfected with plasmids expressing WT AEBP1 and S668A and S668D mutant versions of AEBP1 and nuclear and cytoplasmic fractions were isolated from the transfected cells. Immunoblotting for AEBP1 revealed that at least 80% of AEBP1 was present in the cytoplasmic fraction (results not shown). Mutation of AEBP1 S668 to either alanine or aspartate did not cause any difference in the distribution of AEBP1 when compared to wild type AEBP1 (preliminary results). Since the S668 mutants did not have an altered distribution between the nucleus and cytoplasm, this suggests that phosphorylation does not regulate the nuclear translocation of AEBP1 or at least is not critical for this translocation. It could be that phosphorylation of AEBP1 by MAP kinase regulates the subcellular localization of AEBP1 to other compartments and organelles within the cell which would not be detected by the relatively crude fractionation performed here.

The binding of AEBP1 to DNA and subsequent transcriptional repression could also be affected by phosphorylation. However, luciferase assays analyzing the repression ability of all phosphorylation site mutants, both alanine and aspartate, in comparison to WT AEBP1, did not detect any significant difference between them (results not shown). Because a change in transcriptional repression ability would naturally follow any change in specific DNA binding, the ability of phosphorylation site mutants to bind DNA was not analyzed. Therefore, the function, if any, of this phosphorylation of AEBP1 by MAP kinase remains to be determined.

3.5 DNA binding and Transcriptional Repression by AEBP1.

AEBP1 is a transcriptional repressor that interacts with the AE-1 site within the aP2 gene promoter (*2*). Previous electrophoretic mobility shift assay (EMSA) studies on AEBP1 showed that a C-terminal deletion mutant of AEBP1, AEBP1ΔSty, which lacks the C-terminal 205 amino acids, was unable to bind DNA (Muise and Ro, unpublished results). This suggested that the DNA binding domain of AEBP1 was located in the C-terminal domain, or that this C-terminal deletion caused a conformational change in other domains of the protein necessary for DNA binding. The C-terminus of AEBP1 contains a basic region, rich in arginine, which is predicted to have α-helical secondary structure similar to that found in the DNA binding domains of the basic leucine zipper proteins such as CAAT/enhancer binding protein (C/EBP(*180*), Figure 32), and which may also be necessary for DNA binding by AEBP1.

3.5.1 Both N- and C-termini of AEBP1 are necessary for DNA binding.

In order to better characterize the DNA-binding domains of AEBP1, electrophoretic mobility shift assays (EMSA) were performed in which the ability of N-terminally His6 tagged versions of AEBP1 to bind a ^{32}P-radiolabeled duplex oligonucleotide containing an AEBP1 binding site, AE-1, were assayed. The starting point for these comparisons was AEBP1 protein C-terminally truncated by 18 amino acids and N-terminally truncated by 26 amino acids (AEBP1 27-730), which had been used in all prior studies (*2, 3, 164*). The DNA binding of AEBP1 (27-730) was compared with a version of AEBP1 N-terminally truncated by 122 amino acids. (AEBP1(123-748)). No DNA binding was detected by the AEBP1(123-748) truncation mutant (Figure 33A, lanes 4-6). This lack of binding suggests that an N-terminal domain of AEBP1 is involved in DNA binding, or at least is necessary to retain the proper conformation needed for DNA binding. It could also be that the presence of the C-terminal 18 amino acids of AEBP1 in AEBP1(123-748) prevents DNA binding from occurring.

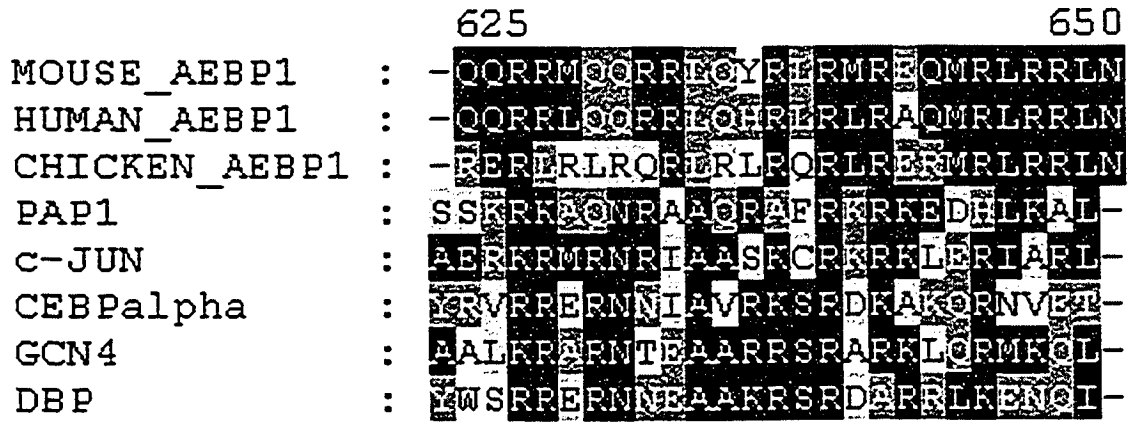
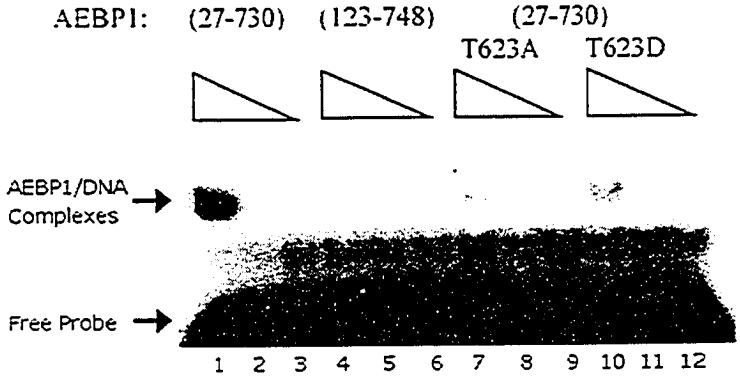


Figure 32. The basic domain of AEBP1 can be compared with that of the basic leucine zipper proteins. An alignment of the DNA binding basic regions of representative proteins of the bZIP family with that of AEBP1. This figure is adapted from Ref. (180). The greyscale shading from black to white indicates progressively lower amino acid identity/similarity.

A.



B.

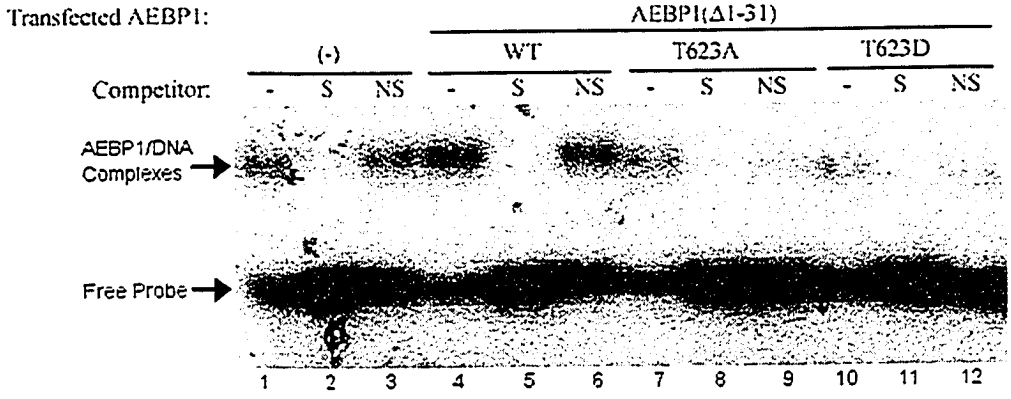


Figure 33. Mutation of AEBP1 threonine 623 results in decreased DNA binding ability. (A) EMSA was performed by incubating 0.15, 0.3 and 0.6 μg of the indicated recombinant versions of the AEBP1 protein with radiolabelled AE-1 duplex oligonucleotide and separating complexes on a non-denaturing gel followed by autoradiography. (B) NIH/3T3 cells were transfected with plasmids expressing the indicated versions of N-terminally HA-tagged AEBP1 or with the AEBP1 ORF inserted in the wrong orientation for expression (-). Nuclear extracts from these cells were incubated with radiolabelled AE-1 duplex oligonucleotide, as well as 50-fold excess specific competitor (lanes 2, 5, 8, and 11) or non-specific competitor (lanes 3, 6, 9, and 12), resolved on a non-denaturing gel, and visualized by autoradiography.

3.5.2 DNA binding activities of AEBP1 are inhibited by MAP kinase phosphorylation site mutation.

The T623 AEBP1 phosphorylation site, the only putative MAP kinase phosphorylation site conserved in mammalian AEBP1, is just N-terminal to the basic helix found in the C-terminus of AEBP1 (Figure 28) which may be important for DNA binding. To test whether phosphorylation of this site might affect the ability of AEBP1 to bind DNA, wild-type and mutant versions of AEBP1 were assayed by EMSA for their ability to bind an oligonucleotide containing the AE-1 binding site. Both the T623A and T623D mutants displayed a greatly reduced DNA binding ability (Figure 33A, lanes 7-9 and 10-12, respectively). This suggests that threonine 623 is required for strong DNA binding by AEBP1. The mutation of this threonine to alanine or aspartate eliminates the hydroxyl group of the threonine that may directly participate in hydrogen bonding with the DNA. Phosphorylation of this threonine might also eliminate this functional group necessary for strong DNA interaction.

In order to see the effect of this mutation on DNA binding *in vivo*, EMSA was performed with nuclear extracts isolated from NIH/3T3 cells transfected with plasmids encoding Δ1-31 and derivative phosphorylation site mutant versions of AEBP1 (Figure 33B). A higher molecular weight band was observed in all lanes where nuclear protein extract was incubated with radiolabeled AE-1 DNA in the absence of unlabeled competitor oligonucleotide. This band disappeared when a specific competitor (S, unlabelled AE-1 DNA) was included but remained in the presence of a non-specific (NS) DNA competitor. The band therefore represented a specific DNA-protein complex and, in the lanes where extracts contained no transfected AEBP1 expression, was assumed to be a complex of endogenous AEBP1 and AE-1 DNA probe. As expected, this complex increased in intensity upon incubation with AEBP1-transfected nuclear extracts. However, it decreased even below endogenous levels upon incubation with T623A and T623D AEBP1-transfected nuclear extracts. In previous EMSA experiments, AEBP1 has been shown to multimerize upon increase in salt concentrations (A Muise, H Ro, unpublished results). The reduced complex formation from the T623A and T623D AEBP1 extracts may result from a requirement for dimerization in order for AEBP1 to bind DNA. The mutant AEBP1, which is unable to bind

DNA strongly, may titrate endogenous AEBP1 away from the DNA probe through a dimerization interaction. Experiments to verify dimerization of AEBP1 through immunoprecipitation have been unsuccessful, possibly because AEBP1 was unable to dimerize in the absence of DNA and/or because of the difficulty in optimizing the conditions for dimerization, DNA interaction, and immunoprecipitation in one reaction.

3.5.3 Transcriptional repression abilities of AEBP1 are inhibited by MAP kinase phosphorylation site mutation in a cell-type specific manner.

The diminished DNA binding activity of mutant derivatives of AEBP1 led me to examine their transcriptional repressor activity. NIH/3T3 and CHO cells were transiently co-transfected with plasmids expressing AEBP1 (Δ1-31) or derivative mutant forms along with a reporter expressing the chloramphenicol acetyl transferase (CAT) gene under the control of the aP2 promoter and a plasmid expressing β-galactosidase under the control of the CMV promoter. The promoter contained 3 copies of the AE-1 binding site, previously characterized as a site to which AEBP1 specifically binds (*2*), upstream of the promoter region of the aP2 gene (*181*). CAT assays were performed to assess the ability of the different forms of AEBP1 to repress transcription of the reporter gene, and normalized for transfection efficiency using results from a β-galactosidase assay (Figure 34). In NIH/3T3 cells, as expected, the decreased DNA binding ability of the T623A and T623D AEBP1 mutants resulted in the elimination of transcriptional repression ability (Figure 34A). Analysis of the localization of AEBP1 showed that mutation of T623 did not affect its nuclear localization or its expression level (data not shown). These results suggested that the transcriptional activity of AEBP1 might be regulated by phosphorylation of threonine 623.

When the assay was repeated in CHO cells there was no apparent effect of mutation of T623 on transcriptional repression by AEBP1 (Figure 34B). Both T623A and T623D showed repression equivalent to wild type. In addition, transcriptional repression by AEBP1 in CHO cells was significantly greater than that detected in NIH/3T3 cells. To determine whether the cell-type specific difference in transcriptional repression by AEBP1 might reflect differences in the expression levels of the AEBP1 proteins in these cells, western blot analysis was

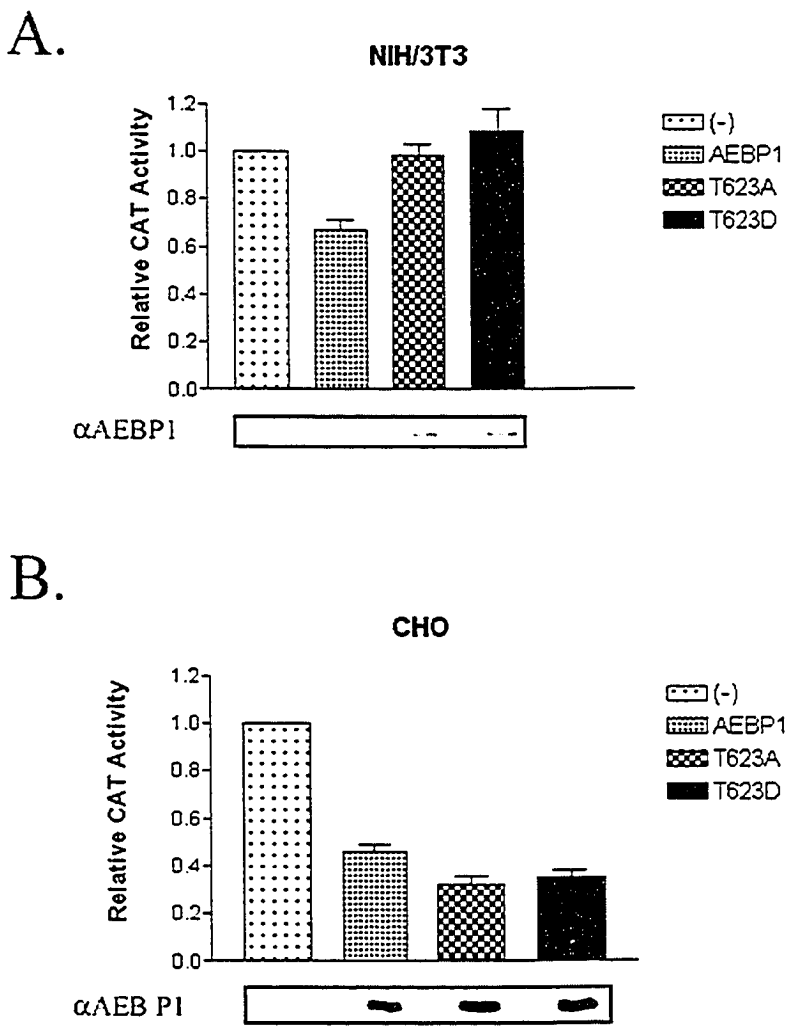


Figure 34. Mutation of AEBP1 threonine 623 results in decreased transcriptional repression ability in NIH/3T3 cells, but not in CHO cells. NIH/3T3 (A) and CHO (B) cells were co-transfected with pJ3H derivative plasmids expressing AEBP1 (Δ1-31) or the indicated mutants of this protein or with the control plasmid pJ3H-AEBP1(-) containing the AEBP1 ORF in the opposite orientation along with the reporter plasmids paP2(3AE-1/-120)CAT and pHermes-β-gal. Transcriptional activity was measured as relative levels of CAT activity, normalized to β-galactosidase activity for transfection efficiency. Data shown is the average (±SEM) of at least three different experiments performed in duplicate (n = 6). Below each histogram is a western blot analysis of AEBP1 levels in proteins extracted from transfected cells.

performed. Levels of AEBP1 in CHO cells were much higher than in NIH/3T3 cells. It could be that the high level of expression in CHO cells allows the mutant forms of AEBP1 to interact with the promoter at a level sufficient to provide repression despite not being as efficient at DNA binding. In order to test this hypothesis, CHO cells were transfected with progressively lower amounts of the plasmid expressing Δ1-31 or mutant forms of AEBP1 along with the reporter plasmid. The repression of reporter gene expression by T623A AEBP1 was no different than that exhibited by wild-type AEBP1 at all levels of protein expression (data not shown). T623D showed a trend towards less transcriptional repression ability, although only one of five data points (levels of T623D AEBP1 expression) exhibited statistically less transcriptional repression ability than wild-type AEBP1. Based on the ability of T623A AEBP1 to repress transcription even at low levels of expression, as well as most levels of T623D AEBP1 to repress transcription, it seems that the ability of AEBP1 to bind DNA strongly in order to repress transcription of the aP2 gene in CHO cells is not necessary even at low levels of AEBP1 expression.

3.5.4 AEBP1 may function as a co-repressor.

One possibility for the cell-type specific effects observed in CHO cells may be that AEBP1 can mediate transcriptional repression in CHO cells by acting as a co-repressor which interacts with another cell-specific factor. In fact, when AEBP1ΔSty, which lacks the C-terminal 205 amino acids of AEBP1 and was previously shown to have no DNA binding ability, was expressed in CHO cells and analyzed by luciferase assay, it was still able to repress transcription 2-fold, somewhat less than wild-type AEBP1 but still to a significant extent (Figure 35). This is similar to the ability of AEBP1 T623 mutants to repress transcription in CHO cells despite not being able to bind DNA strongly. The mechanism for co-repression by AEBP1 has not been determined. However, AEBP1 does possess an LXXLL motif (amino acids 245-249) which is commonly used in many co-repressors as an interaction motif.

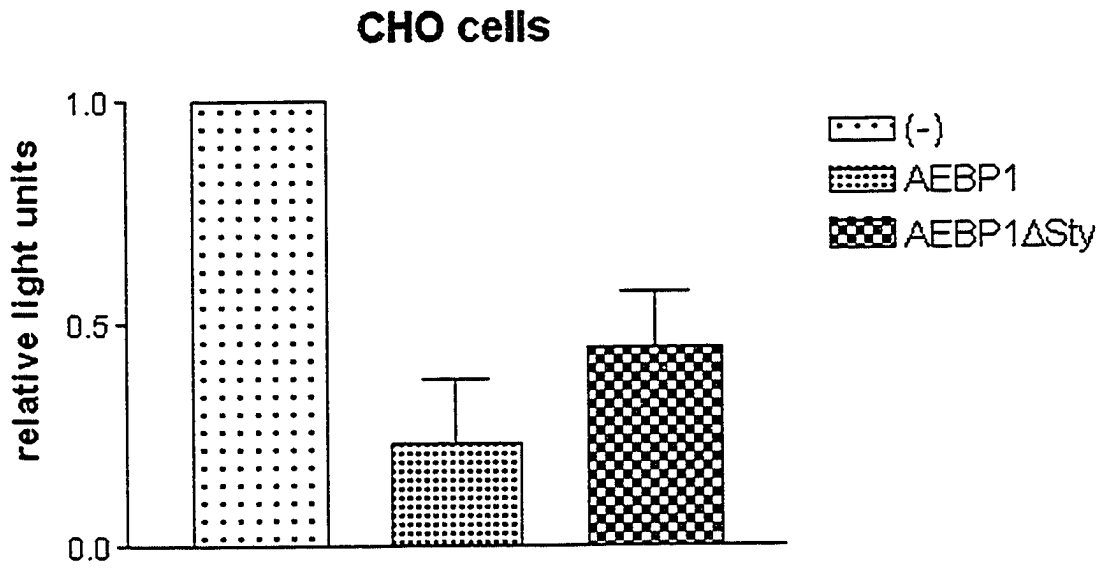


Figure 35. AEBP1ΔSty is able to repress transcription from the aP2 promoter in CHO cells. CHO cells were co-transfected with the luciferase reporter plasmid pGL2-aP2(-168/+21), a β-galactosidase-expressing plasmid, and pJ3H derivative plasmids expressing AEBP1 (amino acids 32-748), AEBP1ΔSty (amino acids 32-543), or a control plasmid (-) containing the AEBP1 ORF in the opposite orientation. Transcriptional activity was measured as relative levels of luciferase activity, normalized to β-galactosidase activity for transfection efficiency. Data shown is the average (±SD) of at least three different transfections (n = 6).

In order to better characterize the ability of AEBP1 to interact with AE-1 DNA oligonucleotide and the relationship of this DNA binding to transcriptional repression by AEBP1, full-length and domain-specific deletions of AEBP1 were used in electrophoretic mobility shift assays (Figure 36). A signal representing the radiolabeled AE-1 DNA duplex bound with AEBP1 protein or a deletion form of AEBP1 was observed for all proteins. WT AEBP1 consistently exhibited a very weak signal representing protein bound to a small fraction of total radiolabeled DNA. ΔN and ΔC (data not shown) AEBP1 also bound DNA weakly, although stronger than WT AEBP1. AEBP1(27-730) and ΔSTP AEBP1 were both much more efficient at DNA binding, with ΔSTP by far having the greatest ability to bind DNA. A second complex with lower mobility was also observed upon interaction of ΔSTP with AE-1 DNA, suggestive of protein dimerization. The weak binding of the ΔC deletion (results not shown) yet very strong binding of the ΔSTP deletion suggests that the primary DNA binding domain of AEBP1 is the basic region within the C-terminus. It appears that both the ΔSTP and the AEBP1(27-730) forms of AEBP1 lack regions that may normally regulate the ability of the protein to interact with DNA. It is possible that interaction of AEBP1 with another protein may cause conformational changes which would allow the putative basic helix-containing DNA-binding domain of AEBP1 to be available for DNA binding *in vivo*.

To determine if the widely varying ability of domain specific deletion mutants of AEBP1 to interact with DNA might correlate with ability to repress transcription, these proteins were expressed in CHO cells and analyzed for their ability to repress transcription of a luciferase reporter gene controlled by the aP2 promoter (Figure 37). A β-galactosidase-expressing plasmid driven by the CMV promoter was also co-transfected as an internal control for transfection efficiency and was expected to be unaffected by expression of AEBP1. When the luciferase assay data was normalized for transfection efficiency with β-galactosidase assay data, both ΔN and ΔC AEBP1 repressed transcription at least as well as, if not better than, WT AEBP1 (Figure 37A). However, it was unexpectedly found that β-galactosidase activity also decreased as cells were transfected with higher levels of plasmid expressing AEBP1 WT or AEBP1 ΔC (Figure 37C). This decrease was not as great as the reduction in aP2-regulated luciferase gene expression (Figure 37B), resulting in a net repression of the aP2 promoter upon normalization, although attenuated compared to the unnormalized luciferase data. ΔN AEBP1 had little effect on β-galactosidase activity, although the same

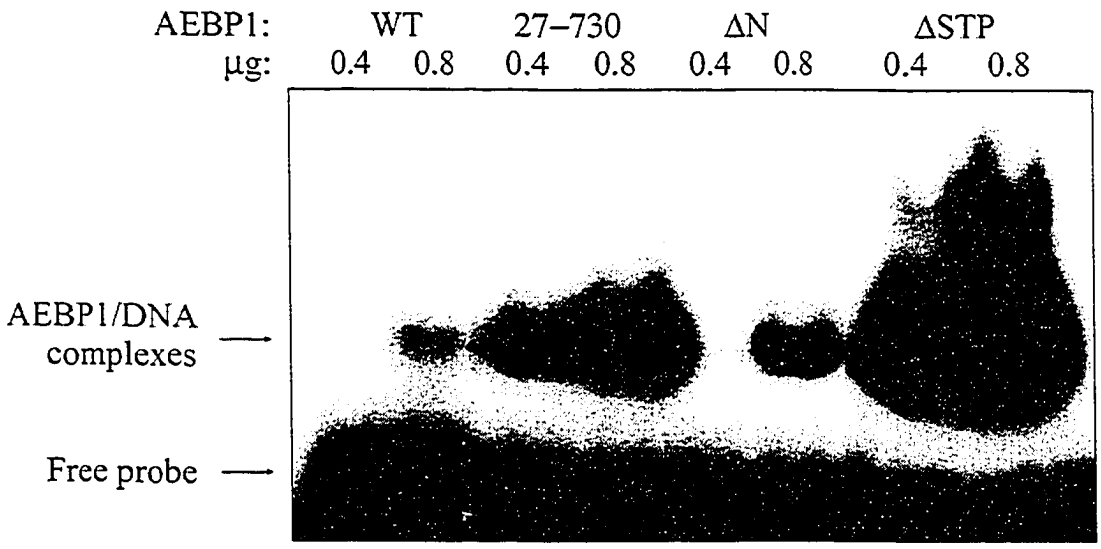
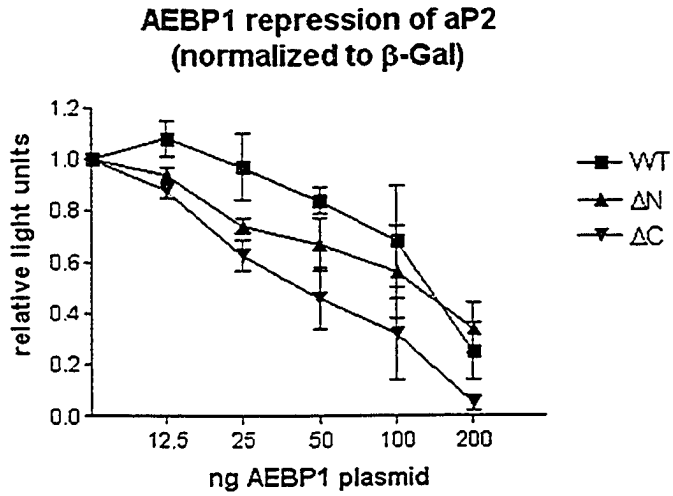


Figure 36. Wild-type AEBP1 binds DNA weakly through its basic region. Electrophoretic mobility shift assays (EMSA) were performed with radiolabeled AE-1 duplex oligonucleotide and the indicated amounts of *E. coli* expressed and purified AEBP1 or its deletion derivatives.

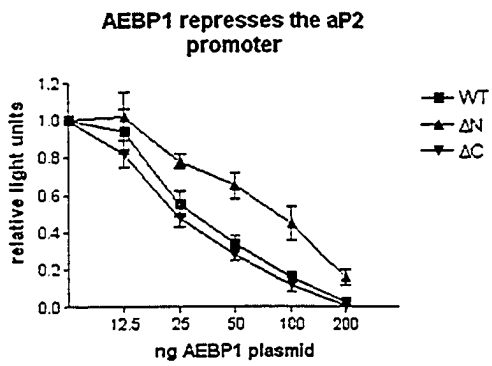
amount of DNA was used under the same conditions. ΔN AEBP1 also had a somewhat reduced effect on aP2-driven luciferase expression when compared to WT and ΔC AEBP1. The dose-dependent decrease in β-galactosidase activity only when increasing amounts of WT AEBP1 or ΔC AEBP1 were expressed suggests that this reduction is due to repression by AEBP1 and not merely due to a decrease in transfection efficiency. The absence of any effect on β-galactosidase expression upon deletion of the amino terminal portion of AEBP1 suggests that this deletion may result in impaired recognition of the CMV promoter which drives expression of the LacZ gene on the expression plasmid, while still allowing regulation of the aP2 promoter. It appears that the N-terminal DLD domain of AEBP1 may be important for transcriptional repression in a promoter-dependent manner.

In contrast to the effects of ΔN AEBP1, WT and ΔC AEBP1 were able to reduce expression from both promoters. Similar results were obtained when empty expression vector was transfected as a control and upon transfection of plasmid from different plasmid preparations, suggesting that the results are not due to a DNA effect or due to plasmid impurities affecting transfection efficiency. The ability of ΔC AEBP1 to repress transcription from both the aP2 and CMV promoters was essentially the same as the ability of WT AEBP1 to repress transcription from these promoters, even when expressing very low levels of protein. This suggests that AEBP1 may not interact with DNA through its C-terminus or that DNA binding through the C-terminus of AEBP1 may not play an important role in transcriptional repression in CHO cells. In fact, luciferase assays analyzing the ability of ΔSTP AEBP1 to repress transcription indicated repression activity similar to wild-type AEBP1 (data not shown), suggesting that the ability of AEBP1 or truncation forms of AEBP1 to interact with DNA *in vitro* may not correlate with their transcriptional repression ability *in vivo*.

A.



B.



C.

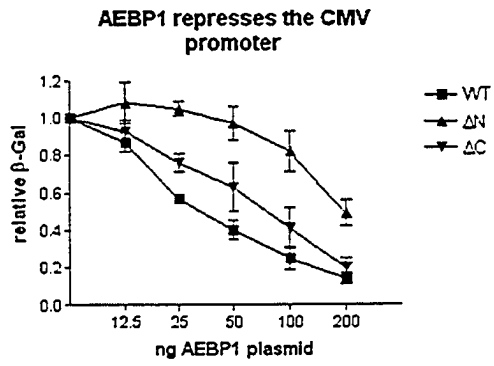


Figure 37. N- and C-terminally truncated AEBP1 repress transcription from the aP2 promoter. CHO cells were transfected with three plasmids, one containing the aP2-luciferase reporter gene, the second expressing β-galactosidase (pHermes-LacZ), and the third expressing AEBP1 WT or the ΔN or ΔC versions of AEBP1. The amount of the plasmid expressing AEBP1 or truncations was varied as indicated and varying amounts of pcDNA empty vector was added to ensure that the total amount of DNA was equal in all transfections. Cells were lysed and transcription of the reporter plasmid measured as light produced normalized to β-galactosidase activity (A). The unnormalized results from the luciferase assay (B) as well as the β-galactosidase assay (C) are also shown. Results are averages (±SD) from four transfections.

Chapter 4

DISCUSSION

4.1 Structure of AEBP1

AEBP1 contains three distinct domains. Recently, X-ray crystal structures of proteins with high sequence identity to AEBP1 have been solved, enabling me to construct three-dimensional structural models for two of the three domains of AEBP1. These models provide us with a structural framework on which to base our understanding of the functions of AEBP1. Further work would greatly benefit from an understanding of how these domains fit together and of how they might move with repect to each other in order to regulate the activity of AEBP1. It is difficult to make any absolute statements as to the relative orientations of these domains. However, we can make some predictions based on the domain models that have been constructed and previously published literature on AEBP1.

Relatively little is known regarding the structure of the C-terminal domain of AEBP1, other than the presence of a MAP kinase phosphorylation site and a basic helix involved in DNA interaction (Figure 36). Studies have shown that interaction of AEBP1 with AE-1 DNA increases the CP activity of AEBP1 (*164*). This suggests that the C-terminus of AEBP1 may block the active site. Interaction of the AEBP1 C-terminal domain with DNA may cause this domain to be moved or changed in conformation such that the active site is more accessible. Amino acids 597-644 of AEBP1, which make up part of the basic helix and the intervening region between this and the CP domain, have some homology to the C-terminus of CPM. Recently the crystal structure of CPM was solved (*114*). In the crystal structure of CPM, these residues are found alongside the CP domain directed toward the active site (see Figure 4B). This suggests the possibility that the homologous residues of AEBP1 also fall into this position, placing the C-terminus in proximity to the putative active site within the CP domain of AEBP1.

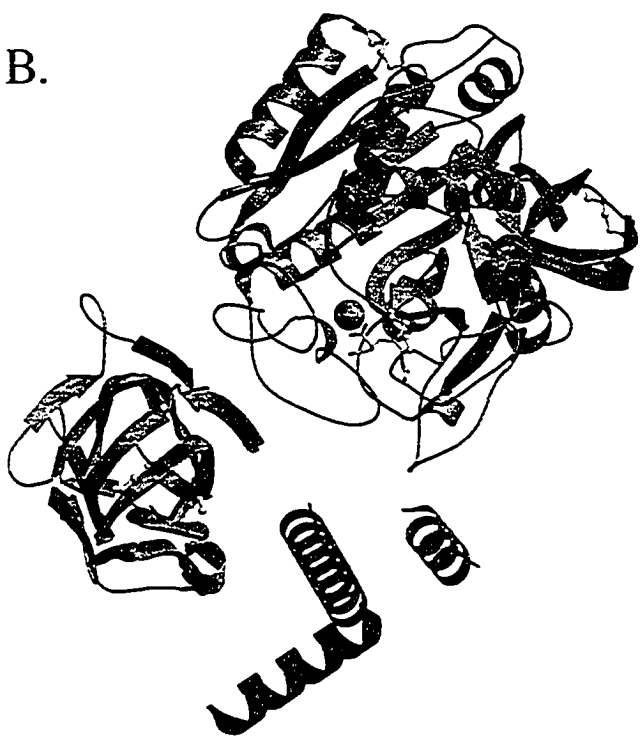
Interactions of AEBP1 with MAP kinase give us more clues as to relative domain orientations. MAP kinase has been shown to interact with the DLD domain of AEBP1 (*7*),

although results shown in Figure 31 also suggest that the C-terminus of AEBP1 can be sufficient for interaction. The study by Kim *et al.* (*7*) also indicated that the C-terminus of AEBP1 is necessary for protection of MAP kinase from dephosphorylation. This suggests that, while the N-terminus is important for strong interaction, the C-terminus also interacts with MAP kinase at the same time to mediate protection from dephosphoryation. Therefore the N- and C-termini of AEBP1 are at least close enough to each other to both interact with MAP kinase at the same time. Interaction of the N- and C-termini of AEBP1 was also suggested by kinase assays described in this thesis, in which removal of the DLD domain of AEBP1 (AEBP1ΔN) resulted in greater *in vitro* phosphorylation by MAP kinase than occured with full-length AEBP1 (Figure 31). The DLD domain may interact with the C-terminus and block accessibility to the phosphorylation site within the C-terminus. Interaction of MAP kinase with AEBP1 has also been shown to increase the ability of AEBP1 to bind DNA (*8*). This further suggests that the DLD and C-terminal domains of AEBP1 may interact with each other and that interaction of the DLD domain with MAP kinase may free the C-terminal domain to interact with DNA. This DNA binding might then allow greater accessibility to the putative CP active site, as mentioned above. Many carboxypeptidases are produced with a propeptide segment that must be removed in order for the enzyme to exhibit full activity. A look at the partially beta-sheet propeptide segment in the crystal structure for proCPA2, 1AYE (Ref. (*110*) and Figure 4A), allows us to imagine the DLD domain of AEBP1 in an analogous position covering the active site. The DLD and C-terminal domains of AEBP1 together might function as a type of "prodomain", which must be moved out of the way for full AEBP1 CP activity. A model of this hypothesis is shown in (Figure 38).

Figure 38. Proposed model for the three-domain structure of AEBP1. The DLD (cyan) and CP domain (brown) homology models are shown in their proposed relative orientations along with three predicted helices. (A) Connected to the CP domain by a brown dotted line are amino acids 391-403, predicted to form an α-helix although not part of the CP homology model. The blue helix indicates the predicted helix (amino acids 625-650) within the basic region and the red helix indicates the predicted helix (amino acids 713-727) within the acidic region of the C-terminal domain of AEBP1. Dotted black lines indicate connecting segments between domains. (B) is the same as (A) but without connecting segments and rotated approximately 90° so that the view is from the top instead of from the side.

A.

B.



In addition to points from the relevant literature mentioned above, certain structural features of each domain, identified by homology modeling, guided my placement of each domain. The DLD domain has two distinct faces; a highly charged hydrophilic face and a more hydrophobic face (see Figure 9). The hydrophobic face was oriented towards the CP domain where it would form hydrophobic contacts with the surface of the CP domain, while the hydrophilic face would be exposed to the aqueous environment at the surface of the protein. The C-terminus of AEBP1 contains two predicted helices; one arginine-rich basic helix and one glutamate-rich acidic helix (see Figure 13). One might predict that these two helices may interact with one another through ionic interactions. It might also be predicted that these helices may interact with the highly charged face of the DLD domain. In my model these helices have been placed in such a manner that the basic helix is sandwiched between the acidic helix, which interacts with the DLD domain, and the helix predicted to form from residues 391 and 403 of the CP domain (Figure 13). This predicted helix is a part of one of the un-modeled loops of the CP domain, and is rather uncharged except for one unconserved glutamate, suggesting that it may not be surface-exposed but may function through hydrophobic interactions. Interestingly, along with the DLD domain of AEBP1, this loop also seems to be important for maintaining high steady-state levels of overexpressed AEBP1 protein (Figure 21). Thus it could be that the DLD domain and amino acids 391-403 are crucial structures for holding the C-terminus in place and maintaining the structural stability of the protein. Since my work has shown that AEBP1 interacts with DNA predominantly through the basic helix, yet WT AEBP1 does not interact with DNA strongly, I have placed the basic helix somewhat behind the acidic helix and the helix comprising amino acids 391-403. Deletion of this acidic helix, as in ΔSTP AEBP1, might then free the basic helix for interactions with DNA.

Several regions of AEBP1 have undefined secondary and tertiary structure and so are not shown in Figure 38. These include the first un-modeled loop of the CP domain (amino acids 316-342), which is likely sandwiched between the CP and DLD domains, and portions of the C-terminus. Although the placement of the extreme C-terminus of AEBP1 is unknown, sequence analysis and purification efforts suggest that it may be tucked away and inaccessible to solvent. The C-terminal 22 amino acids of AEBP1 are largely uncharged except for three negatively charged residues and stand in striking contrast to the long stretch of negatively

charged glutamate located nearby. In fact, a consensus sequence within this hydrophobic stretch with the sequence 733-(Y/F)PΦxxxxxxYT(V/I)(D/N)-744, in which the numbering is taken from the mouse sequence and Φ indicates a hydrophobic amino acid, can be identified with several invariant residues in all orthologues (see Appendix A). This is unlike the majority of the C-terminus of AEBP1, which is not well conserved. The hydrophobic nature of the C-terminus suggests that it may be inserted into the hydrophobic interior of the protein. Attempts to purify soluble AEBP1 supported the idea that the C-terminus is not surface exposed, as very little (~2-5% of the total) wild-type AEBP1 was able to bind to the metal affinity column through the C-terminal His6 tag. On the other hand, it was found that approximately 50% of soluble AEBP1-ΔC and AEBP1-ΔSTP, both lacking the extreme C-terminus of AEBP1, was able to bind the resin (data not shown). The placement of this hydrophobic stretch of AEBP1 into the protein interior may serve as an anchor to hold the C-terminus in place and restrict accessibility to the highly charged basic and acidic helices within the C-terminus. In fact, the presence of a C-terminal "anchor" sequence in both AEBP1-WT and AEBP1(123-748) might be the reason for the lack of DNA binding observed for these two proteins, whereas AEBP1(27-730) and AEBP1-ΔSTP, both lacking this extreme C-terminal sequence, are able to interact strongly with DNA (Figure 36). Additionally, the low solubility exhibited by AEBP1(27-730) expressed in *E. coli* may be due to the C-terminal 18-amino acid "anchor" deletion preventing complete and compact folding of the protein (Figure 17).

4.2 Function of the CP domain of AEBP1

4.2.1 Enzymatic mechanism of AEBP1

The initial characterization of AEBP1 provided evidence suggesting that the CP domain of AEBP1 is critical for its function as a transcriptional repressor (*2*). This was a novel mechanism for regulation of transcription. In order to find substrates or interacting partners for AEBP1, a yeast two-hybrid screen was subsequently carried out. While several interacting proteins were found, no candidate substrates for AEBP1 were discovered (*3*). To further address the role of the CP domain of AEBP1, I undertook a structural analysis of this domain.

Molecular modeling was used to predict the positions of residues with respect to the protein surface and its putative active site. A search was performed of genomic databases for orthologues and homologues of AEBP1 in order to determine residues that are conserved and unique to AEBP1 and therefore possibly important for its function.

A combination of information obtained from both modeling and phylogenetic approaches enabled residues potentially important at the active site of AEBP1 to be identified. Several of these residues were mutated. Unfortunately, CP activity was not able to be measured as it was not detected in my hands. Expression of these mutants in CHO cells resulted in overexpression of all except the AEBP1 R388A mutant and the AEBP1 Δ386-405 deletion, both exhibiting very low levels of protein relative to wild-type AEBP1 (Figure 21). These residues may play a structural role through interdomain interactions as discussed above. All mutants were also analyzed by luciferase assay for their ability to repress transcription from the aP2 promoter. No significant difference was observed for any mutant. The identified residues may not be critical for the transcriptional function of AEBP1 and/or they may have a role in a yet unidentified function of AEBP1. In fact, in retrospect I realized that these residues may not be expected to have any effect on transcriptional repression by AEBP1, as a replacement of the CP domain of AEBP1 with CPM also resulted in transcriptional repression, suggesting that those residues that are conserved between AEBP1 and CPM are sufficient for the transcriptional repression activity of AEBP1 (*2*).

The active site of AEBP1 lacks several residues that are present in most CPs and involved in coordination of the catalytic zinc ion and catalytic activity. Two zinc-coordinating residues, H236 and E239 in AEBP1, are conserved between AEBP1 and other CPs. The remaining zinc-coordinating residue, typically histidine, and the general base, typically glutamate, are substituted with asparagine 374 and tyrosine 485, respectively, in AEBP1. While these are not normally metal coordinating or catalytic residues in carboxypeptidases, there is some precedent for amino acids with carboxamide side chains (asparagine and glutamine) and tyrosine being involved in these functions. The crystal structure for phosphomannose isomerase has indicated that the catalytic zinc ion is coordinated in a distorted trigonal bipyramidal arrangement by two histidines, a glutamate, a water molecule, and the amide oxygen of a glutamine (*182*). The crystal structure of isopenicillin N synthase (IPNS) also indicated a glutamine as a metal coordinating ligand, although it was later found that this

residue was not necessary for IPNS catalytic activity (*183, 184*). An experiment investigating the coordination of the catalytic zinc ion within carbonic anhydrase II (CAII) found that substitution of zinc-coordinating histidines with asparagine or glutamine retained metal coordination ability, although with diminished affinity (*185*). Tyrosine is involved in zinc coordination in the astacin peptidase, a member of the metzincin family of metalloproteases ((*186*), and Figure 4C). In addition to its function in zinc coordination, this tyrosine also plays a role in transition state stabilization similar to Tyr248 of CPA. Another protein of the gluzincin family of peptidases, leukotriene A4 hydrolase (LTA4H) contains a tyrosine at position 383 which, although not necessary for zinc coordination, is critically important as a proton donor in the catalytic mechanism of peptide cleavage (*187, 188*). Figure 12 illustrates the potential coordination of this zinc ion in AEBP1, with both asparagine 374 and tyrosine 485 (through a water molecule) playing a role in zinc coordination.

Of course the possibility remains that N374 and Y485 of AEBP1 may not be involved in zinc coordination or the catalytic mechanism. In fact, the multiple alignment of AEBP1 orthologues indicates that asparagine 374 is substituted with a glutamine in all non-mammalian AEBP1 orthologues, while tyrosine 485 is replaced with a phenylalanine in cow and fish orthologues, suggesting that these residues are not critical. While asparagine and glutamine both contain the same functional group, phenylalanine lacks the hydroxyl group proposed to be involved in zinc coordination. In the protease astacin, which contains a tyrosine at position 149 involved in zinc coordination, mutation of this residue to phenylalanine abolished all but 2.5% of wild-type enzyme activity (*186*). Therefore, if mouse AEBP1 has proteolytic activity dependent on Y485, we would not expect cow or fish AEBP1 to exhibit this same activity.

Other CP residues have been identified as having important roles in substrate binding (see Figure 6). CPD and all members of the N/E family anchor contain an aspartate (position 192 in CPD-II and position 385 in AEBP1) which binds the arginine/lysine side chain at the C-terminus of substrates, hence providing specificity for terminal basic residues. This is conserved in all AEBP1 orthologues (see Appendix A). The crystal structure of CPD-II bound to GEMSA (*127*) indicates the inhibitor carboxylate group of GEMSA which mimicks a peptide substrate C-terminus is bound by N144, R145, and R135, which is also important for the stabilization of the transition state. These residues are substituted in mouse AEBP1 by F310, E311 and L301, respectively. These residues are conserved in mammalian AEBP1

(although F310 is sometimes a tyrosine), but not in non-mammalian orthologues identied thus far. A peptidase from *B. sphaericus* also lacks two of these specific residues, R145 being substituted by an aspartate and R135 being substituted by asparagine (*175-177*). This peptidase is able to hydrolyse C-terminal *meso*-diominopimelic acid, which contains a free amino group in addition to the carboxyl group, or a C-terminal meso-dioaminopimelyl-alanine dipeptide. This suggests that AEBP1 may have specificity towards amino-terminal substrate residues or dipeptides. This possible amino-terminal specificity was not noticed during the course of this study and so was not investigated. Other residues involved in substrate specificity include CPD-II Y250, V252, and G255. These are substituted by N463, R465, and T468, respectively, in AEBP1. However, these residues are not conserved in non-mammalian orthologues. These observations regarding residues involved in substrate recognition suggest that mammalian AEBP1 may interact with N-terminal basic mono- or di-peptides, but that this function may not be conserved in non-mammalian species.

4.2.2 Role of the CP domain in a histone H3 arginine kinase complex

AEBP1 is shown to be the p85 calmodulin-binding component of a Ca^{2+}/calmodulin-dependent histone H3 arginine kinase complex (Figure 25). This interaction of AEBP1 with Ca^{2+}/calmodulin requires the basic helix within the C-terminus of AEBP1 (Figure 26). An interaction of this basic helix with Ca^{2+}/calmodulin is consistent with the current knowledge regarding peptides that interact with Ca^{2+}/calmodulin. While these interacting peptides share no sequence homology, they all potentially fold into an amphipathic α-helix, with large hydrophobic residues either at positions 1-5-10 or 1-8-14 (*189*). The basic region of AEBP1 can potentially fall into the 1-5-10 class of calmodulin-interacting proteins, as residues M640, M644, and L649 of mouse AEBP1 are all large hydrophobic residues (Figure 39). This area of the basic region of AEBP1 is highly conserved across all species. The hydrophobic amino acid in position 1 is not conserved in the fish AEBP1 sequences identified, suggesting that these may not interact with Ca^{2+}/calmodulin or may interact in a different manner. On the other hand, fish AEBP1 may not have the same function as AEBP1 in other vertebrates. Both Fugu and Tetraodon AEBP1 lack a methionine initiation codon in the position where AEBP1 translation is known to initiate from in mice, suggesting that only ACLP may be produced in fish.

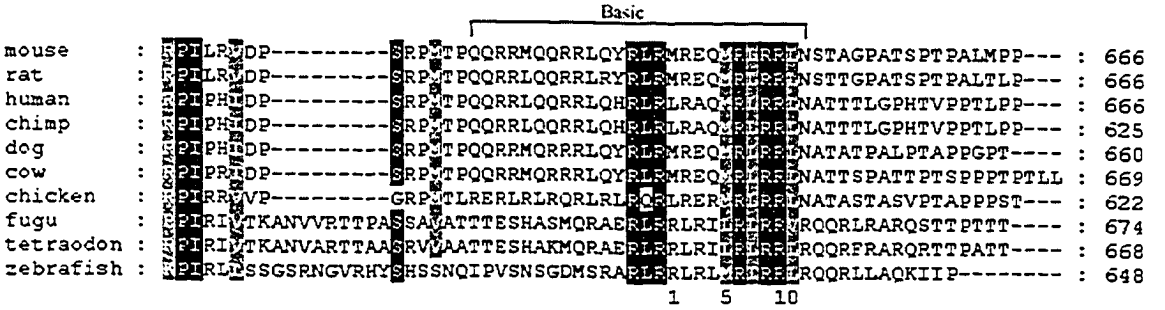


Figure 39. The basic helix of AEBP1 contains the sequence requirements for Ca^{2+}/calmodulin interaction. An alignment of AEBP1 orthologues indicates strong conservation within a portion of the basic domain at the C-terminus of AEBP1. Sequence identity is indicated by black shading, while chemical similarity is indicated by grey shading. The bottom row indicates the residues composing a potential 1-5-10 sequence of hydrophobic amino acids within the the basic helix known to be involved in Ca^{2+}/calmodulin interaction by some proteins.

The Ca^{2+}/calmodulin-dependent arginine kinase activity associated with AEBP1 has been shown to phosphorylate the N- and C-termini of histone H3 at arginines 2, 128, 129, and 131 *in vitro* (*100*). It is proposed here that the CP domain of AEBP1 may interact with the C-terminus of histone H3. A specific interaction of AEBP1 with the 12 C-terminal amino acids of histone H3 was shown by GST pulldown, although interaction of AEBP1 with full-length histone H3 has not yet been shown. AEBP1 may interact with histone H3 and subsequently recruit an arginine kinase to phosphorylate H3. Another possibility previously proposed is that p85/AEBP1 might actually be this arginine kinase or a subunit of it (*101*). The arginine phosphorylation of histone H3 could be another part of the histone code and the mechanism by which AEBP1 represses transcription. CARM1, a histone methyltransferase, also modifies the C-terminus of histone H3 by methylating one or more of the four arginines 128, 129, 131, and/or 134 (*190*). This enzyme, in contrast to the activity of AEBP1, leads to activation of gene transcription (*73*), suggesting that these two modifications may have opposing roles in transcriptional regulation. If AEBP1 is the H3 arginine kinase, then both it and CARM1 may be able to modify similar residues within the C-terminus of histone H3 *in vivo*, possibly through a similar mechanism of interaction with histone H3. A comparison of AEBP1 and CARM1 amino acid sequences did not reveal any obvious similarities, except in one short stretch corresponding to the 20 amino acid loop of AEBP1 (amino acids 386-405) not present in CPD and not modeled with the CP domain of AEBP1 (see Figures 9 and 11 for the location of this loop, and Figure 40 for a comparison with CARM1). This loop is predicted to form an α-helix forming part of the funnel leading to the putative active site of AEBP1 and is necessary for high steady-state levels of AEBP1 in CHO cells (see Figures 12 and 20 and Ref. (*127*). The similar stretch in CARM1 is also predicted to form an arm protruding from the core of the protein. This prediction is based on comparison of CARM1 with the crystal structure for a related cytoplasmic methyltransferase, PRMT3 (Figure 40). It has been shown by crystallographic studies that the hydrophobic arm of PRMT3 can be involved in homodimerization, although gel filtration studies and crosslinking studies suggest that PRMT3 is predominantly a monomer in solution (*191*). It is interesting to speculate that this region in both CARM1 and AEBP1 may be involved in interactions with histone H3 or another factor common to both proteins.

Figure 40. Amino acids 386-405 of AEBP1 are similar to a protruding arm of CARM1. (A) The amino acid sequence of the conserved core of protein arginine methyltransferase 3 (PRMT3) is aligned with the homologous region of coactivator-associated arginine methyltransferase 1 (CARM1). Amino acids 386-405 of AEBP1 are also shown aligned to a similar region of CARM1. Residue shading is either black or grey to indicate sequence identity or similarity, respectively, between two protein sequences. (B) The X-ray crystal structure of PRMT3 (PDB ID 1F3L) is shown, with the co-crystallized reaction product AdoHcy shown as a stick structure and amino acids homologous to CARM1 and AEBP1 386-405 shown in red.

A.

```
PRMT3 : EEVSLPVEKVDVIISEWMGTFLLFESMLDSVLYARSKYLAKGGSVYPDIC : 150
CARM1 : EEVSLP-EQVDIIISEPMGYMLFNERMLESYLHAK-KYLKPSGNMFPTIG : 148
AEBP1 : -------------------------------------------------- :   -

PRMT3 : TISLVAVSEVSKHAD-----PIAFWDDVYGFNMSCMRKAVIPE-----AVV : 191
CARM1 : DVHLAPFTEQLYMEQFTKANFRYTPSFHEVDLSALRGAAVDEYFRQPVV : 198
AEBP1 : ------------------MARTPSLEQILAEALAARE------------ :  20

PRMT3 : EVVDHRTLISDPCDIKHIDCHTTSISDLEFSSDETLRTTKTAVCTAVAGV : 241
CARM1 : DTFLIRILMAKSVKYTVNFLEAKEGDLHRIEIPFKFHMLHSGLVHGLAFW : 248
AEBP1 : -------------------------------------------------- :   -

PRMT3 : EDIYEEKNCHNRVVESTGPQSTKTHCKTIFLLEKEFPVRAGEALKGKIT : 291
CARM1 : ELVASIG-SIMTVWLSTAETEPLTHTYEVRCLFCSFLFARAGDTLSETCL : 297
AEBP1 : -------------------------------------------------- :   -
```

B.

Phosphorylation of arginine within histone H3 might also be involved in cell cycle control through a non-transcriptional mechanism. Wakim *et al.* (*101*) have shown that the Ca^{2+}/calmodulin-dependent histone H3 arginine kinase activity is activated within quiescent rat heart endothelial cells as compared with actively growing cells, while the levels of the major protein component of this complex, p85/AEBP1, do not change. They proposed in this report that p85/AEBP1 may be the H3 arginine kinase or a subunit of this kinase and that it may be involved in cell cycle exit. A role for AEBP1 in cell cycle exit is somewhat contrary to generally held ideas implicating AEBP1 in cell proliferation. This could be resolved by possible cell-type and tissue-specific differences in the function of AEBP1. An earlier paper by Wakim *et al.* (*99*) in which a similar kinase activity was isolated from calf thymus found that attempts to further purify kinase activity following elution from a CaM-Sepharose affinity column were unsuccessful. They proposed that the arginine kinase activity might be part of a large complex, separation of which abolishes kinase activity. Consistent with this concept, gel filtration experiments have been performed showing that AEBP1 is part of complexes ranging in size from 100 to 900 kDa in quiescent 3T3-L1 cells (A. Muise, unpublished). AEBP1 remains associated with these complexes upon cell stimulation with serum or insulin for 10 or 30 minutes, but is associated only with smaller complexes (100-450 kDa) upon serum or insulin stimulation of cells for 2 hours. Changes in protein interactions could be the mechanism by which the activity of this arginine kinase activity, and possibly AEBP1, changes between cell types and between cell quiescence and growth. Phosphorylation of AEBP1, as described in this thesis, might be one mechanism by which t he interaction of AEBP1 with complex components is changed. Phosphorylation of AEBP1 in actively dividing cells might also prevent AEBP1 from interacting with histone H3 and therefore result in a low level of arginine kinase activity.

4.3 DNA binding by AEBP1

There are many conserved domains used by proteins to interact with a specific sequence of DNA. These are often made up of, at least in part, a basic amphipathic helix which inserts itself into the major groove of the DNA with positively charged side chains making contacts with the negatively charged DNA backbone. This is the case with C/EBP and other basic leucine zipper DNA binding proteins (*180*), as well as for homeodomain containing proteins such as Msx-1/2 and others (*192, 193*). Interestingly, there are similarities between both of these groups of proteins and AEBP1. AEBP1 contains a region in its C-terminus that is predicted to form a basic helix and that has sequence similarity with many basic leucine zipper DNA binding proteins (Figure 32). As well as this similarity in the putative binding domain, both proteins interact with the same stretch of DNA, the AE-1 site of the aP2 promoter (*2, 31*). This suggests that they may have somewhat similar mechanisms of DNA binding. However, although there is some evidence for AEBP1 dimerization (A Muise, unpublished), the leucine zipper necessary for C/EBP dimerization is not present in AEBP1, suggesting a different mechanism for AEBP1 dimerization.

Just N-terminal to the basic helix in AEBP1 is a putative MAP kinase phosphorylation site, ^{621}PMTP^{624}. I have shown here that threonine 623 of AEBP1 is important for DNA binding, as mutation to either an alanine or aspartate greatly reduces its ability to bind DNA in a gel-shift assay. Interestingly a very similar mutation in Msx2 (T147A) abolishes DNA binding by this protein (*194*). Msx2 contains a homeodomain DNA binding domain found in many proteins. It is made up of 60 amino acids that form three central helices which interact with the major groove of DNA (helix 3 making the major contribution to this interaction) and an N-terminal extension which lies across the minor groove (*195*). This N-terminal extension contains the crucial T147 as a part of the sequence KPRTP, which in the crystal structure for Msx1 makes a hydrogen bond to the DNA phosphate backbone (*195*). A similar mechanism for DNA binding may be employed by AEBP1, with the basic helix in the major groove as in C/EBP and the RPMTP motif binding to the minor groove as in Msx1.

Other results presented in this thesis suggest that WT AEBP1 may not need to bind DNA in order to repress transcription in some cell types, but that it may instead function as a co-repressor (Figures 35-37). When luciferase reporter assays were performed using AEBP1

with the major DNA binding domain removed (AEBP1 ΔC), transcriptional repression was still observed (Figure 37). Additionally, using bacterially expressed and refolded WT AEBP1, only a very minor shifted band indicating an AEBP1/AE-1 complex is observed upon long exposure (Figure 36). In fact, the stronger DNA binding exhibited by AEBP1(27-730), when compared to the amount of free probe remaining, is quite minor. This suggests that wild-type AEBP1 may not bind DNA. A lack of DNA binding ability is consistent with a recent report describing the role of ACLP in transdifferentiation of preadipocytes into smooth muscle-like cells, in which the authors state that they were unable to detect any binding of AEBP1 to the AE-1 element (*41*) using a version of AEBP1 lacking the N-terminal 30 amino acids. All *in vitro* DNA binding data for AEBP1 has been summarized in Table 4. Only two forms of AEBP1 are able to interact strongly with DNA, AEBP1(27-730) and AEBP1ΔSTP, which both lack the extreme C-terminus but still contain the basic helix within the C-terminal domain. This data suggests that the presence of the C-terminal hydrophobic stretch of the C-terminal domain prevents interaction with DNA that could otherwise occur through the basic helix within the C-terminus.

Table 4. Ability of AEBP1 and deletions to interact with AE-1 duplex DNA.

AEBP1 protein used in *in vitro* EMSA	Source of information	Strength of DNA binding
WT	(Figure 36)	-
(31-748)	Ref. (*41*)	-
(123-748)	(Figure 33)	-
(27-730)	(Figures 33 and 36)	+++
ΔSty (27-543)	(Muise and Ro, unpublished results)	-
ΔN	(Figure 36)	+
ΔC	data not shown	+
ΔSTP	(Figure 36)	++++
CP	data not shown	+

While WT AEBP1 may not efficiently interact with DNA *in vitro*, the possibility exists that AEBP1 may bind DNA *in vivo* following activation by another factor. We have shown that interaction of MAP kinase increases the ability of AEBP1(27-730) to bind DNA (*8*). Although we have not performed this experiment with full-length AEBP1, it could be that this interaction activates the DNA binding ability of full-length AEBP1. This interaction might release the basic domain from a function as an auto-inhibitory domain over the CP domain and make it available for interaction with DNA at the same time. It is worth noting that while WT AEBP1 is not able to interact well with DNA through its basic helix, it is able to interact well with Ca^{2+}/calmodulin through this same basic helix. The C-terminal end of the basic helix may be accessible and able to interact with Ca^{2+}/Calmodulin, while the N-terminal end of the basic helix may be inaccessible but necessary for interaction with DNA. Therefore an initial interaction by Ca^{2+}/calmodulin might be able to "loosen up" the structure, making the rest of the basic helix available for DNA binding.

In the event that AEBP1 does not interact with DNA, but functions solely as a co-repressor, some specific protein-protein interaction with a DNA binding protein is necessary. AEBP1 has been found to interact with PTEN (*178*), MAP kinase (*7*), Gγ5 (*3*), HSP27, and HMGB2 (Muise and Ro, unpublished). Of these, only HMGB2 interacts with DNA, but in a nonspecific manner. An intriguing possibility for a transcription factor which might interact with AEBP1 can be found in initial ACLP research. Layne *et al.* (*4*) discovered ACLP through an expression library screen using the E47 protein, which is a basic helix-loop-helix DNA-binding transcription factor of the E2A family. E47 was used to pull a partial clone out of a human aorta expression library encoding a truncated ACLP protein presumably encoding the C-terminus of ACLP/AEBP1. In further studies Layne *et al.* found that, while the truncated ACLP was able to interact with E47, full-length ACLP protein was not able to interact. The possibility remains, however, that AEBP1 may be able to interact with E47 and/or other related basic helix-loop-helix proteins.

E2A proteins such as E47 are ubiquitously expressed transcription factors (*196, 197*) that interact with E-box (CANNTG) DNA motifs and are important for the regulation of many B lymphocyte and myocyte genes (*197, 198*). They are generally considered to be transcriptional activators, although this activity can be modulated through interaction with many other proteins which inhibit dimerization of E2A proteins or interfere in some way with the

transcriptional activation mechanism. For example, TAL1/SCL heterodimerizes with E proteins leading to a depletion of the E47/HEB activating heterodimer and recruiting the mSIN3A/HDAC1 corepressor complex to many genes (*199*). Recently, E proteins were found to interact with ETO, a transcriptional co-repressor named for an aberrant translocation (8;21, or eight twenty-one) that produces an oncogenic ETO fusion, AML1-ETO (*200*). This AML1-ETO fusion protein is aberrantly expressed in leukemic cells and plays a part in many cases of acute myeloid leukemia (AML) through its interaction with E proteins. This interaction blocks the recruitment of p300/CREB-binding protein (CBP) coactivators to the E proteins and instead recruits ETO-interacting corepressors to the site of E protein binding. It could be that AEBP1 functions in a similar manner through interaction with E2A proteins and inhibition of their activation function. As we don't know exactly what sequence of ACLP was used in the interaction study showing interaction of truncated ACLP with E47, it is difficult to say what domains may be involved. The CP domain of AEBP1 contains an LxxLL motif used in many interactions between transcription factor and coactivators/corepressors (*201*). This may be used by AEBP1 for interactions with E proteins.

The actions of corepressors target specific promoters through interactions with DNA-binding transcription factors such as E2A. This thesis presents some evidence that wild-type AEBP1 is able to repress the CMV promoter driving β-galactosidase expression to a similar extent as the aP2 promoter driving luciferase expression (Figure 37). In fact, AEBP1 has been found to repress a wide variety of promoters, to a greater or lesser extent (data not shown). AEBP1 has even been shown to repress the pGL2-basic promoterless reporter plasmid (data not shown). In an attempt to address this issue and to determine if AEBP1 might affect transfection efficiency I cotransfected a GFP-expressing plasmid, pEGFP-N1, along with constructs expressing various mutants and deletion forms of AEBP1 into CHO cells. It was found that the numbers of transfected cells did not significantly change with or without the expression of AEBP1 (data not shown). However, the brightness of the GFP signal was greatly reduced when an AEBP1 expression vector was tranfected instead of empty vector. In fact, signal brightness in this GFP experiment generally paralleled the luciferase activity seen for any particular version of AEBP1 or negative control shown in (Figure 37). This suggests that AEBP1 affects the expression of GFP and not the number of cells which take up

expression plasmid (transfection efficiency). It also suggests that AEBP1 is able to repress a large number of promoters.

The general repression observed by AEBP1 on a variety of promoters suggests the possibility of global repression. Several examples of factors that regulate genes in a global manner can be mentioned. Among these are some of the histones, which might be expected to exert global effects due to their role in chromatin structure and gene regulation. Deletion of the N-termini of both histone H3 and H4 in yeast has been found to exert a large-scale global regulation of gene expression, with the H3 deletion resulting in activation of most genes while the effects of the H4 deletion were largely balanced between activation and repression (*202*). Histone H5, a variant of the linker histone H1, has also been found to cause global genetic inactivity through chromatin condensation upon H5 phosphorylation (*50*). Histone modifications leading to transcriptional regulation have generally been thought to be directed towards specific promoters through interactions of modifying enzymes with specific transcription factors. However, the results with histone H5 and others suggest that histone modifications can occur globally as well. A study of the yeast histone deacetylase, RPD3, and histone acetyltransferase, GCN5, used chromatin immunoprecipitation to determine the extent of histone acetylation on specific residues following deletion of the modifying enzyme (*203*). Large differences in the acetylation state of histones H3 and H4 were found over extended chromosomal regions covering 22 kb in total, suggesting that acetylation and deacetylation occur globally. The authors proposed that this global modification may function to reduce basal transcription and also to provide a mechanism for the quick return to the initial state of acetylation following a specific local change. Similar to the widespread acetylation found in the above study, methylation of histone H3 at lysine 79 by Dot1p in yeast has been found to occur on 90% of histone H3 proteins, but interestingly, functions to silence only telomeric regions (*204*). It was suggested that H3-K79 methylation marks active regions of chromatin and inhibits the interaction of the Sir silencing proteins. This prevention of promiscuous binding allows the limited supplies of Sir proteins to function where they are needed, at the telomeres (*205*). Non-histone global regulators have also been identified. One of these is NC2, which represses transcription globally through its interaction with TBP within the RNA polymerase II holoenzyme (*206*). The drosophila homolog of NC2 has also been identified and found to affect a wide range of promoters through repression or activation, depending on the core

promoter motif of the particular gene (*207*). Thus there is precedent for global transcriptional regulation. AEBP1 may function in a global manner through interaction with a component of the general transcription machinery as in the case of NC2, or through widespread histone modification mediated by the histone H3 arginine kinase complex of which it is a part, leading to transcriptional repression.

4.4 Phosphorylation of AEBP1 by MAP Kinase

Phosphorylation is a modification used in regulating the activities and interactions of a large number of cellular proteins. AEBP1 was previously found to be phosphorylated (*7*), and one aim of this study was to identify the sites of phosphorylation of AEBP1. Analysis of phosphorylation of AEBP1 using site-directed mutagenesis of predicted MAP kinase phosphorylation sites revealed the major site of phosphorylation by ERK1/2 MAP kinase to be serine 668. Mutation of this residue to an alanine abolished *in vitro* phosphorylation by ERK2 MAP kinase (Figure 30) as well as the *in vivo* mobility shift of AEBP1 upon EGF stimulation (Figure 29). This phosphorylation site is not present in any AEBP1 orthologues except for rat, suggesting that its function may be unique to mouse and rat AEBP1. Mutation of two other predicted phosphorylation sites, T623 and S658, in full-length AEBP1 appeared to cause a small reduction in the level of AEBP1 phosphorylation *in vitro*, but did not cause any change in the mobility shift observed *in vivo*. It has been suggested that phosphorylation of S668 might cause secondary phosphorylation of either T623 or S658 or another phosphorylation site in AEBP1. To investigate this, a kinase assay was performed with both AEBP1 S668A and S668D mutants and exposed for a long period of time (overnight) relative to the usual exposure of 30 minutes (data not shown). This assay revealed that a very minor amount of phosphorylation still remained following mutation of S668 to either alanine or aspartate. Because the phosphorylation was observed in both mutants, which we expected would mimic the phosphorylated and dephosphorylated states, a secondary phosphorylation which requires prior S668 phosphorylation does not likely occur. Additionally, the low abundance of this phosphorylation suggests that it may not be physiologically relevant.

In contrast to kinase assays performed with full-length AEBP1 suggesting that S668 is the only phosphorylation site, my study of AEBP1 T623 using AEBP1(27-730) suggested that

threonine 623 was a major *in vitro* phosphorylation site in AEBP1 (Figure 30(*8*). This appears to be a contradiction. However, the deletion of the 26 N-terminal and 18 C-terminal amino acids of AEBP1 may cause a change in conformation and make T623 more available for phosphorylation. This truncation may mimic a conformational change that may occur in AEBP1 *in vivo* making T623 available for phosphorylation. Alternatively, this truncated protein may not reflect the phosphorylation status of full-length AEBP1 *in vivo*. Although the T623 putative phosphorylation site within AEBP1 is conserved between mouse, rat, human, and cow sequences, recent database searches have also found sequences for chicken and fish AEBP1 which do not conserve this MAP kinase consensus phosphorylation sequence. This suggests that phosphorylation at this residue may not be physiologically relevant or may be important only in a mammalian system.

The function of phosphorylation of AEBP1 at S668 is currently unknown. We now know that AEBP1 is phosphorylated upon stimulation of cells with epidermal growth factor (EGF, Figure 27). Studies of the role of EGF and EGF receptor (EGFR) in the development of the mouse mammary gland have revealed that EGF signaling is critical for many stages of development, as is AEBP1. While EGFR knockout mice are not viable, transplant experiments have shown that stromal EGFR, but not epithelial EGFR, is necessary for ductal outgrowth in the pubertal mouse mammary gland (*208*). Further work has shown that the required ligand for EGFR in this stage of ductal morphogenesis is amphiregulin (AR). While the mammary glands of EGF-null and TGFα-null mice show normal ductal outgrowth during puberty, AR-null mice show severely stunted ductal growth and are unable to nurse their pups (*209*). The role of EGF and EGFR in lobuloalveolar development during pregnancy and lactation is less clear. However, several lines of evidence point to a role in this stage of development. A mutant mouse named *waved-2*, containing an EGFR mutation which impairs its tyrosine kinase activity, develops very sparse lobuloalveoli and reduced milk production (*210*). Additionally, EGFR levels and activation have been shown to peak at late pregnancy and lactation, suggesting a role in the lobuloalveoloar development that occurs at this time (*211*). One more piece of supporting evidence comes from the overexpression of TGFα, another EGFR ligand, in the mammary gland which resulted in precocious alveolar development and delayed involution (*212*), suggesting a role for EGFR in regulating the onset of involution.

In the case of AEBP1, work in our lab has found that stromal AEBP1 is necessary and sufficient to maintain normal mammary gland structure throughout pregnancy and lactation. This correlates well with the necessity for only stromal EGFR for normal ductal outgrowth. However, AEBP1 knockout studies showed that ductal outgrowth and lobuloalveolar development occured normally even without AEBP1 until late pregnancy, when signs of premature involution began, culminating in a mammary gland that was unable to lactate. It is unknown if the possible role of EGFR in lobuloalveolar development is mediated through the stroma only, but it is known that EGFR levels and activation are highest at late pregnancy and lactation. It would be interesting to know if the inactivation of EGFR specifically at late pregnancy would result in a similar phenotype as the premature involution seen in the AEBP1 knockout. To my knowledge this kind of conditional knockout has not been constructed.

Chapter 5

CONCLUSION

Research performed to date on the molecular functions of AEBP1 has focused on its role in adipogenesis. Within the nucleus AEBP1 serves as a repressor of transcription of the aP2 gene encoding fatty acid binding protein (*2*). This repression activity is inhibited by interaction with Gγ5 in 3T3-L1 cells, thus permitting 3T3-L1 cells to express the aP2 gene and differentiate into adipocytes (*3*). AEBP1 is also able to inhibit adipogenesis by protecting MAP kinase from dephosphorylation (*7*). Recent work has indicated that MAP kinase is able to stimulate the DNA binding ability of AEBP1 and to phosphorylate AEBP1 (*8*). A phosphatase with tumour suppressor function, PTEN, is also able to interact with AEBP1 (*178*). The work presented in this thesis has extended the possible molecular functions of AEBP1 and better characterized them. This chapter will summarize this work and comment on potential directions in which studies might be directed in the future.

A search of genomic databases has revealed that AEBP1 is present only in vertebrate species. While other carboxypeptidases do exist in non-vertebrate species, no homologues of AEBP1 or the related CPX-1 and CPX-2 proteins were detected. Knockout studies reveal that a lack of AEBP1 is not embryonic lethal, suggesting that it is not critical for the basic functioning of the cell but may have specialized functions within specific cells unique to vertebrates. In mice it appears to be especially important for the maintenance of the lobuloalveolar structures of the pregnant and lactating mammary gland and for fluid reabsorption within the male reproductive tract. The DLD and CP domains of AEBP1 are highly conserved, but the C-terminal domain and amino acids 386-405 are significantly divergent. One would expect that these divergent regions are not necessary for the fundamental functions of AEBP1 but may regulate its activity in species-specific ways. Interestingly, several identified functions of AEBP1 (substrate for MAP kinase phosphorylation, Ca^{2+}/calmodulin interaction, DNA interaction) are localized to the C-terminal domain and may be involved in regulating the fundamental function of AEBP1.

Despite the DLD and CP domains of AEBP1 being well conserved and now, structurally characterized, their functions and molecular mechanisms still elude us. The CP domain has been reported to be necessary for transcriptional repression (*2*). The DLD domain of AEBP1 also appears to have a role in transcriptional repression, as deletion of it resulted in a reduction in its ability to repress the aP2 promoter and an almost total elimination of its activity at the CMV promoter. Both of these domains, functioning together, may be necessary for the transcriptional activity of AEBP1. While the DLD domain might function as an inhibitory domain, a lack of detectable AEBP1 CP activity in my hands has not allowed me to investigate this possibility. It would be interesting to determine if AEBP1 CP activity could be detected upon mutation of appropriate residues to the homologous residues present in enzymatically active CPs. This could be followed by deletion analysis to examine the effect of deleting the N- and C-terminal domains on this "restored" CP activity. This might give us some clues as to the functions of the three domains of AEBP1. An understanding of the molecular mechanisms and interactions of AEBP1, specificically the DLD and CP domains, will be critical for the characterization of AEBP1 and its functions *in vivo*. It will be difficult to clarify the role of modifications such as phosphorylation of AEBP1 before the basic mechanism of its function is understood.

This thesis, while attempting to determine the role of the CP domain of AEBP1, has also characterized the functions of the C-terminal domain. The ability of AEBP1 to bind DNA relies on the presence of the basic helix within the C-terminal domain. Wild-type AEBP1, however, is not able to efficiently interact with DNA *in vitro*. A compendium of all DNA interaction data for AEBP1 suggests that the presence of the extreme C-terminus of AEBP1 restrains the basic helix so that it is unable to interact with DNA. This raises the question as to whether AEBP1 does interact with DNA *in vivo*. Experiments using CHO cells suggest that the majority of AEBP1 is localized to the cytoplasm and that DNA binding is not necessary for transcriptional repression by AEBP1 in these cells. AEBP1 may not bind DNA *in vivo* or may rely on cell-specific factors to create structural changes in AEBP1 allowing DNA interaction to occur.

The C-terminal domain of AEBP1 is also involved in interaction with Ca^{2+}/calmodulin through the basic helix. Specifically, AEBP1 interacts with Ca^{2+}/calmodulin as part of a Ca^{2+}/calmodulin-activated histone H3 arginine kinase complex, suggesting that AEBP1

activates this arginine kinase through its interaction with Ca^{2+}/calmodulin. The endogenous substrate for this arginine kinase activity has not been identified, although its presence in the nucleus suggests a nuclear substrate and a possible role in transcription. A model for the mechanism of transcriptional repression by AEBP1 is presented in Figure 41, in which interaction of Ca^{2+}/calmodulin with the C-terminal basic helix of AEBP1 leads to activation of histone H3 arginine kinase activity by enabling interaction of AEBP1 with histone H3 and subsequent recruitment of the necessary components of an arginine kinase. Phosphorylation of histone H3 may result in transcriptional repression in a manner similar to other post-translational modifications of histones. Although much of this model is speculative, it might serve as a useful starting point for further investigations into the mechanism of transcriptional repression by AEBP1.

Both the interaction of AEBP1 with calmodulin and its role in arginine phosphorylation present two fascinating avenues of research to pursue. Calmodulin is a ubiquitous protein involved in transducing signals associated with changes in cellular calcium levels which regulate many aspects of cellular life. While Ca^{2+}/calmodulin appears to regulate an arginine kinase-related activity of AEBP1 in the nucleus, the presence of ≥50% of AEBP1 in the cytoplasm suggests that Ca^{2+}/calmodulin might also regulate the cytoplasmic functions of AEBP1. Additionally, further analysis of this arginine kinase activity could open up a largely unentered area of cell signaling through phosphorylation of basic residues such as arginine.

Finally, a comparison of the cellular effects of AEBP1 with those of related proteins such as CPX-1 and CPX-2 might give us some insight into the function of AEBP1. The very similar sequences of these proteins, yet somewhat different tissue-specific expression, suggest that they may perform similar roles in different tissues. Transgenic mouse models lacking the expression of these proteins might reveal striking similarities in function. In fact, one wonders if there are any similarities between the mammary gland and testes phenotypes of the AEBP1 knockout mouse model that might suggest a similar function for AEBP1 in both systems. Unfortunately, we do not know at this time the cell-type specific expression of AEBP1 within the testes to compare with the stromal expression of AEBP1 within the mammary gland.

Many questions remain to be answered. This thesis provides an introduction to the relationship of AEBP1 to calmodulin and an arginine kinase activity and clarifies the

capabilities of AEBP1 in DNA binding and transcriptional repression. It also provides a clearer understanding of the structure of AEBP1 which is the basis of its function.

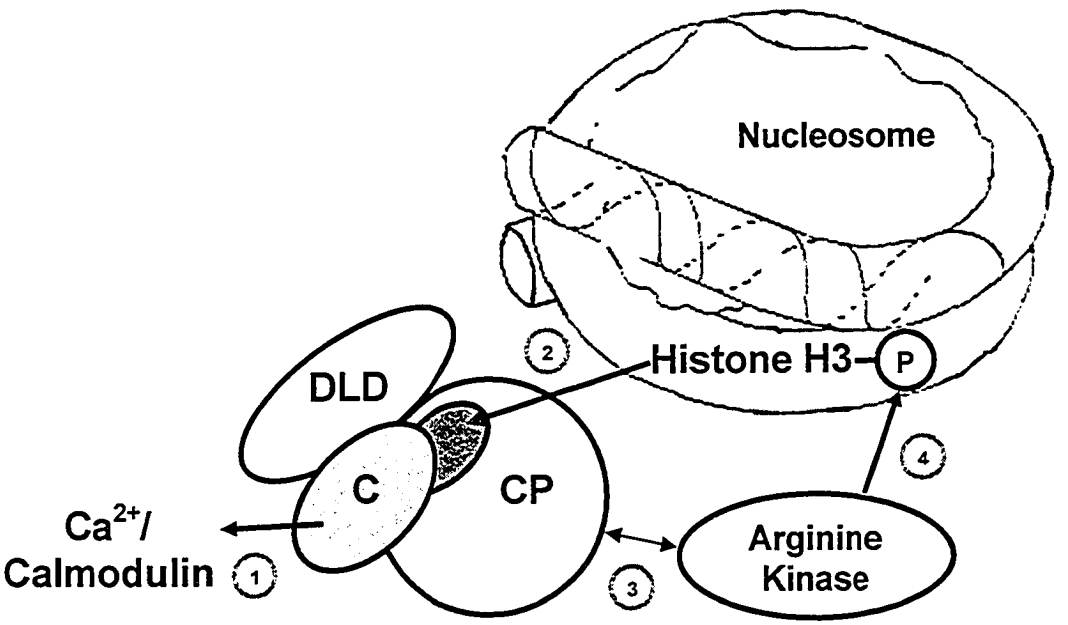
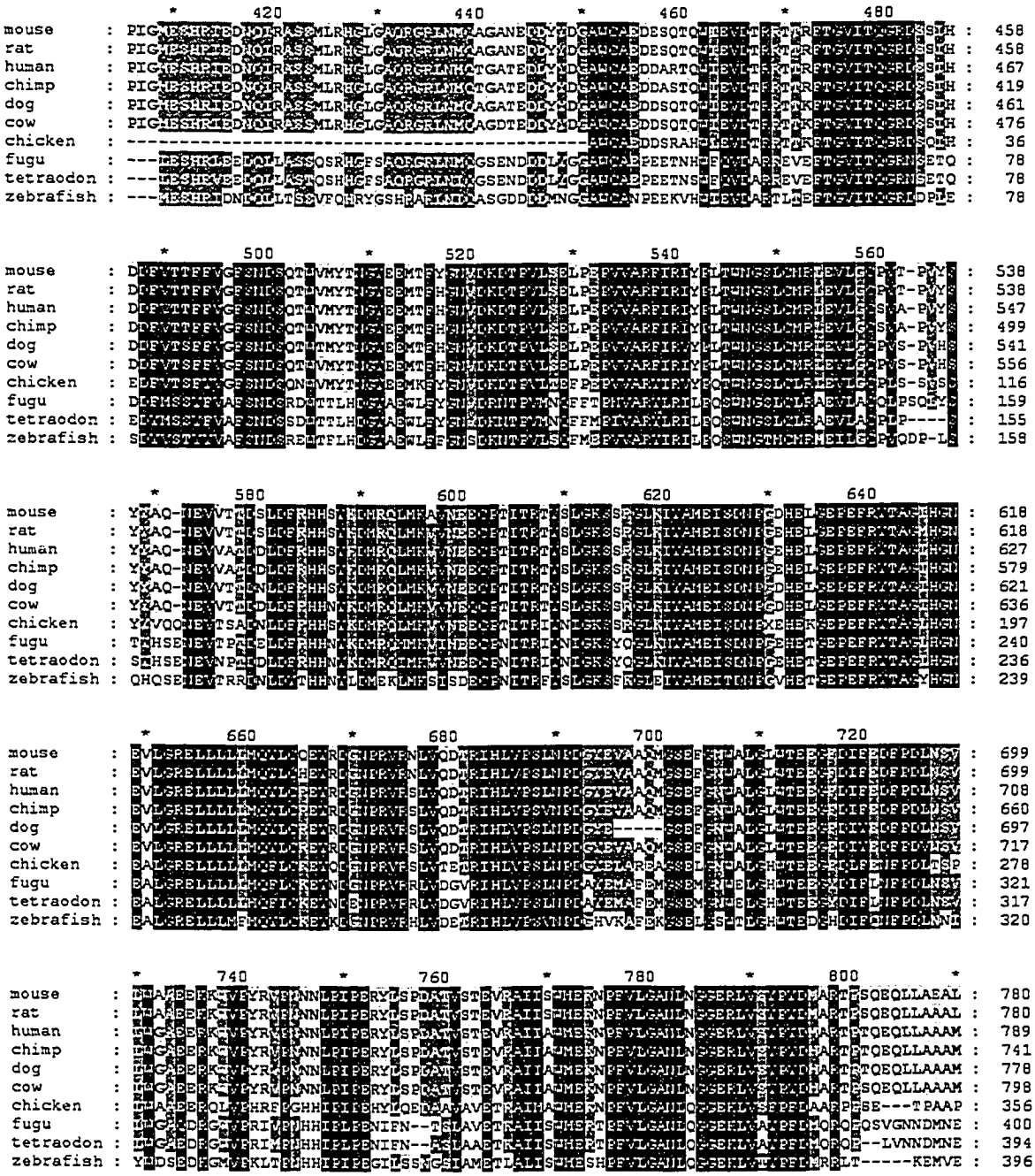


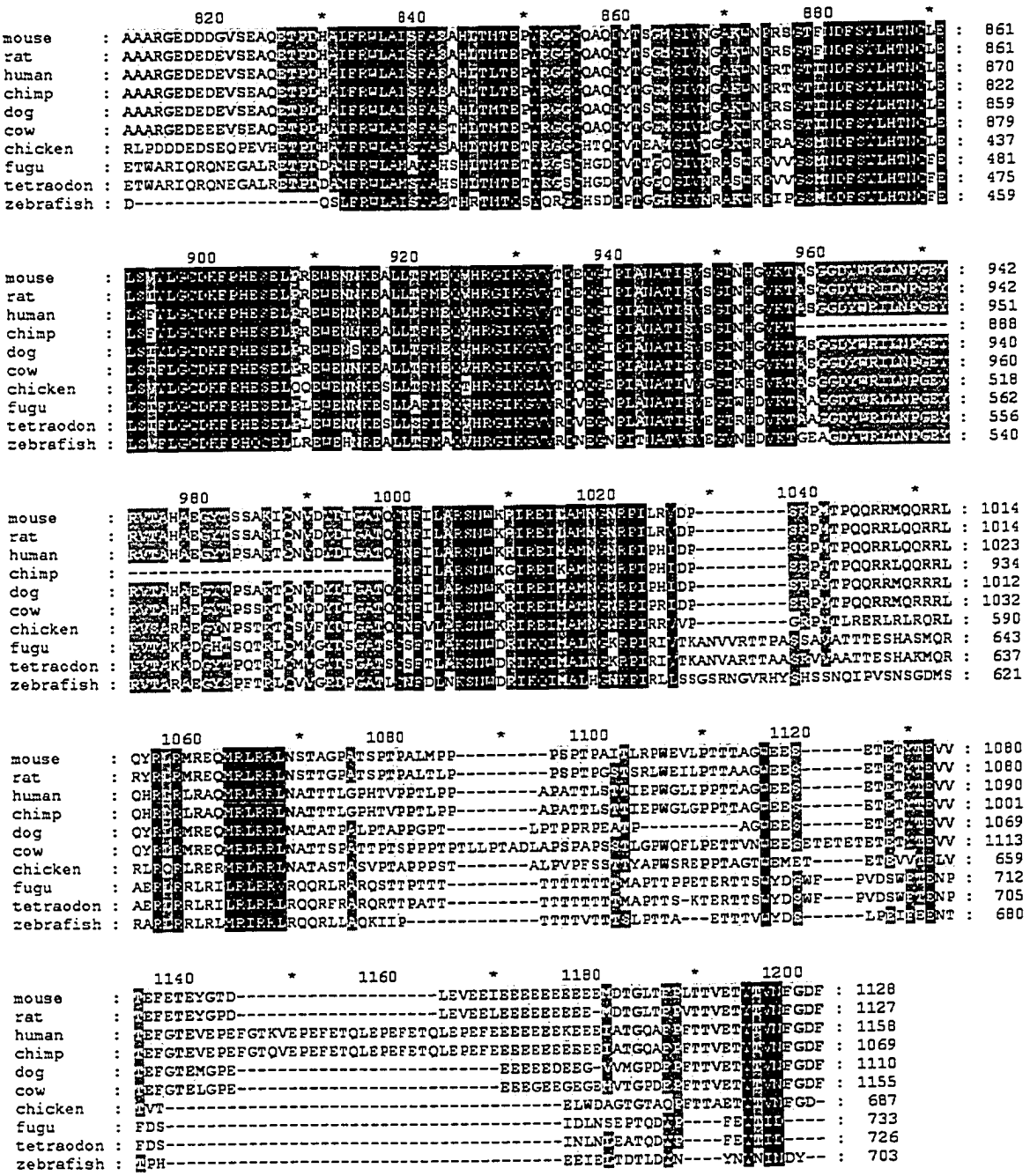
Figure 41. A model of a possible mechanism for transcriptional repression by AEBP1. In this model, interaction of Ca^{2+}/calmodulin with the C-terminus of AEBP1 causes a conformational change (1) allowing interaction of AEBP1 with histone H3 (2) within a nucleosomal structure. An arginine kinase is then recruited to AEBP1 (3) in order to phosphorylate histone H3 (4), leading to repression of transcription from the promoter involved.

Appendix A

Multiple alignment of all identified AEBP1 orthologues. For all mammalian orthologues, sequence includes additional N-terminal residues found in ACLP. Accurate identification of ACLP residues in fish and chicken orthologues was not possible due to probable sequence divergence. Shading indicates identity (black) or similarity (grey).

```
                    *        20         *        40         *        60         *        80
mouse      : MAPVRTASLLCGLLALLTLCPEGNPQTVLTDDEIEEFLEGFLSELETQSPPREDDVEVQPLP-EPTQRPRKSKAGGKQR-- :   78
rat        : MAAVRTASLLCGLLALLALCPEGSPQTVLTDDEIQEFLEGFLSEFETQSPPREDDVEAQPLP-EPTQRARKSKAGGKPR-- :   78
human      : MAAVRGAPLLSCLLALLALCPGGRPQTVLTDDEIEEFLEGFLSELEPE--PREDDVEAPPPP-EPTPRVRKAQAGGKPGKR :   78
chimp      : MAAVRGAPLLSCLLALLALCPGGRPQTVLTDDEIEEFLEGFLSELEPE--PREDDVEAPPPP-EPTPRVRKAQAGGKPGKR :   78
dog        : MAAARGAPLLGCLLALLALCPGGRPQTVLTDDEIEEFLEGFLSELESE--TREDDLEAPPPPPEPTLRDRKAQTGG----- :   74
cow        : MAALLGAPLLGCLLVLLALCPGGRPQTVLTDDEIEEFLEGFLSELESE--PREDDVEAPPPP-EPTLPARKPQTGGKPGAR :   78
chicken    : ------------------------------------------------------------------------------- :    -
fugu       : ------------------------------------------------------------------------------- :    -
tetraodon  : ------------------------------------------------------------------------------- :    -
zebrafish  : ------------------------------------------------------------------------------- :    -

                        *       100         *       120         *       140         *       160
mouse      : --ADVEVPPEKNKDKEKKGKKDKGPKATK-PLEGSTRPTKKPKEKPPK-----ATKKPKEKPPKATKKPKEKPPKATKKPK :  151
rat        : --ADAEAPPEKNKDKEKKGKKDKGPKAAK-HLEGSTRPTKKPKEKPPK-----ATKKPKEKPPKATKKPKEKPPKATKKPK :  151
human      : PGTAAEVPPEKTKDKGKKGKKDKGPKVPKESLEGSPRPPKKGKEKPPK-----ATKKPKEKPPKATKKPKEKPPKATKKPK :  154
chimp      : PGTAAEVPPEKTKDKGKKGKKDKGPKVPKESLEGSPRLPKKGKEKPPK-----ATKKPKEKPPKATKKPKEKPPKATKKPK :  154
dog        : ------TPLEKVKDKGKKGKKDKAPKATKKPLEGSPKPSKKPKEKPPK-----ATKKPKEKPPKATKKPKEKPPKATKKPK :  144
cow        : PGVAEEVPPGKAKDKGKKGKKDKSPKATKQPPEGSSRPPKKPKEKPPKEKPPKATKKPKEKPPKATKKPKEKPPKATKKPK :  159
chicken    : ------------------------------------------------------------------------------- :    -
fugu       : ------------------------------------------------------------------------------- :    -
tetraodon  : ------------------------------------------------------------------------------- :    -
zebrafish  : ------------------------------------------------------------------------------- :    -

                    *       180         *       200         *       220         *       240
mouse      : EKPPKATKRPSAGKKPSTVAPLETLDRLLPSPSNPSAQELPQKRDTPFPNAWQGQGEETQVEAKQPR---PEPEEETEMPT :  229
rat        : EKPPKATKRPSAGKRPSTVAPLETPERSLTSPSNPGTRELPEERGRTSLNTWQGQGEETQVEARQHR---PEPEEETEMPT :  229
human      : EKPPKATKKPPSGKRPPILAPSETLEWPLPPPPSPGPEELPQEGGAPLSNNWQNPGEETHVEAREHQ---PEPEEETEQPT :  232
chimp      : EKPPKATKKPPSGKRPPILAPSETLEWPLPPPPSPGPKELPQE---------------------------------------- :  197
dog        : EKPPKATKKPKE--KPPTLTPSEAPWWPLPVPPSPGPEELPQEGG-PPQIHWLGPGGETDVERQAATALPPEPEEETEPPT :  222
cow        : EKPPKATKRPLAGKKLSIPTPSESPQWPLPPLLSPDHEELPQEGGGPLPDPWGGPGEETDVEHQ------PEPEEETEQPT :  234
chicken    : ------------------------------------------------------------------------------- :    -
fugu       : ------------------------------------------------------------------------------- :    -
tetraodon  : ------------------------------------------------------------------------------- :    -
zebrafish  : ------------------------------------------------------------------------------- :    -

                    *       260         *       280         *       300         *       320
mouse      : LDYNDQIEKEDYEDFEYIRRQKQPRPTPSRRR----LWPERPEEKTEEPEER-----KEVEPP-LKPLLPP---DYGDSYV :  297
rat        : LDYNDQIEREDYEDCDTSRRQKQPRPTPSRKR----IWPEPPEEKTQEPEER-----KEVDPP-LKPLLPP---DYGDGYL :  297
human      : LDYNDQIEREDYEDFEYIRRQKQPRPPPSRRRRPERVWPEPPEEKAPAPAP-----EERIEPP-VKPLLPPLPPDYGDGYV :  307
chimp      : ---------------EYIRRQKQPRPPPSRRRRPERVWPEPPEEKAPAPAPA---PEERIEPP-VKSLLPPLPPDYGDGYV :  259
dog        : LDYNDQIEREDYEDFEYIRRQKQPRPSPSRRR-PERLWPERPEEKAELPGPGLEAPEERIEPP-LKPLPPPPPPDYGDGYL :  301
cow        : LDYNDQIEREDYEDFEYIRRQKQPRPPPSRRP-PERPWPERPEEKAEPPGRGFEAPEEKIEPPPLKPLPPPLLPDYEDGYV :  314
chicken    : ------------------------------------------------------------------------------- :    -
fugu       : ------------------------------------------------------------------------------- :    -
tetraodon  : ------------------------------------------------------------------------------- :    -
zebrafish  : ------------------------------------------------------------------------------- :    -

                    *       340         *       360         *       380         *       400
mouse      : IPNYDDLDYYFPHPPPQKPDVGQEVDEEKEEMKKPKKEGSSPKEDT-EDKWTVEKNKDHKGPRKGEELEEEWAPVEKIKCP :  377
rat        : IPNYDDLDYYFPHPPPQKPDVGQEVDEEKEELKKPKKEGSSPKEDT-EDKWAAEKNKDHKGPRKGEELEEEWGPVEKIKCP :  377
human      : IPNYDDMDYYFGPPPPQKPDAERQTDEEKEELKKPKKEDSSPKEET--DKWAVEKGKDHKEPRKGEELEEEWTPTEKVKCP :  386
chimp      : IPNYDDMDYYFGPPPPQKPDAERQTDEEKEELKKPKKEDSSPKEET--DKWAVEKGKDHKEPRKGEEVEEEWMPTEKVKCP :  338
dog        : IKHYDDMDYYFGPPRPKKPHTDLETDEEKEELKKPKKEGR--KEEETDDKWTVEKGKDHKGPRKGEDLEEEWAPTERVKCP :  380
cow        : IPNYDDMDYYPRPPKPQKPDTGLETDEEKEELKKPKKEGSSPKEEETDDKWTVEKGKDPKGPRKGEESEEEWAPVEKIKCP :  395
chicken    : ------------------------------------------------------------------------------- :    -
fugu       : ------------------------------------------------------------------------------- :    -
tetraodon  : ------------------------------------------------------------------------------- :    -
zebrafish  : ------------------------------------------------------------------------------- :    -
```


mouse
rat
human
chimp
dog
cow
chicken
fugu
tetraodon
zebrafish


820 * 840 * 860 * 880 *
mouse : AAARGEDDDGVSEAQ : 861
rat : AAARGEDEDEVSEAQ : 861
human : AAARGEDEDEVSEAQ : 870
chimp : AAARGEDEDEVSEAQ : 822
dog : AAARGEDEDEVSEAQ : 859
cow : AAARGEDEEEVSEAQ : 879
chicken : RLPDDDEDSEQPEVH : 437
fugu : ETWARIQRQNEGALR : 481
tetraodon : ETWARIQRQNEGALR : 475
zebrafish : D : 459
900 * 920 * 940 * 960 *
mouse : 942
rat : 942
human : 951
chimp : 888
dog : 940
cow : 960
chicken : 518
fugu : 562
tetraodon : 556
zebrafish : 540
980 * 1000 * 1020 * 1040 *
mouse : 1014
rat : 1014
human : 1023
chimp : 934
dog : 1032
cow : 590
chicken : 643
fugu : 637
tetraodon : 621
zebrafish : 621
1060 * 1080 * 1100 * 1120 *
mouse : 1080
rat : 1080
human : 1090
chimp : 1069
dog : 1113
cow : 659
chicken : 712
fugu : 705
tetraodon : 680
1140 * 1160 * 1180 * 1200
mouse : 1128
rat : 1127
human : 1158
chimp : 1069
dog : 1110
cow : 1155
chicken : 687
fugu : 733
tetraodon : 726
zebrafish : 703

Appendix B

Assembly of Chicken AEBP1 EST clones.

```
                        *        20         *        40         *        60         *
mouseAEBP1       : CCCACCTATTGGGATGGAGTCACACCGCATTGAGGACAACCAGATCCGTGCCTCCTCCATGCTGCGCCACGGCCT :   75
pgf1n.pk006.h24  : ------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------- :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ------------------------------------------------------------------------- :    -
040024.2         : ------------------------------------------------------------------------- :    -

                     80         *       100         *       120         *       140         *
mouseAEBP1       : CGGAGCCCAGCGGGCCGGCTCAACATGCAGGCTGGTGCCAATGAAGATGACTACTATGACGGGGCATGGTGTGCT :  150
pgf1n.pk006.h24  : ----------------------------------------------------------CGCGTGGTGCGCA :   13
pgp1n.pk012.m18  : ------------------------------------------------------------------------- :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ------------------------------------------------------------------------- :    -
040024.2         : ------------------------------------------------------------------------- :    -

                        160         *       180         *       200         *       220
mouseAEBP1       : GAGGACGAGTCGCAGACCCAGTGGATCGAGGTGGACACCCGAAGGACAACTCGGTTCACGGGCGTCATCACTCAG :  225
pgf1n.pk006.h24  : GAGGACGACAGCCGCGCGCATTGGCTGGAGGTGGACACGCGCCGCACCACCAAATTCACCGGGCGTCATCACGCAG :   88
pgp1n.pk012.m18  : ----------------------------------------------------------CANGCAG :    7
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ------------------------------------------------------------------------- :    -
040024.2         : ------------------------------------------------------------------------- :    -

                     *       240         *       260         *       280         *       300
mouseAEBP1       : GGCCGTGACTCCAGCATCCATGACGACTTCGTGACTACCTTCTTTGTGGGCTTCAGCAATGACAGCCAGACCTGG :  300
pgf1n.pk006.h24  : GGACGCGACTCCCAGATCCATGAGGACTTCGTCACCAGCTTCTATGTGGGCTTCAGCAACGACAGCCAGAACTGG :  163
pgp1n.pk012.m18  : GGACGCGACTCCCAGATCNNTGAGGACTTCGTCACCAGCTTCTATGTGGGCTTCAGCAACGACAGCCNNAACTGG :   82
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ------------------------------------------------------------------------- :    -
040024.2         : ------------------------------------------------------------------------- :    -

                        *       320         *       340         *       360         *
mouseAEBP1       : GTGATGTACACCAATGGCTACGAGGAAATGACCTTCTATGGAAATGTGGACAAGGACACACCTGTGCTGAGCGAG :  375
pgf1n.pk006.h24  : GTGATGTACACCAACGGCTACGAGGAGATGAAGTTCTATGGCAACGTGGACAAGGACACGCCGGTGCTGACCGAG :  238
pgp1n.pk012.m18  : GTGATGTACACCAACGGCTACGAGGAGATGAAGTTCTATGGCAACGTGGACAAGGACACGCCGGTGCTGACCGAG :  157
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ------------------------------------------------------------------------- :    -
040024.2         : ------------------------------------------------------------------------- :    -

                     380         *       400         *       420         *       440         *
mouseAEBP1       : CTCCCTGAGCCAGTTGTGGCCCGTTTCATCCGCATCTATCCACTCACCTGGAACGGTAGCCTGTGCATGCGCCTG :  450
pgf1n.pk006.h24  : TTCCCCGAGCCCGTGGTGGCCCGCTACATCCGCGTGTACCCGCAGACGTGGAACGGCAGCCTCTGCCTGCGGCTG :  313
pgp1n.pk012.m18  : TTCCCCGAGCCCGTGGTGGCCCGCTACATCCGCGTGTACCCGCAGACGTGGAACGGCAGCCTCTGCCTGCGGCTG :  232
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ------------------------------------------------------------------------- :    -
040024.2         : ------------------------------------------------------------------------- :    -
```

```
                          460         *       480         *       500         *       520
mouseAEBP1       : GAGGTGCTAGGCTGCCCCGTGACCCCTGTCTACAGCTACTACGCACAG---AATGAGGTGGTAACTACTGACAGC :  522
pgf1n.pk006.h24  : GAGGTGCTGGGCTGTCCCCTCTCCTCTGTCAGCAGTTACTACGTGCAGCAGAACGAGGTGACCTCAGCCGACAAC :  388
pgp1n.pk012.m18  : GAGGTGCTGGGCTGTCCCCTCTCCTCTGTCAGCAGTTACTACGCGCAGCAGAACGAGGTGACCTCAGCCGACAAC :  307
pgf2n.pk001.c7   : ---------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ---------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ---------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ---------------------------------------------------------------------------- :    -
040024.2         : ---------------------------------------------------------------------------- :    -

                      *       540         *       560         *       580         *       600
mouseAEBP1       : CTGGACTTCCGGCACCACAGCTACAAGGACATGCGCCAGCTG---ATGAAGGCTGTCAATGAGGAGTGCCCCACA :  594
pgf1n.pk006.h24  : CTGGACTTCCGTCACCACAGCTACAAGGACATGAGGCAGGTGGGTATGGCGGCGGCGGNNNNNNGGTGTCCCCGC :  463
pgp1n.pk012.m18  : CTGGACTTCCGTCACCACAGCTACAAGGACATGAGGCAGCTG---ATGAAGGTGGTGAATGAGGAGTGCCCCACC :  379
pgf2n.pk001.c7   : ---------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ---------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ---------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ---------------------------------------------------------------------------- :    -
040024.2         : ---------------------------------------------------------------------------- :    -

                          *       620         *       640         *       660         *
mouseAEBP1       : ATCACTCGCACATACAGCCTGGGCAAGAGTTCACGAGGGCTCAAGATCTACGCAATGGAAATCTCAGACAACCCT :  669
pgf1n.pk006.h24  : GCTGCCTCCCAGCCCCCGTCCTCTTCCTCTGCCCCCCTCCCCGCAGCTGATGAA-------------------- :  517
pgp1n.pk012.m18  : ATCACCCGCATCTACAACATCGGCAAGAGCTCACGTGGGCTGAAGATCTACGCCATGGAGATCTCTGACAACCCG :  454
pgf2n.pk001.c7   : ------------------------------------------------------------CACACACCCCATAC :   15
pgm2n.pk007.g13  : ---------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ---------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ---------------------------------------------------------------------------- :    -
040024.2         : ---------------------------------------------------------------------------- :    -

                      680         *       700         *       720         *       740         *
mouseAEBP1       : GGGGATCATGAACTGGGGGAGCCCGAGTTCCGCTACACAGCCGGGATCCACGGCAATGAGGTGCTAGGCCGAGAG :  744
pgf1n.pk006.h24  : ---------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : NNNGAGCACGAGAAGGGAGAGCCCGAGTTCCGCTACACGGCGNNNCTGCACGGCAATGAGGCGCTGGGCCGTGAG :  529
pgf2n.pk001.c7   : CCCTGGGTCCCCGCAGGAGAGCCCGAGTTCCGCTACACGGCGGGGCTGCACGGCAATGAGGCGCTGGGCCGTGAG :   90
pgm2n.pk007.g13  : ---------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ---------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ---------------------------------------------------------------------------- :    -
040024.2         : ---------------------------------------------------------------------------- :    -

                          760         *       780         *       800         *       820
mouseAEBP1       : CTCCTGCTCCTGCTCATGCAATACCTATGCCAGGAGTACCGCGATGGGAACCCGAGAGTGCGCAACCTGGTGCAG :  819
pgf1n.pk006.h24  : ---------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : CTGCTGCTGCTGCTGATGCAGTTCCTGTGCNNNNNGTACCANGACNNCAACCCCCGCGTGCGCAGCCTCGT------ :  600
pgf2n.pk001.c7   : CTGCTGCTGCTGCTGATGCAGTTCCTGTGCAAGGAGTACCAGGACGGCAACCCCCGCGTGCGCAGCCTCGTCACC :  165
pgm2n.pk007.g13  : ------------------------------------------CGGCAACCCCCGCGTGCGCAGCCTCGTCACC :   31
pgf1n.pk006.o3   : ---------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ---------------------------------------------------------------------------- :    -
040024.2         : ---------------------------------------------------------------------------- :    -

                      *       840         *       860         *       880         *       900
mouseAEBP1       : GACACACGCATCCACCTGGTGCCCTCGCTGAACCCTGATGGCTATGAGGTGGCAGCGCAGATGGGCTCAGAGTTT :  894
pgf1n.pk006.h24  : ---------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : ---------------------------------------------------------------------------- :    -
pgf2n.pk001.c7   : GAGACACGCATCCACCTCGTGCCCTCGCTCAACCCCGACGGCTACGAGCTGGCACGGGAGGCGGGCTCCGAGCTG :  240
pgm2n.pk007.g13  : GAGACACGCATCCACCTCGTGCCCTCGCTCAACCCCGACGGCTACGAGCTGGCACGGGAGGCGGGCTCCGAGCTG :  106
pgf1n.pk006.o3   : ---------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : ---------------------------------------------------------------------------- :    -
040024.2         : ---------------------------------------------------------------------------- :    -

                          *       920         *       940         *       960         *
mouseAEBP1       : GGGAACTGGGCACTGGGGCTGTGGACTGAGGAGGGCTTTGACATCTTCGAGGACTTCCCAGATCTCAACTCTGTG :  969
pgf1n.pk006.h24  : ---------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : ---------------------------------------------------------------------------- :    -
pgf2n.pk001.c7   : GGCAACTGGGCGCTGGGCCACTGGACGGAGGAGGGCTTCGATCTCTTTGAGAACTTCCCCGACCTGACGTCACCG :  315
pgm2n.pk007.g13  : GGCAACTGGGCGCTGGGCCACTGGACGGAGGAGGGCTTCGATCTCTTTGAGAACTTCCCCGACCTGACGTCACCG :  181
pgf1n.pk006.o3   : ------------------------------------CGATCTCTTTGAGAACTTCCCCGACCTGACGTCACCG :   37
pgf1n.pk009.p24  : ---------------------------------------------------------------------------- :    -
040024.2         : ---------------------------------------------------------------------------- :    -
```

```
                    980         *       1000         *       1020         *       1040         *
mouseAEBP1      : CTCTGGGCAGCTGAGGAGAAGAAATGGGTCCCCTACAGGGTCCCAAACAATAACTTGCCAATCCCTGAACGTTAC : 1044
pgf1n.pk006.h24 : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18 : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7  : CTGTGGGCCGCCGAGGAGCGGCAGTTGGTGCCGCACCGCTTCCCCGGCCACCACATCCCCATCCCGGAGCACTAC : 390
pgm2n.pk007.g13 : CTGTGGGCCGCCGAGGAGCGGCAGTTGGTGCCGCACCGCTTCCCCGGCCACCACATCCCCATCCCGGAGCACTAC : 256
pgf1n.pk006.o3  : CTGTGGGCCGCCGAGGAGCGGCAGTTGGTGCCGCACCGCTTCCCCGGCCACCACATCCCCATCCCGGAGCACTAC : 112
pgf1n.pk009.p24 : ------------------------------------------------------------------------------ :    -
040024.2        : ------------------------------------------------------------------------------ :    -

                         1060         *       1080         *       1100         *       1120
mouseAEBP1      : CTGTCCCCAGATGCCACGGTCTCCACAGAAGTCCGGGCCATTATTTCCTGGATGGAGAAGAACCCCTTTGTGCTG : 1119
pgf1n.pk006.h24 : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18 : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7  : CTGCAGGAGGACGCCGCGGTGGCCGTGGAGACGCGAGCCATCATGGCCTGGATGGAGAAGAACCCCTTCGTGCTG :  465
pgm2n.pk007.g13 : CTGCAGGAGGACGCCGCGGTGGCCGTGGAGACGCGAGCCATCATGGCCTGGATGGAGAAGAACCCCTTCGTGCTG :  331
pgf1n.pk006.o3  : CTGCAGGAGGACGCCGCGGTGGCCGTGGAGACGCGAGCCATCATGGCCTGGATGGAGAAGAACCCCTTCGTGCTG :  187
pgf1n.pk009.p24 : ------------------------------------------------------------------------------ :    -
040024.2        : ------------------------------------------------------------------------------ :    -

                    *       1140         *       1160         *       1180         *       1200
mouseAEBP1      : GGTGCAAATCTGAACGGTGGTGAGCGGCTTGTGTCTTATCCCTATGACATGGCCCGGACACCTAGCCAGGAGCAG : 1194
pgf1n.pk006.h24 : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18 : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7  : GGAGCCAACCTGCAGGGCGGGGAGAAGTTGGTCTCCTTCCCCTTCGACGCAGCGCGCCCCCCCAGCGAGNCCCCT :  540
pgm2n.pk007.g13 : GGAGCCAACCTGCAGGGCGGGGAGAAGTTGGTCTCCTTCCCCTTCGACGCAGCGCGCCCCCCCAGCGAGACCCCT :  406
pgf1n.pk006.o3  : GGAGCCAACCTGCAGGGCGGGGAGAAGTTGGTCTCCTTCCCCTTCGACGCAGCGCGCCCCCCCAGCGAGACCCCT :  262
pgf1n.pk009.p24 : ------------------------------------------------------------------------------ :    -
040024.2        : ------------------------------------------------------------------------------ :    -

                         *       1220         *       1240         *       1260         *
mouseAEBP1      : CTGTTGGCCGAGGCACTGGCAGCTGCCCGCGGAGAAGATGATGACGGGGTGTCTGAGGCCCAGGAGACTCCAGAT : 1269
pgf1n.pk006.h24 : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18 : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7  : GCAGCCCCGCGCCTGCCGGACGACGATGAGGACAGCGAGCAGCCCGNNNTGCACGNNNCGCCCGNNNACGCCG--  :  613
pgm2n.pk007.g13 : GCAGCCCCGCGCCTGCCGGACGACGATGAGGACAGCGAGCAGCCCGAGGTGCACGAGACGCCCGACCACGCCG--  :  479
pgf1n.pk006.o3  : GCAGCCCCGCGCCTGCCGGACGACGATGAGGACAGCGAGCAGCCCGAGGTGCACGAGACGCCCGACCACGCCG--  :  335
pgf1n.pk009.p24 : ------------------------------------------------------------------------------ :    -
040024.2        : ------------------------------------------------------------------------------ :    -

                    1280         *       1300         *       1320         *       1340         *
mouseAEBP1      : CACGCTATTTTCCGCTGGCTGGCCATCTCATTTGCCTCCGCCCATCTCACCATGACGGAGCCCTACCGGGGAGGG : 1344
pgf1n.pk006.h24 : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18 : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7  : -------TGTTCCGCTGGCTGGCCATCTCCTACGCCTCGGCACACCT---------------------------- :  653
pgm2n.pk007.g13 : -------TGTTCCGCTGGCTGGCCATCTCCTACGCCTCGGCACACCTCACCATGACCGAGACGTTCCGCGNNNGC :  547
pgf1n.pk006.o3  : -------TGTTCCGCTGGCTGGCCATCTCCTACGCCTCGGCACACCTCACCATGACCGAGACGTTCCGCGGGGGC :  403
pgf1n.pk009.p24 : ------------------------------------------------------------------------------ :    -
040024.2        : ------------------------------------------------------------------------------ :    -

                         1360         *       1380         *       1400         *       1420
mouseAEBP1      : TGCCAGGCCCAGGACTACACCAGCGGCATGGGCATTGTCAACGGGGCCAAGTGGAATCCTCGCTCTGGGACTTTC : 1419
pgf1n.pk006.h24 : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18 : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7  : ------------------------------------------------------------------------------ :    -
pgm2n.pk007.g13 : TGCCACACGCAGGATGTCACCGAGGCCATGGGCATCGTGCAGGGGGCCNNGTGGCGGCCCCGCGCCGGCAGCATG :  622
pgf1n.pk006.o3  : TGCCACACGCAGGATGTCACCGAGGCCATGGGCATCGTGCAGGGGGCCAAGTGGCGGCCCCGCGCCGGCAGCATG :  478
pgf1n.pk009.p24 : ----------------------------------------------------------GGACGCGTGGGG :   13
040024.2        : ---------------------------------------------------------------AGCATG :    6

                    *       1440         *       1460         *       1480         *       1500
mouseAEBP1      : AATGACTTTAGCTACCTGCACACAAACTGTCTGGAGCTCTCCGTATACCTGGGCTGTGACAAGTTCCCCCACGAG : 1494
pgf1n.pk006.h24 : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18 : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7  : ------------------------------------------------------------------------------ :    -
pgm2n.pk007.g13 : AATGACTTCAGCTACCTGCACACCAACT------------------------------------------------- :  650
pgf1n.pk006.o3  : AATGACTTCAGCTACCTGCACACCAACTGCCTGGAGCTCTCCGTCTACCTGGGCTGCGACAAGTTCCCCCATGAG :  553
pgf1n.pk009.p24 : AATGACTTCAGCTACCTGCACACCAACTGCCTGGAGCTCTCCGTCTACCTGGGCTGCGACAAGTTCCCCCATGAG :   88
040024.2        : AATGACTTCAGCTACCTGCACACCAACTGCCTGGAGCTCTCCGTCTACCTGGGCTGCGACAAGTTCCCCCATGAG :   81
```

```
                          *      1520          *      1540          *      1560          *
mouseAEBP1       : AGTGAGCTACCCCGAGAATGGGAGAACAACAAAGAAGCGCTGCTCACCTTCATGGAGCAGGTGCACCGTGGCATT : 1569
pgf1n.pk006.h24  : ------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------- :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : AGCGAGCTGNN------------------------------------------------------------- :  564
pgf1n.pk009.p24  : AGCGAGCTGCAGCAGGAGTGGGAGAACAACAAAGAGTCACTGCTCACCTTCATGGAGCAGACCCACCGCGGGATT :  163
040024.2         : AGCGAGCTGCAGCAGGAGTGGGAGAACAACAAAGAGTCACTGCTCACCTTCATGGAGCAGACCCACCGCGGGATC :  156

                     1580          *      1600          *      1620          *      1640          *
mouseAEBP1       : AAGGGTGTGGTGACAGATGAGCAAGGCATCCCCATTGCCAATGCCACCATCTCTGTGAGTGGCATCAACCATGGT : 1644
pgf1n.pk006.h24  : ------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------- :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : AAAGGTCTGGTGACAGACCAGCAGGGAGAGCCCATCGCCAACGCCACCATCGTGGTGGGTGGCATCAAGCACAGC :  238
040024.2         : AAAGGTTTGGTGACAGACCAGCAGGGAGAGCCCATCGCCAACGCCACCATCGTGGTGGGCGGCATCAAGCACAGC :  231

                          1660          *      1680          *      1700          *      1720
mouseAEBP1       : GTGAAGACAGCAAGTGGAGGTGACTACTGGCGCATTCTGAACCCGGGTGAGTACCGTGTGACAGCTCACGCAGAG : 1719
pgf1n.pk006.h24  : ------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------- :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : GTCAGGACAGCCAGCGGTGGGGACTACTGGCGCATCCTGAACCCCGGTGAGTACCGCGTGTCGGCGCGGGCCGAG :  313
040024.2         : GTCAGGACAGCCAGCGGTGGGGACTACTGGCGCATCCTGAACCCCGGTGAGTACCGCGTGTCGGCGCGGGCCGAG :  306

                     *      1740          *      1760          *      1780          *      1800
mouseAEBP1       : GGCTACACCTCAAGTGCCAAGATCTGCAATGTGGACTACGATATTGGGGCCACTCAGTGCAACTTCATCCTGGCT : 1794
pgf1n.pk006.h24  : ------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------- :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : GGCTACAACCCCAGCACCAAGACCTGCAGCGTCTTCTACGACATCGGGGCCACCCAGTGCAACTTCGTGCTGGCC :  388
040024.2         : GGCTACAACCCCAGCACCAAGACCTGCAGCGTCTTCTACGACATCGGGGCCACCCAGTGCAACTTCGTGCTGGCC :  381

                          *      1820          *      1840          *      1860          *
mouseAEBP1       : CGATCCAACTGGAAGCGCATTCGGGAGATCTTGGCTATGAACGGGAACCGTCCCATTCTCCGAGTTGACCCCTCA : 1869
pgf1n.pk006.h24  : ------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------- :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : CGCTCCAACTGGAAGCGCATCCGTGAGATCATGGCCATGAACGGGAACCGACCCATCCGCCGCGTGGTGCCCGGC :  463
040024.2         : CGCTCCAACTGGAAGCGCATCCGCGAGATCATGGCCATGAACGGGAACCGACCCATCCGCCGCGTGGTGCCCGGC :  456

                     1880          *      1900          *      1920          *      1940          *
mouseAEBP1       : CGACCCATGACCCCCCAGCAGCGGCGCATGCAGCAGCGCCGTCTACAGTACCGGCTCCGCATGAGGGAACAGATG : 1944
pgf1n.pk006.h24  : ------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------- :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : CGCCCCATGACGCTCCGCGAGCGCCTGCGGCTGCGCCAGCGCCTGCGGCTCCGGCAGCGGCTCCGCGAGCGGATG :  538
040024.2         : CGCCCCATGACGCTCCGCGAGCGCCTGCGGCTGCGCCAGCGCCTGCGGCTCCGGCAGCGGCTCCGCGAGCGGATG :  531

                          1960          *      1980          *      2000          *      2020
mouseAEBP1       : CGACTGCGTCGCCTCAATTCTACCGCAGGCCCTGCCACAAGCCCCACTCCTGCCCTTATGCCTCCCCCTTCCCCT : 2019
pgf1n.pk006.h24  : ------------------------------------------------------------------------- :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------- :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------- :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------- :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------- :    -
pgf1n.pk009.p24  : NNNNTGCGGAGACTGAATGCCACCGCCAGCACCGCCAGCGTCCCCACGGCCCCCCCGCCCAGCACGGCGCTGCCC :  613
040024.2         : AGGCTGCGGAGACTGAATGCCACCGCCAGCACCGCCAGCGTCCCCACGGCCCCCCCGCCCAGCACGGCGCTGCCC :  606
```

```
                        *      2040          *      2060          *      2080          *      2100
mouseAEBP1       : ACACCAGCCATTACCTTGAGGCCCTGGGAAGTTCTACCCACTACCACTGCAGGCTGGGAGGAGTCAGAGACTGAG : 2094
pgf1n.pk006.h24  : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------------ :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------------ :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------------ :    -
pgf1n.pk009.p24  : GCCCCGTTCAGCAGCACCACGTATG---------------------------------------------------- :  638
040024.2         : GTCCCGTTCAGCAGCACCACGTATGCACCATGGAGCCGGGAGCCGCCCACAGCGGGCACCTGGGAGATGGAGACC :  681

                              *      2120          *      2140          *      2160          *
mouseAEBP1       : ACCTATACAGAAGTAGTGACAGAGTTTGAGACAGAGTATGGGACTGACCTAGAGGTGGAAGAGATAGAGGAGGAG : 2169
pgf1n.pk006.h24  : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------------ :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------------ :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------------ :    -
pgf1n.pk009.p24  : ------------------------------------------------------------------------------ :    -
040024.2         : GAAACGGAGGTGGTGACGGAGCTGGTGACGGTGACGGAGCTGTGGGATGCGGGCACAGGGACGGCGCAGCCCTTC :  756

                    2180          *      2200          *      2220          *      2240          *
mouseAEBP1       : GAGGAGGAGGAGGAGGAAGAGATGGACACAGGCCTTACATTTCCACTCACAACAGTGGAGACCTACACAGTGAAC : 2244
pgf1n.pk006.h24  : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------------ :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------------ :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------------ :    -
pgf1n.pk009.p24  : ------------------------------------------------------------------------------ :    -
040024.2         : ACCACCGCAGAGACCTACACTGTGAACTTTGGCGATTAGGGGACGCCCGTCCCCTTGCTCCCGCCTCCGGGTGCT :  831

                          2260          *      2280          *      2300          *      2320
mouseAEBP1       : TTTGGGGACTTCTGAGACTGGGATCTCAAAGCCCTGCCCAATTCAAACTAAGGCAGCACCTCCCAAGCCTGTGCC : 2319
pgf1n.pk006.h24  : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------------ :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------------ :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------------ :    -
pgf1n.pk009.p24  : ------------------------------------------------------------------------------ :    -
040024.2         : CCACGGGGGCACCATCCCGTGCTCCTGGAGCCCCTAGGCCAGGCACACAACACACAAGACACACAATCCCGTTGC :  906

                        *      2340          *      2360          *      2380          *      2400
mouseAEBP1       : AGCAGACACATAGCCATCAGATGTCCCTGGGTGGACCCCACTCCCCCAGTGTGGGACATCAAAGCTACCGGGACT : 2394
pgf1n.pk006.h24  : ------------------------------------------------------------------------------ :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------------------ :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------------------ :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------------------ :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------------------ :    -
pgf1n.pk009.p24  : ------------------------------------------------------------------------------ :    -
040024.2         : CCAGGCAGCTGGCGACGGCACTGCGGCTGGGCACGGAGGGAAGCAGGGCAGGGCCCTGGCGTGTTTATTTTGTTA :  981

                              *      2420          *      2440          *      2460
mouseAEBP1       : CTGCATAGACTCTGGTCTACCCGCCCCAGCTCTACCTGCCAGCCTTTGGGGAGGGGCAGGCAAGGA : 2460
pgf1n.pk006.h24  : ------------------------------------------------------------------ :    -
pgp1n.pk012.m18  : ------------------------------------------------------------------ :    -
pgf2n.pk001.c7   : ------------------------------------------------------------------ :    -
pgm2n.pk007.g13  : ------------------------------------------------------------------ :    -
pgf1n.pk006.o3   : ------------------------------------------------------------------ :    -
pgf1n.pk009.p24  : ------------------------------------------------------------------ :    -
040024.2         : CAAATAAACAAAGCTGTATTGG------------------------------------------- : 1003
```

REFERENCES

(1) E. Pennisi. (2004) Searching for the genome's second code. *Science 306*, 632-5.

(2) G. P. He, Muise, A., Li, A. W. and Ro, H. S. (1995) A eukaryotic transcriptional repressor with carboxypeptidase activity. *Nature 378*, 92-6.

(3) J. G. Park, Muise, A., He, G. P., Kim, S. W. and Ro, H. S. (1999) Transcriptional regulation by the gamma5 subunit of a heterotrimeric G protein during adipogenesis. *EMBO J 18*, 4004-12.

(4) M. D. Layne, Endege, W. O., Jain, M. K., Yet, S. F., Hsieh, C. M., Chin, M. T., Perrella, M. A., Blanar, M. A., Haber, E. and Lee, M. E. (1998) Aortic carboxypeptidase-like protein, a novel protein with discoidin and carboxypeptidase-like domains, is up-regulated during vascular smooth muscle cell differentiation. *J Biol Chem 273*, 15654-60.

(5) M. D. Layne, Yet, S. F., Maemura, K., Hsieh, C. M., Bernfield, M., Perrella, M. A. and Lee, M. E. (2001) Impaired abdominal wall development and deficient wound healing in mice lacking aortic carboxypeptidase-like protein. *Mol Cell Biol 21*, 5256-61.

(6) A. Gagnon, Abaiian, K. J., Crapper, T., Layne, M. D. and Sorisky, A. (2002) Down-regulation of aortic carboxypeptidase-like protein during the early phase of 3T3-L1 adipogenesis. *Endocrinology 143*, 2478-85.

(7) S. W. Kim, Muise, A. M., Lyons, P. J. and Ro, H. S. (2001) Regulation of adipogenesis by a transcriptional repressor that modulates MAPK activation. *J Biol Chem 276*, 10199-206.

(8) P. J. Lyons, Muise, A. M. and Ro, H. S. (2005) MAPK modulates the DNA binding of adipocyte enhancer-binding protein 1. *Biochemistry 44*, 926-31.

(9) G. A. Bray. (2003) Risks of obesity. *Endocrinol Metab Clin North Am 32*, 787-804, viii.

(10) J. M. Ntambi and Young-Cheul, K. (2000) Adipocyte differentiation and gene expression. *J Nutr 130*, 3122S-3126S.

(11) A. Soukas, Socci, N. D., Saatkamp, B. D., Novelli, S. and Friedman, J. M. (2001) Distinct transcriptional profiles of adipogenesis in vivo and in vitro. *J Biol Chem 276*, 34167-74.

(12) L. Fajas, Fruchart, J. C. and Auwerx, J. (1998) Transcriptional control of adipogenesis. *Curr Opin Cell Biol 10*, 165-73.

(13) F. M. Gregoire, Smas, C. M. and Sul, H. S. (1998) Understanding adipocyte differentiation. *Physiol Rev 78*, 783-809.

(14) E. D. Rosen, Walkey, C. J., Puigserver, P. and Spiegelman, B. M. (2000) Transcriptional regulation of adipogenesis. *Genes Dev 14*, 1293-307.

(15) S. R. Farmer. (2005) Regulation of PPARgamma activity during adipogenesis. *Int J Obes Relat Metab Disord 29 Suppl 1*, S13-6.

(16) B. M. Spiegelman and Flier, J. S. (1996) Adipogenesis and obesity: rounding out the big picture. *Cell 87*, 377-89.

(17) P. Tontonoz, Hu, E. and Spiegelman, B. M. (1994) Stimulation of adipogenesis in fibroblasts by PPAR gamma 2, a lipid-activated transcription factor. *Cell 79*, 1147-56.

(18) Z. Wu, Rosen, E. D., Brun, R., Hauser, S., Adelmant, G., Troy, A. E., McKeon, C., Darlington, G. J. and Spiegelman, B. M. (1999) Cross-regulation of C/EBP alpha and PPAR gamma controls the transcriptional pathway of adipogenesis and insulin sensitivity. *Mol Cell 3*, 151-8.

(19) E. D. Rosen, Hsu, C. H., Wang, X., Sakai, S., Freeman, M. W., Gonzalez, F. J. and Spiegelman, B. M. (2002) C/EBPalpha induces adipogenesis through PPARgamma: a unified pathway. *Genes Dev 16*, 22-6.

(20) E. D. Rosen and Spiegelman, B. M. (2001) PPARgamma: a nuclear regulator of metabolism, differentiation, and cell growth. *J Biol Chem 276*, 37731-4.

(21) D. Ren, Collingwood, T. N., Rebar, E. J., Wolffe, A. P. and Camp, H. S. (2002) PPARgamma knockdown by engineered transcription factors: exogenous PPARgamma2 but not PPARgamma1 reactivates adipogenesis. *Genes Dev 16*, 27-32.

(22) S. M. Rangwala and Lazar, M. A. (2004) Peroxisome proliferator-activated receptor gamma in diabetes and metabolism. *Trends Pharmacol Sci 25*, 331-6.

(23) D. L. Gerhold, Liu, F., Jiang, G., Li, Z., Xu, J., Lu, M., Sachs, J. R., Bagchi, A., Fridman, A., Holder, D. J., Doebber, T. W., Berger, J., Elbrecht, A., Moller, D. E. and Zhang, B. B. (2002) Gene expression profile of adipocyte differentiation and its regulation by peroxisome proliferator-activated receptor-gamma agonists. *Endocrinology 143*, 2106-18.

(24) T. Albrektsen, Frederiksen, K. S., Holmes, W. E., Boel, E., Taylor, K. and Fleckner, J. (2002) Novel genes regulated by the insulin sensitizer rosiglitazone during adipocyte differentiation. *Diabetes 51*, 1042-51.

(25) P. Tontonoz, Hu, E., Devine, J., Beale, E. G. and Spiegelman, B. M. (1995) PPAR gamma 2 regulates adipose expression of the phosphoenolpyruvate carboxykinase gene. *Mol Cell Biol 15*, 351-7.

(26) M. Armoni, Kritz, N., Harel, C., Bar-Yoseph, F., Chen, H., Quon, M. J. and Karnieli, E. (2003) Peroxisome proliferator-activated receptor-gamma represses GLUT4 promoter activity in primary adipocytes, and rosiglitazone alleviates this effect. *J Biol Chem 278*, 30614-23.

(27) D. Kramer, Shapiro, R., Adler, A., Bush, E. and Rondinone, C. M. (2001) Insulin-sensitizing effect of rosiglitazone (BRL-49653) by regulation of glucose transporters in muscle and fat of Zucker rats. *Metabolism 50*, 1294-300.

(28) J. M. Way, Gorgun, C. Z., Tong, Q., Uysal, K. T., Brown, K. K., Harrington, W. W., Oliver, W. R., Jr., Willson, T. M., Kliewer, S. A. and Hotamisligil, G. S. (2001) Adipose tissue resistin expression is severely suppressed in obesity and stimulated by peroxisome proliferator-activated receptor gamma agonists. *J Biol Chem 276*, 25651-3.

(29) P. Tontonoz, Hu, E., Graves, R. A., Budavari, A. I. and Spiegelman, B. M. (1994) mPPAR gamma 2: tissue-specific regulator of an adipocyte enhancer. *Genes Dev 8*, 1224-34.

(30) R. Herrera, Ro, H. S., Robinson, G. S., Xanthopoulos, K. G. and Spiegelman, B. M. (1989) A direct role for C/EBP and the AP-I-binding site in gene expression linked to adipocyte differentiation. *Mol Cell Biol 9*, 5331-9.

(31) H. S. Ro and Roncari, D. A. (1991) The C/EBP-binding region and adjacent sites regulate expression of the adipose P2 gene in human preadipocytes. *Mol Cell Biol 11*, 2303-6.

(32) H. S. Ro, Kim, S. W., Wu, D., Webber, C. and Nicholson, T. E. (2001) Gene structure and expression of the mouse adipocyte enhancer-binding protein. *Gene 280*, 123-33.

(33) M. Benito, Porras, A., Nebreda, A. R. and Santos, E. (1991) Differentiation of 3T3-L1 fibroblasts to adipocytes induced by transfection of ras oncogenes. *Science 253*, 565-8.

(34) E. M. Sale, Atkinson, P. G. and Sale, G. J. (1995) Requirement of MAP kinase for differentiation of fibroblasts to adipocytes, for insulin activation of p90 S6 kinase and for insulin or serum stimulation of DNA synthesis. *EMBO J 14*, 674-84.

(35) J. Font de Mora, Porras, A., Ahn, N. and Santos, E. (1997) Mitogen-activated protein kinase activation is not necessary for, but antagonizes, 3T3-L1 adipocytic differentiation. *Mol Cell Biol 17*, 6068-75.

(36) H. S. Camp and Tafuri, S. R. (1997) Regulation of peroxisome proliferator-activated receptor gamma activity by mitogen-activated protein kinase. *J Biol Chem 272*, 10811-6.

(37) E. Hu, Kim, J. B., Sarraf, P. and Spiegelman, B. M. (1996) Inhibition of adipogenesis through MAP kinase-mediated phosphorylation of PPARgamma. *Science 274*, 2100-3.

(38) Q. Q. Tang, Otto, T. C. and Lane, M. D. (2003) Mitotic clonal expansion: a synchronous process required for adipogenesis. *Proc Natl Acad Sci U S A 100*, 44-9.

(39) J. W. Edmunds and Mahadevan, L. C. (2004) MAP kinases as structural adaptors and enzymatic activators in transcription complexes. *J Cell Sci 117*, 3715-23.

(40) A. D. Bradshaw, Graves, D. C., Motamed, K. and Sage, E. H. (2003) SPARC-null mice exhibit increased adiposity without significant differences in overall body weight. *Proc Natl Acad Sci U S A 100*, 6045-50.

(41) A. Abderrahim-Ferkoune, Bezy, O., Astri-Roques, S., Elabd, C., Ailhaud, G. and Amri, E. Z. (2004) Transdifferentiation of preadipose cells into smooth muscle-like cells: role of aortic carboxypeptidase-like protein. *Exp Cell Res 293*, 219-28.

(42) A. Gagnon, Landry, A., Proulx, J., Layne, M. D. and Sorisky, A. (2005) Aortic carboxypeptidase-like protein is regulated by transforming growth factor beta in 3T3-L1 preadipocytes. *Exp Cell Res.*

(43) M. D. Layne, Yet, S. F., Maemura, K., Hsieh, C. M., Liu, X., Ith, B., Lee, M. E. and Perrella, M. A. (2002) Characterization of the mouse aortic carboxypeptidase-like protein promoter reveals activity in differentiated and dedifferentiated vascular smooth muscle cells. *Circ Res 90*, 728-36.

(44) Q. Zhou, Clarke, L., Nie, R., Carnes, K., Lai, L. W., Lien, Y. H., Verkman, A., Lubahn, D., Fisher, J. S., Katzenellenbogen, B. S. and Hess, R. A. (2001) Estrogen action and male fertility: roles of the sodium/hydrogen exchanger-3 and fluid reabsorption in reproductive tract function. *Proc Natl Acad Sci U S A 98*, 14132-7.

(45) S. Hahn. (2004) Structure and mechanism of the RNA polymerase II transcription machinery. *Nat Struct Mol Biol 11*, 394-403.

(46) G. Lazennec, Canaple, L., Saugy, D. and Wahli, W. (2000) Activation of peroxisome proliferator-activated receptors (PPARs) by their ligands and protein kinase A activators. *Mol Endocrinol 14*, 1962-75.

(47) A. Flaus and Owen-Hughes, T. (2004) Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? *Curr Opin Genet Dev 14*, 165-73.

(48) A. Saha, Wittmeyer, J. and Cairns, B. R. (2002) Chromatin remodeling by RSC involves ATP-dependent DNA translocation. *Genes Dev 16*, 2120-34.

(49) G. Mizuguchi, Shen, X., Landry, J., Wu, W. H., Sen, S. and Wu, C. (2004) ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science 303*, 343-8.

(50) T. E. Wagner, Hartford, J. B., Serra, M., Vandegrift, V. and Sung, M. T. (1977) Phosphorylation and dephosphorylation of histone (V (H5): controlled condensation of avian erythrocyte chromatin. Appendix: Phosphorylation and dephosphorylation of histone H5. II. Circular dichroic studies. *Biochemistry 16*, 286-90.

(51) C. Costanzi and Pehrson, J. R. (1998) Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature 393*, 599-601.

(52) A. G. Ladurner. (2003) Inactivating chromosomes: a macro domain that minimizes transcription. *Mol Cell 12*, 1-3.

(53) M. D. Allen, Buckle, A. M., Cordell, S. C., Lowe, J. and Bycroft, M. (2003) The crystal structure of AF1521 a protein from Archaeoglobus fulgidus with homology to the non-histone domain of macroH2A. *J Mol Biol 330*, 503-11.

(54) D. Angelov, Molla, A., Perche, P. Y., Hans, F., Cote, J., Khochbin, S., Bouvet, P. and Dimitrov, S. (2003) The histone variant macroH2A interferes with transcription factor binding and SWI/SNF nucleosome remodeling. *Mol Cell 11*, 1033-41.

(55) B. P. Chadwick and Willard, H. F. (2001) A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J Cell Biol 152*, 375-84.

(56) D. Angelov, Verdel, A., An, W., Bondarenko, V., Hans, F., Doyen, C. M., Studitsky, V. M., Hamiche, A., Roeder, R. G., Bouvet, P. and Dimitrov, S. (2004) SWI/SNF remodeling and p300-dependent transcription of histone variant H2ABbd nucleosomal arrays. *Embo J 23*, 3815-24.

(57) R. T. Kamakaka and Biggins, S. (2005) Histone variants: deviants? *Genes Dev 19*, 295-310.

(58) E. McKittrick, Gafken, P. R., Ahmad, K. and Henikoff, S. (2004) Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc Natl Acad Sci U S A 101*, 1525-30.

(59) C. M. Chow, Georgiou, A., Szutorisz, H., Maia e Silva, A., Pombo, A., Barahona, I., Dargelos, E., Canzonetta, C. and Dillon, N. (2005) Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. *EMBO Rep 6*, 354-60.

(60) K. Luger, Mader, A. W., Richmond, R. K., Sargent, D. F. and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature 389*, 251-60.

(61) B. M. Turner. (2002) Cellular memory and the histone code. *Cell 111*, 285-91.

(62) X. de la Cruz, Lois, S., Sanchez-Molina, S. and Martinez-Balbas, M. A. (2005) Do protein motifs read the histone code? *Bioessays 27*, 164-75.

(63) J. E. Brownell and Allis, C. D. (1995) An activity gel assay detects a single, catalytically active histone acetyltransferase subunit in Tetrahymena macronuclei. *Proc Natl Acad Sci U S A 92*, 6364-8.

(64) J. E. Brownell, Zhou, J., Ranalli, T., Kobayashi, R., Edmondson, D. G., Roth, S. Y. and Allis, C. D. (1996) Tetrahymena histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell 84*, 843-51.

(65) J. Taunton, Hassig, C. A. and Schreiber, S. L. (1996) A mammalian histone deacetylase related to the yeast transcriptional regulator Rpd3p. *Science 272*, 408-11.

(66) C. L. Peterson and Laniel, M. A. (2004) Histones and histone modifications. *Curr Biol 14*, R546-51.

(67) J. Wu and Grunstein, M. (2000) 25 years after the nucleosome model: chromatin modifications. *Trends Biochem Sci 25*, 619-23.

(68) P. Puigserver, Wu, Z., Park, C. W., Graves, R., Wright, M. and Spiegelman, B. M. (1998) A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell 92*, 829-39.

(69) A. E. Wallberg, Yamamura, S., Malik, S., Spiegelman, B. M. and Roeder, R. G. (2003) Coordination of p300-mediated chromatin remodeling and TRAP/mediator function through coactivator PGC-1alpha. *Mol Cell 12*, 1137-49.

(70) C. Yu, Markan, K., Temple, K. A., Deplewski, D., Brady, M. J. and Cohen, R. N. (2005) The nuclear receptor corepressors NCoR and SMRT decrease peroxisome proliferator-activated receptor gamma transcriptional activity and repress 3T3-L1 adipogenesis. *J Biol Chem 280*, 13600-5.

(71) K. Jepsen and Rosenfeld, M. G. (2002) Biological roles and mechanistic actions of co-repressor complexes. *J Cell Sci 115*, 689-98.

(72) R. L. Erickson, Hemati, N., Ross, S. E. and MacDougald, O. A. (2001) p300 coactivates the adipogenic transcription factor CCAAT/enhancer-binding protein alpha. *J Biol Chem 276*, 16348-55.

(73) D. Chen, Ma, H., Hong, H., Koh, S. S., Huang, S. M., Schurter, B. T., Aswad, D. W. and Stallcup, M. R. (1999) Regulation of transcription by a protein methyltransferase. *Science 284*, 2174-7.

(74) S. Rea, Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D. and Jenuwein, T. (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature 406*, 593-9.

(75) Y. Zhang and Reinberg, D. (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev 15*, 2343-60.

(76) K. Nishioka, Chuikov, S., Sarma, K., Erdjument-Bromage, H., Allis, C. D., Tempst, P. and Reinberg, D. (2002) Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. *Genes Dev 16*, 479-89.

(77) G. L. Cuthbert, Daujat, S., Snowden, A. W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P. D., Tempst, P., Bannister, A. J. and Kouzarides, T. (2004) Histone deimination antagonizes arginine methylation. *Cell 118*, 545-53.

(78) Y. Wang, Wysocka, J., Sayegh, J., Lee, Y. H., Perlin, J. R., Leonelli, L., Sonbuchner, L. S., McDonald, C. H., Cook, R. G., Dou, Y., Roeder, R. G., Clarke, S., Stallcup, M. R., Allis, C. D. and Coonrod, S. A. (2004) Human PAD4 regulates histone arginine methylation levels via demethylimination. *Science 306*, 279-83.

(79) Y. Shi, Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A. and Shi, Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell 119*, 941-53.

(80) G. Gill. (2004) SUMO and ubiquitin in the nucleus: different functions, similar mechanisms? *Genes Dev 18*, 2046-59.

(81) L. Levinger and Varshavsky, A. (1982) Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the Drosophila genome. *Cell 28*, 375-85.

(82) H. Wang, Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S. and Zhang, Y. (2004) Role of histone H2A ubiquitination in Polycomb silencing. *Nature 431*, 873-8.

(83) Z. W. Sun and Allis, C. D. (2002) Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature 418*, 104-8.

(84) W. W. Hwang, Venkatasubrahmanyam, S., Ianculescu, A. G., Tong, A., Boone, C. and Madhani, H. D. (2003) A conserved RING finger protein required for histone H2B monoubiquitination and cell size control. *Mol Cell 11*, 261-6.

(85) H. H. Ng, Xu, R. M., Zhang, Y. and Struhl, K. (2002) Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79. *J Biol Chem 277*, 34655-7.

(86) Y. Shiio and Eisenman, R. N. (2003) Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A 100*, 13225-30.

(87) L. R. Gurley, D'Anna, J. A., Barham, S. S., Deaven, L. L. and Tobey, R. A. (1978) Histone phosphorylation and chromatin structure during mitosis in Chinese hamster cells. *Eur J Biochem 84*, 1-15.

(88) S. J. Nowak and Corces, V. G. (2004) Phosphorylation of histone H3: a balancing act between chromosome condensation and transcriptional activation. *Trends Genet 20*, 214-20.

(89) W. S. Lo, Duggan, L., Emre, N. C., Belotserkovskya, R., Lane, W. S., Shiekhattar, R. and Berger, S. L. (2001) Snf1--a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription. *Science 293*, 1142-6.

(90) C. C. Chen, Bruegger, B. B., Kern, C. W., Lin, Y. C., Halpern, R. M. and Smith, R. A. (1977) Phosphorylation of nuclear proteins in rat regenerating liver. *Biochemistry 16*, 4852-5.

(91) D. L. Smith, Chen, C. C., Bruegger, B. B., Holtz, S. L., Halpern, R. M. and Smith, R. A. (1974) Characterization of protein kinases forming acid-labile histone phosphates in Walker-256 carcinosarcoma cell nuclei. *Biochemistry 13*, 3780-5.

(92) D. L. Smith, Bruegger, B. B., Halpern, R. M. and Smith, R. A. (1973) New histone kinases in nuclei of rat tissues. *Nature 246*, 103-4.

(93) C. C. Chen, Smith, D. L., Bruegger, B. B., Halpern, R. M. and Smith, R. A. (1974) Occurrence and distribution of acid-labile histone phosphates in regenerating rat liver. *Biochemistry 13*, 3785-9.

(94) J. M. Huang, Wei, Y. F., Kim, Y. H., Osterberg, L. and Matthews, H. R. (1991) Purification of a protein histidine kinase from the yeast Saccharomyces cerevisiae. The first member of this class of protein kinases. *J Biol Chem 266*, 9023-31.

(95) V. D. Huebner and Matthews, H. R. (1985) Phosphorylation of histidine in proteins by a nuclear extract of Physarum polycephalum plasmodia. *J Biol Chem 260*, 16106-13.

(96) P. S. Steeg, Palmieri, D., Ouatas, T. and Salerno, M. (2003) Histidine kinases and histidine phosphorylated proteins in mammalian cell biology, signal transduction and cancer. *Cancer Lett 190*, 1-12.

(97) E. Tan, Besant, P. G. and Attwood, P. V. (2002) Mammalian histidine kinases: do they REALLY exist? *Biochemistry 41*, 3843-51.

(98) F. Levy-Favatier, Delpech, M. and Kruh, J. (1987) Characterization of an arginine-specific protein kinase tightly bound to rat liver DNA. *Eur J Biochem 166*, 617-21.

(99) B. T. Wakim, Picken, M. M. and DeLange, R. J. (1990) Identification and partial purification of a chromatin bound calmodulin activated histone 3 kinase from calf thymus. *Biochem Biophys Res Commun 171*, 84-90.

(100) B. T. Wakim and Aswad, G. D. (1994) Ca(2+)-calmodulin-dependent phosphorylation of arginine in histone 3 by a nuclear kinase from mouse leukemia cells. *J Biol Chem 269*, 2722-7.

(101) B. T. Wakim, Grutkoski, P. S., Vaughan, A. T. and Engelmann, G. L. (1995) Stimulation of a Ca(2+)-calmodulin-activated histone 3 arginine kinase in quiescent rat heart endothelial cells compared to actively dividing cells. *J Biol Chem 270*, 23155-8.

(102) J. P. Th'ng, Sung, R., Ye, M. and Hendzel, M. J. (2005) H1 family histones in the nucleus: Control of binding and localization by the C-terminal domain. *J Biol Chem.*

(103) B. T. Schurter, Koh, S. S., Chen, D., Bunick, G. J., Harp, J. M., Hanson, B. L., Henschen-Edman, A., Mackay, D. R., Stallcup, M. R. and Aswad, D. W. (2001) Methylation of histone H3 by coactivator-associated arginine methyltransferase 1. *Biochemistry 40*, 5747-56.

(104) N. M. Hooper. (1994) Families of zinc metalloproteases. *FEBS Lett 354*, 1-6.

(105) H. M. Kim, Shin, D. R., Yoo, O. J., Lee, H. and Lee, J. O. (2003) Crystal structure of Drosophila angiotensin I-converting enzyme bound to captopril and lisinopril. *FEBS Lett 538*, 65-70.

(106) R. Natesh, Schwager, S. L., Sturrock, E. D. and Acharya, K. R. (2003) Crystal structure of the human angiotensin-converting enzyme-lisinopril complex. *Nature 421*, 551-4.

(107) R. Natesh, Schwager, S. L., Evans, H. R., Sturrock, E. D. and Acharya, K. R. (2004) Structural details on the binding of antihypertensive drugs captopril and enalaprilat to human testicular angiotensin I-converting enzyme. *Biochemistry 43*, 8718-24.

(108) P. Towler, Staker, B., Prasad, S. G., Menon, S., Tang, J., Parsons, T., Ryan, D., Fisher, M., Williams, D., Dales, N. A., Patane, M. A. and Pantoliano, M. W. (2004) ACE2 X-ray structures reveal a large hinge-bending motion important for inhibitor binding and catalysis. *J Biol Chem 279*, 17996-8007.

(109) J. Vendrell, Querol, E. and Aviles, F. X. (2000) Metallocarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties. *Biochim Biophys Acta 1477*, 284-98.

(110) I. Garcia-Saez, Reverter, D., Vendrell, J., Aviles, F. X. and Coll, M. (1997) The three-dimensional structure of human procarboxypeptidase A2. Deciphering the basis of the inhibition, activation and intrinsic activity of the zymogen. *EMBO J 16*, 6906-13.

(111) S. R. Nalamachu, Song, L. and Fricker, L. D. (1994) Regulation of carboxypeptidase E. Effect of Ca2+ on enzyme activity and stability. *J Biol Chem 269*, 11192-5.

(112) Y. Qian, Varlamov, O. and Fricker, L. D. (1999) Glu300 of rat carboxypeptidase E is essential for enzymatic activity but not substrate binding or routing to the regulated secretory pathway. *J Biol Chem 274*, 11582-6.

(113) F. X. Gomis-Ruth, Companys, V., Qian, Y., Fricker, L. D., Vendrell, J., Aviles, F. X. and Coll, M. (1999) Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallocarboxypeptidase subfamily. *EMBO J 18*, 5817-26.

(114) D. Reverter, Maskos, K., Tan, F., Skidgel, R. A. and Bode, W. (2004) Crystal structure of human carboxypeptidase M, a membrane-bound enzyme that regulates peptide hormone activity. *J Mol Biol 338*, 257-69.

(115) Y. Lei, Xin, X., Morgan, D., Pintar, J. E. and Fricker, L. D. (1999) Identification of mouse CPX-1, a novel member of the metallocarboxypeptidase gene family with highest similarity to CPX-2. *DNA Cell Biol 18*, 175-85.

(116) X. Xin, Day, R., Dong, W., Lei, Y. and Fricker, L. D. (1998) Identification of mouse CPX-2, a novel member of the metallocarboxypeptidase gene family: cDNA cloning, mRNA distribution, and protein expression and characterization. *DNA Cell Biol 17*, 897-909.

(117) F. Tan, Balsitis, S., Black, J. K., Blochl, A., Mao, J. F., Becker, R. P., Schacht, D. and Skidgel, R. A. (2003) Effect of mutation of two critical glutamic acid residues on the activity and stability of human carboxypeptidase M and characterization of its signal for glycosylphosphatidylinositol anchoring. *Biochem J 370*, 567-78.

(118) L. Song and Fricker, L. D. (1997) Cloning and expression of human carboxypeptidase Z, a novel metallocarboxypeptidase. *J Biol Chem 272*, 10543-50.

(119) G. N. Reeke, Hartsuck, J. A., Ludwig, M. L., Quiocho, F. A., Steitz, T. A. and Lipscomb, W. N. (1967) The Structure of Carboxypeptidase A, VI. Some Results at 2.0-angstrom Resolution, and the Complex with Glycyl-Tyrosine at 2.8- angstrom Resolution. *Proc Natl Acad Sci U S A 58*, 2220-2226.

(120) L. Stryer (1995) *Biochemistry*, Vol., Fourth ed., W. H. Freeman and Company, New York.

(121) K. W. Matthews, Mueller-Ortiz, S. L. and Wetsel, R. A. (2004) Carboxypeptidase N: a pleiotropic regulator of inflammation. *Mol Immunol 40*, 785-93.

(122) L. D. Fricker and Leiter, E. H. (1999) Peptides, enzymes and obesity: new insights from a 'dead' enzyme. *Trends Biochem Sci 24*, 390-3.

(123) N. X. Cawley, Zhou, J., Hill, J. M., Abebe, D., Romboz, S., Yanik, T., Rodriguiz, R. M., Wetsel, W. C. and Loh, Y. P. (2004) The carboxypeptidase E knockout mouse exhibits endocrinological and behavioral deficits. *Endocrinology 145*, 5807-19.

(124) F. Tan, Rehli, M., Krause, S. W. and Skidgel, R. A. (1997) Sequence of human carboxypeptidase D reveals it to be a member of the regulatory carboxypeptidase family with three tandem active site domains. *Biochem J 327 (Pt 1)*, 81-7.

(125) E. G. Novikova, Eng, F. J., Yan, L., Qian, Y. and Fricker, L. D. (1999) Characterization of the enzymatic properties of the first and second domains of metallocarboxypeptidase D. *J Biol Chem 274*, 28887-92.

(126) S. E. Reznik and Fricker, L. D. (2001) Carboxypeptidases from A to Z: implications in embryonic development and Wnt binding. *Cell Mol Life Sci 58*, 1790-804.

(127) P. Aloy, Companys, V., Vendrell, J., Aviles, F. X., Fricker, L. D., Coll, M. and Gomis-Ruth, F. X. (2001) The crystal structure of the inhibitor-complexed carboxypeptidase D domain II and the modeling of regulatory carboxypeptidases. *J Biol Chem 276*, 16177-84.

(128) C. K. Too, Vickaryous, N., Boudreau, R. T. and Sangster, S. M. (2001) Identification and nuclear localization of a novel prolactin and cytokine-responsive carboxypeptidase D. *Endocrinology 142*, 1357-67.

(129) M. Rehli, Krause, S. W. and Andreesen, R. (2000) The membrane-bound ectopeptidase CPM as a marker of macrophage maturation in vitro and in vivo. *Adv Exp Med Biol 477*, 205-16.

(130) R. A. Skidgel, Davis, R. M. and Tan, F. (1989) Human carboxypeptidase M. Purification and characterization of a membrane-bound carboxypeptidase that cleaves peptide hormones. *J Biol Chem 264*, 2236-41.

(131) E. G. Novikova, Reznik, S. E., Varlamov, O. and Fricker, L. D. (2000) Carboxypeptidase Z is present in the regulated secretory pathway and extracellular matrix in cultured cells and in human tissues. *J Biol Chem 275*, 4865-70.

(132) C. Moeller, Swindell, E. C., Kispert, A. and Eichele, G. (2003) Carboxypeptidase Z (CPZ) modulates Wnt signaling and regulates the development of skeletal elements in the chicken. *Development 130*, 5103-11.

(133) E. J. Chang, Kwak, H. B., Kim, H., Park, J. C., Lee, Z. H. and Kim, H. H. (2004) Elucidation of CPX-1 involvement in RANKL-induced osteoclastogenesis by a proteomics approach. *FEBS Lett 564*, 166-70.

(134) F. J. Eng, Novikova, E. G., Kuroki, K., Ganem, D. and Fricker, L. D. (1998) gp180, a protein that binds duck hepatitis B virus particles, has metallocarboxypeptidase D-like enzymatic activity. *J Biol Chem 273*, 8382-8.

(135) A. Harris, Morgan, J. I., Pecot, M., Soumare, A., Osborne, A. and Soares, H. D. (2000) Regenerating motor neurons express Nna1, a novel ATP/GTP-binding protein related to zinc carboxypeptidases. *Mol Cell Neurosci 16*, 578-96.

(136) A. Fernandez-Gonzalez, La Spada, A. R., Treadaway, J., Higdon, J. C., Harris, B. S., Sidman, R. L., Morgan, J. I. and Zuo, J. (2002) Purkinje cell degeneration (pcd) phenotypes caused by mutations in the axotomy-induced gene, Nna1. *Science 295*, 1904-6.

(137) R. J. Mullen, Eicher, E. M. and Sidman, R. L. (1976) Purkinje cell degeneration, a new neurological mutation in the mouse. *Proc Natl Acad Sci U S A 73*, 208-12.

(138) S. Macedo-Ribeiro, Bode, W., Huber, R., Quinn-Allen, M. A., Kim, S. W., Ortel, T. L., Bourenkov, G. P., Bartunik, H. D., Stubbs, M. T., Kane, W. H. and Fuentes-Prior, P. (1999) Crystal structures of the membrane-binding C2 domain of human coagulation factor V. *Nature 402*, 434-9.

(139) K. P. Pratt, Shen, B. W., Takeshima, K., Davie, E. W., Fujikawa, K. and Stoddard, B. L. (1999) Structure of the C2 domain of human factor VIII at 1.5 A resolution. *Nature 402*, 439-42.

(140) S. F. Altschul, Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res 25*, 3389-402.

(141) A. Fiser, Do, R. K. and Sali, A. (2000) Modeling of loops in protein structures. *Protein Sci 9*, 1753-73.

(142) A. Sali and Blundell, T. L. (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol 234*, 779-815.

(143) M. J. Sippl. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins 17*, 355-62.

(144) R. A. Laskowski, MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst. 26*, 283-291.

(145) A. A. Canutescu, Shelenkov, A. A. and Dunbrack, R. L., Jr. (2003) A graph-theory algorithm for rapid protein side-chain prediction. *Protein Sci 12*, 2001-14.

(146) P. J. Kraulis. (1991) MOLSCRIPT: A Program to Produce Both Detailed and Schematic Plots of Protein Structures. *J. Appl. Cryst. 24*, 946-950.

(147) E. A. Merritt and Bacon, D. J. (1997) Raster3D Photorealistic Molecular Graphics. *Meth. Enzymol. 277*, 505-524.

(148) J. D. Thompson, Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*, 4673-80.

(149) K. B. Nicholas, Nicholas H.B. Jr., and Deerfield, D.W. II. (1997) GeneDoc: Analysis and Visualization of Genetic Variation. *EMBNEW.NEWS 4.14.*

(150) R. D. Page. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci 12*, 357-8.

(151) M. B. Yaffe, Leparc, G. G., Lai, J., Obata, T., Volinia, S. and Cantley, L. C. (2001) A motif-based profile scanning approach for genome-wide prediction of signaling pathways. *Nat Biotechnol 19*, 348-53.

(152) B. Rost and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J Mol Biol 232*, 584-99.

(153) J. A. Cuff, Clamp, M. E., Siddiqui, A. S., Finlay, M. and Barton, G. J. (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics 14*, 892-3.

(154) G. N. Gill, Chen, W. S., Lazar, C. S., Glenney, J. R., Jr., Wiley, H. S., Ingraham, H. A. and Rosenfeld, M. G. (1988) Role of intrinsic protein tyrosine kinase in function and metabolism of the epidermal growth factor receptor. *Cold Spring Harb Symp Quant Biol 53 Pt 1*, 467-76.

(155) A. Hoffmann, Chiang, C. M., Oelgeschlager, T., Xie, X., Burley, S. K., Nakatani, Y. and Roeder, R. G. (1996) A histone octamer-like structure within TFIID. *Nature 380*, 356-9.

(156) M. P. Myers, Stolarov, J. P., Eng, C., Li, J., Wang, S. I., Wigler, M. H., Parsons, R. and Tonks, N. K. (1997) P-TEN, the tumor suppressor from human chromosome 10q23, is a dual-specificity phosphatase. *Proc Natl Acad Sci U S A 94*, 9052-7.

(157) M. M. Bradford. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem 72*, 248-54.

(158) C. M. Gorman, Moffat, L. F. and Howard, B. H. (1982) Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol Cell Biol 2*, 1044-51.

(159) D. Baker and Sali, A. (2001) Protein structure prediction and structural genomics. *Science 294*, 93-6.

(160) B. Contreras-Moreira, Fitzjohn, P. W. and Bates, P. A. (2002) Comparative modelling: an essential methodology for protein structure prediction in the post-genomic era. *Appl Bioinformatics 1*, 177-90.

(161) M. Saraste, Sibbald, P. R. and Wittinghofer, A. (1990) The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci 15*, 430-4.

(162) T. Duda, Goraczniak, R. M. and Sharma, R. K. (1993) The glycine residue of ATP regulatory module in receptor guanylate cyclases that is essential in natriuretic factor signaling. *FEBS Lett 335*, 309-14.

(163) R. K. Sharma, Yadav, P. and Duda, T. (2001) Allosteric regulatory step and configuration of the ATP-binding pocket in atrial natriuretic factor receptor guanylate cyclase transduction mechanism. *Can J Physiol Pharmacol 79*, 682-91.

(164) A. M. Muise and Ro, H. S. (1999) Enzymic characterization of a novel member of the regulatory B-like carboxypeptidase with transcriptional repression function: stimulation of enzymic activity by its target DNA. *Biochem J 343 Pt 2*, 341-5.

(165) A. Teplyakov, Polyakov, K., Obmolova, G., Strokopytov, B., Kuranova, I., Osterman, A., Grishin, N., Smulevitch, S., Zagnitko, O. and Galperina, O. (1992) Crystal structure of carboxypeptidase T from Thermoactinomyces vulgaris. *Eur J Biochem 208*, 281-8.

(166) L. Zhao, Buckman, B., Seto, M., Morser, J. and Nagashima, M. (2003) Mutations in the substrate binding site of thrombin-activatable fibrinolysis inhibitor (TAFI) alter its substrate specificity. *J Biol Chem 278*, 32359-66.

(167) (2002) *pET System Manual*, Vol., 10 ed., Novagen.

(168) G. Georgiou, Valax, P., Ostermeier, M. and Horowitz, P. M. (1994) Folding and aggregation of TEM beta-lactamase: analogies with the formation of inclusion bodies in Escherichia coli. *Protein Sci 3*, 1953-60.

(169) G. Georgiou and Valax, P. (1996) Expression of correctly folded proteins in Escherichia coli. *Curr Opin Biotechnol 7*, 190-7.

(170) D. L. Wilkinson and Harrison, R. G. (1991) Predicting the solubility of recombinant proteins in Escherichia coli. *Biotechnology (N Y) 9*, 443-8.

(171) J. Grodberg and Dunn, J. J. (1988) ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. *J Bacteriol 170*, 1245-53.

(172) W. A. Prinz, Aslund, F., Holmgren, A. and Beckwith, J. (1997) The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the Escherichia coli cytoplasm. *J Biol Chem 272*, 15661-7.

(173) P. H. Bessette, Aslund, F., Beckwith, J. and Georgiou, G. (1999) Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proc Natl Acad Sci U S A 96*, 13703-8.

(174) M. Donoghue, Hsieh, F., Baronas, E., Godbout, K., Gosselin, M., Stagliano, N., Donovan, M., Woolf, B., Robison, K., Jeyaseelan, R., Breitbart, R. E. and Acton, S. (2000) A novel angiotensin-converting enzyme-related carboxypeptidase (ACE2) converts angiotensin I to angiotensin 1-9. *Circ Res 87*, E1-9.

(175) F. Arminjon, Guinand, M., Vacheron, M. J. and Michel, G. (1977) Specificity profiles of the membrane-bound gamma-D-glutamyl-(L)meso- diaminopimelateendopeptidase and LD-carboxypeptidase from Bacillus sphaericus 9602. *Eur J Biochem 73*, 557-65.

(176) M. Garnier, Vacheron, M. J., Guinand, M. and Michel, G. (1985) Purification and partial characterization of the extracellular gamma-D- glutamyl-(L)meso-diaminopimelate endopeptidase I, from Bacillus sphaericus NCTC 9602. *Eur J Biochem 148*, 539-43.

(177) M. L. Hourdou, Guinand, M., Vacheron, M. J., Michel, G., Denoroy, L., Duez, C., Englebert, S., Joris, B., Weber, G. and Ghuysen, J. M. (1993) Characterization of the sporulation-related gamma-D-glutamyl-(L)meso- diaminopimelic-acid-hydrolysing peptidase I of Bacillus sphaericus NCTC 9602 as a member of the metallo(zinc) carboxypeptidase A family. Modular design of the protein. *Biochem J 292*, 563-70.

(178) O. Gorbenko, Kuznetsov, V., Kukharenko, O., Zhyvoloup, A., Panasyuk, G., Nemazanyy, I., Filonenko, V. and Gout, I. (2004) Identification of a novel binding partners for tumor suppressor PTEN by a yeast two-hybrid approach. *Eksp Onkol 26*, 15-9.

(179) C. Widmann, Gibson, S., Jarpe, M. B. and Johnson, G. L. (1999) Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol Rev 79*, 143-80.

(180) M. Miller, Shuman, J. D., Sebastian, T., Dauter, Z. and Johnson, P. F. (2003) Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. *J Biol Chem 278*, 15178-84.

(181) G. P. He, Kim, S. and Ro, H. S. (1999) Cloning and characterization of a novel zinc finger transcriptional repressor. A direct role of the zinc finger motif in repression. *J Biol Chem 274*, 14678-84.

(182) A. Cleasby, Wonacott, A., Skarzynski, T., Hubbard, R. E., Davies, G. J., Proudfoot, A. E., Bernard, A. R., Payton, M. A. and Wells, T. N. (1996) The x-ray crystal structure of phosphomannose isomerase from Candida albicans at 1.7 angstrom resolution. *Nat Struct Biol 3*, 470-9.

(183) P. L. Roach, Clifton, I. J., Fulop, V., Harlos, K., Barton, G. J., Hajdu, J., Andersson, I., Schofield, C. J. and Baldwin, J. E. (1995) Crystal structure of isopenicillin N synthase is the first from a new structural family of enzymes. *Nature 375*, 700-4.

(184) M. Sami, Brown, T. J., Roach, P. L., Schofield, C. J. and Baldwin, J. E. (1997) Glutamine-330 is not essential for activity in isopenicillin N synthase from Aspergillus nidulans. *FEBS Lett 405*, 191-4.

(185) C. A. Lesburg, Huang, C., Christianson, D. W. and Fierke, C. A. (1997) Histidine --> carboxamide ligand substitutions in the zinc binding site of carbonic anhydrase II alter metal coordination geometry but retain catalytic activity. *Biochemistry 36*, 15780-91.

(186) I. Yiallouros, Grosse Berkhoff, E. and Stocker, W. (2000) The roles of Glu93 and Tyr149 in astacin-like zinc peptidases. *FEBS Lett 484*, 224-8.

(187) M. Blomster, Wetterholm, A., Mueller, M. J. and Haeggstrom, J. Z. (1995) Evidence for a catalytic role of tyrosine 383 in the peptidase reaction of leukotriene A4 hydrolase. *Eur J Biochem 231*, 528-34.

(188) M. M. Thunnissen, Andersson, B., Samuelsson, B., Wong, C. H. and Haeggstrom, J. Z. (2002) Crystal structures of leukotriene A4 hydrolase in complex with captopril and two competitive tight-binding inhibitors. *Faseb J 16*, 1648-50.

(189) S. W. Vetter and Leclerc, E. (2003) Novel aspects of calmodulin target recognition and activation. *Eur J Biochem 270*, 404-14.

(190) B. T. Schurter, Koh, S. S., Chen, D., Bunick, G. J., Harp, J. M., Hanson, B. L., Henschen-Edman, A., Mackay, D. R., Stallcup, M. R. and Aswad, D. W. (2001) Methylation of histone H3 by coactivator-associated arginine methyltransferase 1. *Biochemistry 40*, 5747-56.

(191) X. Zhang, Zhou, L. and Cheng, X. (2000) Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. *EMBO J 19*, 3509-19.

(192) B. Aufiero, Neufeld, E. J. and Orkin, S. H. (1994) Sequence-specific DNA binding of individual cut repeats of the human CCAAT displacement/cut homeodomain protein. *Proc Natl Acad Sci U S A 91*, 7757-61.

(193) Z. Shang, Isaac, V. E., Li, H., Patel, L., Catron, K. M., Curran, T., Montelione, G. T. and Abate, C. (1994) Design of a "minimAl" homeodomain: the N-terminal arm modulates DNA binding affinity and stabilizes homeodomain structure. *Proc Natl Acad Sci U S A 91*, 8373-7.

(194) E. P. Newberry, Latifi, T., Battaile, J. T. and Towler, D. A. (1997) Structure-function analysis of Msx2-mediated transcriptional suppression. *Biochemistry 36*, 10451-62.

(195) S. Hovde, Abate-Shen, C. and Geiger, J. H. (2001) Crystal structure of the Msx-1 homeodomain/DNA complex. *Biochemistry 40*, 12013-21.

(196) T. M. Conlon and Meyer, K. B. (2004) Cloning and functional characterisation of avian transcription factor E2A. *BMC Immunol 5*, 11.

(197) F. Lluis, Ballestar, E., Suelves, M., Esteller, M. and Munoz-Canoves, P. (2005) E47 phosphorylation by p38 MAPK promotes MyoD/E47 association and muscle-specific gene transcription. *Embo J 24*, 974-84.

(198) M. W. Quong, Romanow, W. J. and Murre, C. (2002) E protein function in lymphocyte development. *Annu Rev Immunol 20*, 301-22.

(199) J. O'Neil, Shank, J., Cusson, N., Murre, C. and Kelliher, M. (2004) TAL1/SCL induces leukemia by inhibiting the transcriptional activity of E47/HEB. *Cancer Cell 5*, 587-96.

(200) J. Zhang, Kalkum, M., Yamamura, S., Chait, B. T. and Roeder, R. G. (2004) E protein silencing by the leukemogenic AML1-ETO fusion protein. *Science 305*, 1286-9.

(201) M. J. Plevin, Mills, M. M. and Ikura, M. (2005) The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem Sci 30*, 66-9.

(202) N. Sabet, Tong, F., Madigan, J. P., Volo, S., Smith, M. M. and Morse, R. H. (2003) Global and specific transcriptional repression by the histone H3 amino terminus in yeast. *Proc Natl Acad Sci U S A 100*, 4084-9.

(203) M. Vogelauer, Wu, J., Suka, N. and Grunstein, M. (2000) Global histone acetylation and deacetylation in yeast. *Nature 408*, 495-8.

(204) F. van Leeuwen, Gafken, P. R. and Gottschling, D. E. (2002) Dot1p modulates silencing in yeast by methylation of the nucleosome core. *Cell 109*, 745-56.

(205) F. van Leeuwen and Gottschling, D. E. (2002) Genome-wide histone modifications: gaining specificity by preventing promiscuity. *Curr Opin Cell Biol 14*, 756-62.

(206) E. L. Gadbois, Chao, D. M., Reese, J. C., Green, M. R. and Young, R. A. (1997) Functional antagonism between RNA polymerase II holoenzyme and global negative regulator NC2 in vivo. *Proc Natl Acad Sci U S A 94*, 3145-50.

(207) P. J. Willy, Kobayashi, R. and Kadonaga, J. T. (2000) A basal transcription factor that activates or represses transcription. *Science 290*, 982-5.

(208) J. F. Wiesen, Young, P., Werb, Z. and Cunha, G. R. (1999) Signaling through the stromal epidermal growth factor receptor is necessary for mammary ductal development. *Development 126*, 335-44.

(209) N. C. Luetteke, Qiu, T. H., Fenton, S. E., Troyer, K. L., Riedel, R. F., Chang, A. and Lee, D. C. (1999) Targeted inactivation of the EGF and amphiregulin genes reveals distinct roles for EGF receptor ligands in mouse mammary gland development. *Development 126*, 2739-50.

(210) K. J. Fowler, Walker, F., Alexander, W., Hibbs, M. L., Nice, E. C., Bohmer, R. M., Mann, G. B., Thumwood, C., Maglitto, R., Danks, J. A. and et al. (1995) A mutation in the epidermal growth factor receptor in waved-2 mice has a profound effect on receptor biochemistry that results in impaired lactation. *Proc Natl Acad Sci U S A 92*, 1465-9.

(211) J. A. Schroeder and Lee, D. C. (1998) Dynamic expression and activation of ERBB receptors in the developing mouse mammary gland. *Cell Growth Differ 9*, 451-64.

(212) E. P. Sandgren, Schroeder, J. A., Qui, T. H., Palmiter, R. D., Brinster, R. L. and Lee, D. C. (1995) Inhibition of mammary gland involution is associated with transforming growth factor alpha but not c-myc-induced tumorigenesis in transgenic mice. *Cancer Res 55*, 3915-27.