

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.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

**Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600**


UMI[®]



Université d'Ottawa • University of Ottawa

**Molecular Cloning of Three Genes Encoding
SNF1/AMPK Protein Kinases from
*Drosophila melanogaster***

By

 Eduardo N. Taboada

Thesis submitted to the
School of Graduate Studies and Research
in partial fulfillment of the requirements for the
Doctor of Philosophy degree in Biology

University of Ottawa



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

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

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

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

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

0-612-52279-2

Canada

ACKNOWLEDGEMENTS

I would like to express a most sincere thank you to my supervisor Dr. Donal Hickey, for providing me with a rich learning experience, for providing invaluable support and guidance, for his patience and encouragement, and above all, for going above and beyond the call of duty in looking out for the best interests of myself and my fellow students.

I am also thankful to the members of my research committee, Dr. B. Benkel, Dr. G. Carmody and Dr. T. Moon for their helpful advice.

I would also like to pass many thanks to my various labmates past and present, who have over the years provided excellent assistance, ideas, and discussions, scientific and otherwise.

I am also very thankful to the Taboada, Scarinci-Taboada and Mykytczuk families, for providing me with encouragement, strength, patience, love and support.

Lastly, I want to extend my warmest thoughts to Oksana Mykytczuk. While the road we have traveled has meandered greatly, the trip has certainly been memorable.

ABSTRACT

Drosophila melanogaster provides a rare example of glucose repression in a metazoan and there is evidence for the functional conservation between the yeast and *Drosophila* glucose repression pathways. In an effort to characterize potentially conserved molecular components of the *Drosophila* glucose repression pathway, possible *Drosophila* homologues of genes known to play a role in yeast glucose repression were sought.

A number of different genes are known to play a role in the yeast glucose repression pathway and the *SNF1* gene, encoding a serine-threonine protein kinase, has the essential function of releasing genes from glucose repression in response to glucose starvation. *SNF1* homologues have been found in a number of different organisms and recent molecular and biochemical evidence has revealed that the mammalian AMP-activated protein kinases (AMPKs), known to regulate fatty acid and sterol metabolism in response to stress, represent the mammalian homologues of *SNF1*. *SNF1* kinase homologues have been found across diverse eukaryotic lineages thus making them ideal candidates for cloning. We therefore took aim at cloning possible *SNF1* homologues in *Drosophila*.

Three novel *Drosophila* protein kinases were cloned and found to represent homologues of previously reported protein kinases. Two of the *Drosophila* protein kinase genes reported here, *DM_SNF1B-1* and *DM_SNF1B-2* are found clustered in tandem within a 4-kb region of the *Drosophila* genome. The encoded polypeptides share intermediate levels of sequence similarity with *SNF1* but share higher levels of sequence similarity with the putative human protein kinase KP78, a marker protein lost in chemically induced primary carcinoma of the pancreas.

The third *Drosophila* protein kinase gene *DM_SNF1A*, which is found in a single copy in the *Drosophila* genome, was found to share high similarity with the yeast *SNF1* and an even higher level of similarity (92% similarity within the catalytic domain) with the mammalian AMPKs. Although the non-catalytic regulatory domain of *Dm_snf1a* contains a number of features held in common with the two mammalian *Ampk- α* isoforms, it also contains features held in common with

yeast Snf1 and which are missing in the Ampk- α isoforms. Some of these features may be the basis for known functional differences between the yeast and mammalian kinases. Given the conservation between the yeast and *Drosophila* glucose repression systems the *DM_SNF1A* may play a dual role in both glucose repression and the regulation of lipid synthesis in response to stress.

In addition to well-established SNF1 homologues, a growing number of “SNF1-related” kinase sequences exist in the various molecular sequence databases. Phylogenetic analysis of these sequences, along with those of the three novel kinases reported here, revealed that *DM_SNF1A* is found on a well-supported branch that includes well-known SNF1-homologues. *DM_SNF1B-1* and *DM_SNF1B-2* branch along with kinases related to the human *KP78* kinase. These kinases, along with the bulk of other ‘SNF1-related’ kinases in the databases represent sequences that are too divergent from the main SNF1/AMPK homologues for the sequence similarity to have any functional significance. Thus functional inferences in the literature that have been made based on sequence similarity between these ‘SNF1-related’ kinases and the SNF1/AMPK kinases must be re-examined.

The identification of a *Drosophila* homologue of the yeast SNF1 kinase helps to extend the list of functional similarities between the *Drosophila* and yeast glucose repression systems. The cloning of the *Drosophila DM_SNF1A* presents us with the opportunity for future comparative studies to shed light on the evolution of both the SNF1/AMPK family of protein kinases and the signalling pathways that they mediate.

TABLE OF CONTENTS

AKNOWLEDGEMENTS	i
ABSTRACT	ii
LIST OF FIGURES	ix
LIST OF TABLES.....	xi
LIST OF ABBREVIATIONS.....	xii
GENE NAMING CONVENTIONS.....	xii
1. <u>LITERATURE REVIEW</u>.....	1
1.1 <u>Introduction</u>	1
1.2 <u>Signal transduction</u>	3
1.2.1- The role of signal transduction pathways.....	3
1.2.2- The components of a signal transduction pathway.....	3
1.3 <u>An overview of protein phosphorylation</u>	7
1.3.1- The “eukaryotic” protein kinase superfamily.....	8
1.3.2- Protein kinases in Bacteria, Archaea.	10
1.3.3- Homologous kinase domains in the ePK superfamily	13
1.3.4- Protein kinase cascades and their role in signal transduction	15
1.3.5- The phosphoprotein phosphatases	16
1.4 <u>The SNF1/AMPK family of protein kinases. A historical overview.</u>	19
1.4.1- The role of SNF1 in the yeast glucose repression	20
1.4.1.1- Glucose repression in yeast metabolism	20

TABLE OF CONTENTS (cont)

1.4.1.2- The glucose repression pathway: generation of the glucose signal.....	22
1.4.1.3- The glucose repression pathway : the downstream effectors	23
1.4.1.4- The glucose repression pathway: the role of SNF1 in derepression	24
1.4.1.5- The SNF1 kinase is a heterotrimeric complex	26
1.4.1.6- Glucose repression: targets of the SNF1 kinase complex.....	29
1.4.1.7- The role of protein phosphatases in glucose repression.....	33
1.4.2- The role of the AMPK in mammalian lipid metabolism	33
1.4.3- Molecular characterization of the mammalian AMPKs.....	35
1.4.4- Regulation of the mammalian AMPKs	38
1.4.5- The physiological role of mammalian AMPKs	40
1.4.6- Similarities between SNF1 and AMPK protein kinases.....	44
1.4.7- Phylogenetic distribution of the SNF1/AMPK protein kinases	46
1.5- <u>The glucose repression system in <i>Drosophila</i></u>	49
1.6- <u>Thesis project : description and goals</u>	51
2. <u>MATERIALS AND METHODS</u>	52
2.1- <u>Degenerate PCR primer design</u>	52
2.1.1- SNF1/AMPK biased primers.....	52

TABLE OF CONTENTS (cont)

2.1.2- Primers biased towards SNF1 and its plant homologues.....	52
2.2- <u>PCR cloning of SNF1-like sequences</u>	53
2.3- <u>Probe construction and <i>Drosophila</i> library screening</u>	55
2.4- <u>Subcloning and DNA sequencing of genomic clones</u>	56
2.5- <u>Northern and genomic Southern analysis</u>	57
2.6- <u>Cloning of full length cDNA of DM SNF1A</u>	58
2.7- <u>Sequence analysis and sequence comparisons</u>	60
2.8- <u>Molecular sequence accession numbers</u>	61
3. <u>RESULTS</u>	
3.1- <u>PCR identification of novel kinase domains</u>	62
3.1.1- Identification of novel <i>Drosophila</i> SNF1-like kinase domains.....	62
3.1.2- Identification of novel human SNF1-like kinase domains.....	64
3.1.2.1- Expression studies on novel human SNF1-like kinase domains	68
3.2- <u>Cloning of novel protein kinases from <i>Drosophila</i></u>	69
3.2.1- Restriction mapping and Southern analysis of clones.....	69
3.2.2- Characterization of <i>DM_SNF1B-1</i> and <i>DM_SNF1B-2</i>	69
3.2.2.1- Restriction mapping and sequencing of the <i>DM_SNF1B- 1/DM_SNF1B-2</i> gene cluster.....	72

TABLE OF CONTENTS (cont)

3.2.3	Characterization of <i>DM_SNF1A</i>	76
3.2.3.1-	Restriction mapping and sequencing of the <i>DM_SNF1A</i> gene.....	76
3.2.3.2-	Cloning of a cDNA copy of <i>DM_SNF1A</i>	80
3.2.3.3-	Analysis of the sequence divergence between <i>DM_SNF1A</i> and its human counterparts.....	83
3.2.3.4-	The <i>DM_SNF1A</i> gene appears to be in single copy in the <i>Drosophila</i> genome.....	83
3.2.4-	Expression studies on the novel <i>Drosophila</i> kinases.....	86
3.3	<u>Phylogenetic survey of the SNF1/AMPK protein kinase family</u>	88
3.3.1-	<i>DM_SNF1B</i> : relationship to other related kinases.....	88
3.3.1.1-	Protein sequence alignments.....	88
3.3.1.2-	Phylogenetic analysis of sequences closely related to <i>DM_SNF1B-1</i> and <i>DM_SNF1B-2</i>	90
3.3.2-	<i>DM_SNF1A</i> : relationship to SNF1 and other close relatives.....	92
3.3.2.1-	Protein sequence alignments.....	92
3.3.2.2-	Phylogenetic analysis of closely-related SNF1/AMPK sequences.....	93
3.3.3-	Relationship between SNF1, AMPK and other SnRKs.....	96

TABLE OF CONTENTS (cont)

3.3.3.1- Protein sequence alignments.....	96
3.3.3.2- Phylogenetic analysis of the SNF1/AMPK and SNF1- related kinases.....	97
4. <u>DISCUSSION AND CONCLUSIONS</u>	
4.1 <u>Homology-based PCR cloning of members of a protein kinase subfamily.....</u>	101
4.2 <u>Evolutionary conservation of <i>DM SNF1B-1</i> and <i>DM SNF1B-2</i>.....</u>	105
4.3 <u><i>Drosophila SNF1A</i>: a putative homologue of the mammalian AMPKs.....</u>	108
4.4 <u>Molecular evolution of the SNF1/AMPK protein kinase family.....</u>	114
4.5 <u>Molecular conservation of the glucose repression cascade in <i>Drosophila</i>.....</u>	119
4.6 <u>Evolution of SNF1/AMPK-mediated cascades.....</u>	123
4.7 <u>Conclusions.....</u>	126
5. <u>REFERENCES</u>	129
APPENDIX-I	
SNF1/AMPK and SNF1-related kinases used in phylogenetic.....	150
analysis	

LIST OF FIGURES

<u>Figure</u>		<u>Page</u>
I	An overview of the yeast glucose repression system.	24
II	Structure of the SNF1 kinase complex and glucose-regulated interactions.	28
1	Identification of SNF1/AMPK-like sequences from <i>Drosophila melanogaster</i> .	63
2	PCR amplification of SNF1/AMPK-like sequences from <i>Drosophila melanogaster</i> .	65
3	PCR identification of a novel human AMPK isoform.	66
4	Restriction mapping and subcloning of the <i>Drosophila DM_SNF1B</i> gene cluster	70
5	Sequencing strategy used on subclone pET193-2	71
6	Sequence of a 6.5 Kb region bearing the <i>DM_SNF1B-1/DM_SNF1B-2</i> gene cluster.	74
7	Protein sequence alignment of the proteins encoded by <i>DM_SNF1B-1</i> and <i>DM_SNF1B-2</i> and the human kinase KP78.	75
8	Restriction mapping and subcloning of the region containing the <i>DM_SNF1A</i> gene.	78
9	Sequencing strategy used to obtain the <i>DM_SNF1A</i> gene.	79
10	Sequence of a 3.8 Kb region bearing the complete <i>DM_SNF1A</i> gene.	81
11	Protein sequence alignments featuring <i>Drosophila DM_Snf1a</i> and the human and yeast homologues.	82
12	Protein sequence alignment of a group of kinases similar to <i>DM_SNF1B-1</i> and <i>SNF1B-2</i> .	89
13	Protein parsimony tree derived from an alignment of kinases sharing similarity with <i>DM_SNF1B-1</i> and <i>DM_SNF1B-2</i> .	91

LIST OF FIGURES (cont)

14	Protein sequence alignment of a group of sixteen highly related SNF1/AMPK-like protein kinases	94
15	Protein parsimony tree derived from an alignment of SNF1/AMPK-like protein kinases.	95
16	Protein sequence alignment of the catalytic domains of thirty-one SNF1/AMPK-related protein kinases.	99
17	Protein parsimony tree derived from an alignment of kinases sharing similarity with <i>DM_SNF1B-1</i> and <i>DM_SNF1B-2</i> .	100

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1	Percentage of sequence identity and sequence similarity between <i>DM_SNF1A</i> and the human AMPK isoforms α -1 and α -2	84
2	Distance matrix for the catalytic and non-catalytic domains of seven protein kinases	85
3	SNF1/AMPK and SNF1-related kinases used in phylogenetic analysis	150

LIST OF ABBREVIATIONS

aa: amino acid	Indels: insertions/deletions
ACC: acetyl-CoA carboxylase	Kb: kilobase pair(s)
ACCK: acetyl-CoA carboxylase kinase	kDa: kilo Dalton
AMPK: AMP-activated protein kinase	MAP: mitogen activated protein (kinase)
AMPKK: AMP-activated protein kinase-kinase	M/mM: molar/millimolar
AMP/ADP/ATP: adenosine mono/di-/tri-phosphate	mCi: milli-Curie
bp: base pair(s)	min: minute
°C: degree Celsius	mg/ml: milligram/millilitre
cAMP: cyclic adenosine monophosphate	mRNA: messenger RNA
cDNA: complementary DNA	ng: nanogram
cGMP: cyclic guanosine monophosphate	nt: nucleotide
CK: creatine kinase	PCr: phosphocreatine
CoA: coenzyme A	PCR: polymerase chain reaction
Cr: creatine	PFU: plaque forming units
C-terminus(al): carboxy-terminus(al)	PKA: cAMP activated protein kinase
DNA: deoxyribonucleic acid	pmol: picomole
dNTP: deoxy nucleotide triphosphate	PPI: protein phosphatase 1
DTT: dithiothreitol	RNA: ribonucleic acid
EDTA: ethylenediaminetetraacetic acid	RT-PCR: reverse transcriptase PCR
ePKs: eukaryotic protein kinases	SAMS: peptide substrate for measuring AMPK activity
g: gram(s)	SDS: sodium dodecyl sulfate
GPAT: <i>sn</i> -glycerol-3-phosphate acetyltransferase	Snf1-RD: SNF1 kinase regulatory domain
GTP: guanosine triphosphate	Snf1-CD: SNF1 kinase catalytic domain
HMGR: hydroxylmethylglutaryl-CoA reductase	U: unit
HSL: hormone sensitive lipase	µg/µl: microgram/microliter
	UV: ultraviolet
	vs: versus

GENE NAMING CONVENTIONS

I have tried to adhere to the following naming conventions for genes and their proteins:

For example, for *SNF1* vs. **Snf1** vs. *snf1* vs. **SNF1**:

- The gene name is given in upper case italics (eg. *SNF1*)
- The encoded protein is given in lower case with the first letter capitalized (eg. **Snf1**)
- The mutant of the gene is given in lower case italics (eg. *snf1*)
- The multi-protein kinase complex is given in upper case (eg. **SNF1**)

1. LITERATURE REVIEW

1.1 Introduction.

Adaptation is one of the most important principles in biology and the evolutionary process is fuelled by the pressure on organisms to adapt to ever changing conditions. In order for organisms to survive and thrive, they must be able to sense changes in their environment and to be able to respond accordingly. The process of adaptive evolution has forced organisms to devote a great deal of resources towards sensing and processing biological information and towards utilizing this information to make changes. These changes may range from biochemical changes at the cellular level to behavioural changes at the level of an individual organism or even a population.

Signal transduction, the transfer of biological information within or between cells, plays a crucial role in the regulation of cellular functions, including cell differentiation, and cell proliferation. Since different signal transduction pathways interact with one another to orchestrate every cellular response, the elucidation of signal transduction pathways has become one of the most important areas of research in recent years. In section 1.2, I will give an overview of basic concepts in signal transduction as well as discuss significant findings in signal transduction research. Although signal transduction pathways can underlie events at the level of the cell, the organism, or a population, I shall restrict my discussion to events at the cellular level since my research has dealt with components of a eukaryotic signal transduction pathway which effects intracellular metabolic changes in response to changes in nutritional status.

Protein kinases and the phosphorylation cascades for which they are responsible, relay a majority of the biological information processed by the eukaryotic cell and can be thought of as the major currency of signal transduction pathways. This is reflected by the number of protein kinase genes encoded by eukaryotic genomes and by the different number of biological processes that are regulated by protein phosphorylation. I will devote section 1.3 to an overview of protein kinases and their role as signal transducers in the cell and to the discussion of the evolutionary conservation that is observed within the protein kinase superfamily.

One of the areas of interest in our lab has been the study of gene expression as well as the regulation of gene expression in response to different signals. The presence of a possible evolutionary link between our model system and signalling systems involved in yeast glucose repression and mammalian lipid metabolism has led my research towards the study of the family of protein kinases that act as the major signal transducers in each of these signalling pathways. In section 1.4, I will give some background on this protein kinase family, the AMP-activated/SNF1 family, and discuss current knowledge on the role played by these protein kinases in different organisms. In particular, I will discuss both the work that uncovered a role for SNF1 in yeast glucose repression as well as the work that has established the AMP-activated protein kinases (AMPK) as major players in regulating mammalian fatty acid and cholesterol metabolism in response to a compromised energy status.

Finally, in section 1.5, I will provide background on *Drosophila* glucose repression and how the emerging similarities it shares with the yeast glucose repression pathway led us to the study of SNF1/AMPK kinases in *Drosophila*. I will then describe how my thesis project fits in with current developments in research on the SNF1/AMPK family of protein kinases.

1.2 Signal Transduction.

1.2.1 The role of signal transduction pathways.

The concept of signal transduction pathways is crucial towards understanding how organisms sense, relay and interpret biological information. All cells, in general, can be regarded as signal transduction machines whereby extracellular signals are translated into cellular responses via intracellular communication. The nature of extracellular stimuli is most often a change in concentration of a signalling molecule that ultimately induces a number of responses such as cell growth, differentiation, division, secretion, movement toward or away from a stimulus and altered metabolism.

In order for a stimulus to elicit an appropriate cellular response, a signal must be able to pass across the membrane and reach the appropriate cellular compartment that will act on the signal and effect a response. Some signals, hydrophobic molecules such as steroid hormones, can cross the plasma membrane and interact with intracellular components directly. The majority of external signalling molecules, or first messengers, require interaction with a cell surface receptor, a membrane protein that can transduce the signal to a membrane-bound effector enzyme that then produces a second messenger. The second messenger, a small diffusible molecule or ion, carries the signal to the appropriate intracellular destination: the cytosol, an organelle, or the nucleus. Although details are still being elucidated in many cases, one universal theme in signal transduction is that generation of the second messenger will almost invariably lead to the activation of protein kinases that can then phosphorylate, and modulate the activity of, different protein targets.

1.2.2 The components of a signal transduction pathway.

There are three main types of signal reception mechanisms in eukaryotic cells: sequence-

specific DNA-binding receptor proteins that are found in the cytoplasm; enzymes in intermediary metabolic pathways; and finally, receptors located on the external surface of the plasma membrane (Barrit, 1992). Sequence-specific DNA-binding receptor proteins bind to and form a complex with hormones or morphogens in the cytoplasm, such as thyroid hormones and retinoic acid, respectively (Giguere *et al.*, 1987). The hormone- or morphogen- DNA-binding receptor complex translocates to the nucleus and, ultimately, induces an increased transcription rate in a specific set of genes (reviewed by Ribeiro *et al.*, 1995; Katzenellenbogen and Katzenellenbogen, 1996). The ecdysterone DNA-binding receptor in *Drosophila*, for example, when complexed with its ligand, binds to a hormone-response element (an enhancer sequence) on target DNA and activates transcription of genes essential for adult development (Gronemeyer *et al.*, 1980).

Similar to hormone induced signal transduction, extracellular metabolites may also act as external stimuli although in most cases their main role is to regulate cell metabolism. A cytoplasmic enzyme initially receives the external signal and catalyzes a reaction that produces other intermediary metabolic products. Some intermediary metabolites alter the concentration of second messengers and thus, induce specific cellular responses in response to the extracellular signal (Barrit, 1992). The metabolism of glucose and other metabolites in the pancreas, for instance, stimulates insulin secretion from β -cells. Although the exact mechanism by which the induction of insulin secretion occurs has yet to be determined, the metabolism of nutrients that stimulate this pathway ultimately causes activation of a voltage-gated Ca^{2+} channels, leading to an increase in free Ca^{2+} concentration in the cytoplasm of β -cells, which then leads to the migration of insulin secretory granules to the plasma membrane (Streb *et al.*, 1983; Ashcroft *et al.*, 1994).

The most common mechanism by which signals are received by cells, however, is via plasma membrane receptors. Because a response to a specific signal (a hormone, growth factor or a metabolite) is dependent on a cell bearing the appropriate cell surface receptors on its plasma

membrane, each cell is able to respond to only a select number of stimuli at a given point in time. Specificity of signal reception is dependent on the vast diversity and stereospecificity of receptors, ligands and transducers at the surface of the cell. In addition, there are several mechanisms through which plasma membrane receptors are coupled to effector molecules. For instance, receptors may either be coupled with enzymes or ion channels through GTP-binding proteins, they may be ion channels themselves or the cytoplasmic domain of the receptor may have protein-tyrosine kinase activity (Barrit, 1992). Transduction of external signals to the intracellular space, however, is carried out by common pathways that use only a few known effector enzymes and second messengers.

The main components of intracellular communication are: GTP-binding regulatory proteins in the plasma membrane, mobile intracellular messengers and metabolites, the network of protein kinases, sequence-specific DNA binding proteins and the proteins synthesized in response to extracellular signals (Barrit, 1992). GTP-binding proteins, or G-proteins, are peripheral membrane protein transducers that link extracellular signals to intracellular machinery (Lefkowitz, 1993). In particular, G-proteins directly couple plasma membrane receptors to effector enzymes or ion channels (Lefkowitz, 1993). Mobile intracellular messengers are mostly water-soluble molecules or ions, such as cyclic-AMP (cAMP) and Ca^{2+} , respectively. The mobility of water-soluble intracellular messengers is limited to the cytoplasm of a cell whereas the lipid soluble intracellular messengers, such as diacylglycerol, are confined to the plasma membrane (Barrit, 1992). As mentioned earlier, the role of intracellular messengers is to activate protein kinases that, in turn exert effects on target proteins. The protein kinase network includes the cyclic nucleotide-activated kinases (cAMP and cGMP), the Ca^{2+} activated kinases, the diacylglycerol activated-protein kinase C, the various cell-division control kinases and the vast number of receptor protein-tyrosine kinases in the plasma membrane, among others (reviewed by Hunter and Cooper, 1985; Barrit, 1992;

Graves and Krebs, 1999). Sequence-specific DNA binding proteins are then involved in conveying information from the cytoplasm to particular genes within the nucleus (Mitchell and Tjian, 1989). Since the localization and DNA-binding activity of transcription factors can be affected by phosphorylation, they can act as the downstream effectors of kinase signalling pathways (Karin, 1994). The main purpose of signal transduction pathways is the production of a cellular response to the initial stimulus; in many cases this involves the synthesis of proteins. The type, final destination and concentration of the newly synthesized protein will determine the specific response elicited by the cell.

Most signal transduction pathways are not only in charge of detecting and passing down extracellular signals but are also involved in amplifying the signal. For example, a signal originating from one single hormone-receptor complex can activate more than one transducer G-protein by the lateral diffusion of both G-proteins and receptors in the plasma membrane (Lefkowitz, 1993). In addition, each membrane-bound effector enzyme can produce many second messenger molecules upon its activation. Similarly, protein kinases further down the signalling pathway may phosphorylate numerous target proteins. Each component of the pathway has the potential to amplify the effect of the initial signal and thus the term “cascade” is used to describe this series of events (Barrit, 1992).

1.3 An overview of protein phosphorylation

Phosphate has been known to be associated with protein for over 100 years. The first phospho-amino acid was isolated from protein in 1933 and since then, it has become common knowledge that a vast number of proteins are covalently modified by the addition of phosphate (reviewed by Hunter, 1991; Hunter and Hanks, 1995). Most of these phosphate modifications are known to occur on serine and threonine residues with smaller amounts being added to tyrosine, histidine, lysine, arginine, aspartic acid, glutamic acid, and cysteine residues (Hunter and Cooper, 1985). The reaction is catalyzed by protein kinases and occurs through the addition of a phosphate moiety to the side-chains of amino acid residues in target proteins. Addition of the highly negatively charged phosphate group can modify the conformation of the target protein and this may either alter its activity or its interaction with other proteins (Barrit, 1992). The reverse reaction can also occur and removal of the phosphate moiety is catalyzed by phosphoprotein phosphatases. Removal of the phosphate group restores the conformation and properties of the native protein (Barrit, 1992).

It was in the 1950s that a number of investigators involved in studying the regulation of glycogen degradation and biosynthesis first recognized that proteins were reversibly phosphorylated and dephosphorylated as part of normal cell regulation processes. Results obtained by Krebs and collaborators provided evidence that the active and inactive forms of glycogen phosphorylase (phosphorylase *a* and phosphorylase *b*, respectively), previously discovered by Cori and Cori in 1943, differed only in that the active form was phosphorylated. Soon after, they provided evidence that the enzyme responsible for the activation of glycogen phosphorylase, phosphorylase kinase, is itself regulated through reversible phosphorylation. By the late 1960s, pathways other than those

involved in glycogen metabolism had also been shown to be regulated by reversible phosphorylation (Linn *et al.*, 1966) and the protein kinase which mediates the multiple actions of the second messenger cAMP, the cAMP-dependent protein kinase (PKA), was isolated (Walsh *et al.*, 1969). In the 1970s, numerous reports cited the discovery of novel protein kinases and of different metabolic pathways displaying enzyme regulation through reversible phosphorylation, further establishing the important role played by protein kinases and their antagonists, the protein phosphatases, in regulating cellular processes.

1.3.1 The “eukaryotic” protein kinase superfamily.

Eukaryotic protein kinases catalyze the transfer of orthophosphate to amino acid residues in target proteins. They use the γ -phosphate from ATP or GTP and catalyze an esterification reaction by utilizing the hydroxyl groups found in the alcohol sidechains of serine and threonine, or the phenolic group of tyrosine, as phosphate group acceptors (Hunter, 1991). As a result, the protein is covalently modified and becomes a monophosphate ester while the donor nucleotide triphosphate loses a single phosphate to become a nucleotide diphosphate (Barrit, 1992). With few exceptions, all of the serine/threonine and tyrosine protein kinases found in eukaryotes are evolutionarily related and form the largest known superfamily of proteins (Hanks *et al.*, 1988; Hanks and Quinn, 1991; Hanks and Hunter, 1995). The region which is conserved in all superfamily protein kinases, the core catalytic domain, is responsible for catalyzing the phosphorylation reaction and deletion studies on different protein kinases have shown that it is the only element required for constitutive catalysis (Hanks and Hunter, 1995). I will use the terms “superfamily kinases” or “eukaryotic protein kinases” (ePKs) to describe protein kinases which bear the conserved core catalytic domain

which defines this large group of related kinases. Protein kinases that are unrelated to the superfamily kinases do exist in eukaryotes although, as I will discuss later, these are few in number and most show similarity to the bacterial protein histidine kinases, which are unrelated to the eukaryotic superfamily kinases (Hunter, 1991). Predictions that about 1 to 2 percent of all eukaryotic genes would turn out to be protein kinases (Hunter, 1988; Hunter, 1994) have been confirmed with the completion of the *Saccharomyces cerevisiae* genome sequencing project (reviewed by Zagulski *et al.*, 1998). It has been established that in *S. cerevisiae* there are 113 protein kinase genes, roughly 2 % of the known and putative genes in the entire genome (Hunter and Plowman, 1997). Although this total is a little lower than had initially been predicted, by extrapolating from these results, as well as from data from the *Caenorhabditis elegans* sequencing project, it is still expected that mammalian genomes will contain well over 1000 protein kinase genes (Hunter and Plowman, 1997).

Protein kinases have been found in every eukaryotic lineage and, thus, the ancestral protein kinase gene can be traced back to a time early in eukaryotic evolution before the evolutionary split of the major eukaryotic kingdoms (Hanks and Quinn, 1991; Hanks and Hunter, 1995). This view is supported by the fact that protein phosphorylation pathways play crucial roles in the regulation of basic cellular processes shared by all eukaryotes and by the fact that many phosphorylation cascades have been conserved between the unicellular eukaryotes, plants, animals and fungi. This would suggest that protein phosphorylation pathways became an integral part of early eukaryotic evolution (Hanks and Hunter, 1995). In addition, phosphorylation pathways are also involved in processes that are specific to higher eukaryotes. For instance, many morphogenetic and cellular differentiation processes unique to multicellular eukaryotes require protein phosphorylation

pathways. Many protein tyrosine kinases act as hormone and growth factor receptors in multicellular differentiated organisms. In contrast, there are no superfamily protein tyrosine kinases present in the *S. cerevisiae* genome and protein tyrosine phosphorylation occurs only in low levels in Bacteria and Archaea (Hunter and Plowman, 1997). Given the necessity for cell-cell signalling in any multicellular organism, a strong case has been made for the co-evolution of protein tyrosine kinases and multicellularity in higher eukaryotes (Hunter and Cooper, 1985). Interestingly, most protein tyrosine kinases are plasma membrane associated, which enables them to act in the sensing and transducing of extracellular signals (Hunter and Cooper, 1985).

1.3.2 Protein kinases in Bacteria, Archaea.

Just as in eukaryotes, protein phosphorylation is equally important to the regulation of cellular processes in Bacteria and Archaea. The protein kinases that were first found in bacterial species can be divided into three general classes (reviewed by Saier, 1993). The first, and largest, is formed by the protein-histidine kinases of two-component regulatory systems. The protein-histidine kinases auto-phosphorylate at the 1- or 3- position of a histidine residue (autophosphorylation can also occur, though less commonly, at the guanido group of arginine or at the ϵ -NH₂ group of lysine) before transferring the phosphate group to an aspartate residue on a sensor protein or sensor domain fused with the kinase, termed the response-regulator protein (Falke *et al.*, 1997). The two other main groups of bacterial kinases are smaller and less well characterized and are composed of the phosphoenol-pyruvate-dependent phosphotransferases and serine/threonine kinases such as the isocitrate dehydrogenase kinase/phosphatase (Hunter, 1985). Neither of these families is related to each other nor to the ePK superfamily based on protein sequence analysis.

For many years the prevailing view was that although all three domains of life make use of protein phosphorylation, phosphorylation was mediated by families of kinases specific to each lineage. As a logical extension of this view the ePKs, and especially protein-tyrosine kinases, were thought to occur only in eukaryotes, while protein-histidine kinases were restricted to Bacteria, with Archaea having an unknown protein phosphorylation system status (Kennelly and Potts, 1996). This exclusivity has been slowly eroded by emerging reports outlining the discovery of small numbers of superfamily-like protein kinases in Bacteria and Archaea, the discovery of protein-histidine-like kinases in Archaea and eukaryotes, and the detection of phospho-tyrosine in Bacteria and Archaea (reviewed by Kennelly and Potts, 1996). In one of the most provocative findings at the time, Munoz-Dorado *et al.* (1992) first described several ePK-like kinases in the myxobacterium *Myxococcus xanthus*, a soil bacterium which is able to undergo cellular differentiation leading to the formation of fruiting bodies similar to those of the eukaryotic slime mold *Dictyostelium discoideum*. Northern blot analysis revealed that one of these ePK-like kinases, Pkn1, was developmentally regulated and its deletion results in premature differentiation and spore formation, thus implying a role in the regulation of developmental cycle in *M. xanthus* (Munoz-Dorado *et al.*, 1992). These results fuelled the speculation that ePK based protein phosphorylation had evolved with the emergence of multicellularity and cellular differentiation. However, Leonard *et al.* (1998) have recently surveyed the sequences of a number of completed genomes (including bacterial, archaeal and *S. cerevisiae* genomes, and partial data from the then-incomplete *C. elegans* sequencing project) and have detected the presence of 5 novel ePK-like gene families with bacterial and archaeal representatives and which bear similarity to the eukaryotic superfamily kinases

(Leonard *et al.*, 1998). One of these families represents sequences which are most similar to the ePK-like kinases of *M. xanthus* discovered by Munoz-Dorado *et al.* (1992), while the other four represent novel families of putative kinases. In addition to eukaryotic-like protein kinases, there have also been reports of eukaryotic-like protein phosphatases and tyrosine-phosphorylated proteins in several bacterial and archaeal species (Kennelly and Potts, 1998).

Just as eukaryotic superfamily-like protein kinases have been found in Bacteria and Archaea, protein-histidine kinase-like genes have been detected in eukaryotes and Archaea. Popov and co-workers cloned the rat mitochondrial branched-chain α -ketoacid dehydrogenase kinase gene (Popov *et al.*, 1992) and the rat pyruvate dehydrogenase kinase gene (Popov *et al.*, 1993) and showed that their closest homologues were not the ePKs but rather the bacterial protein-histidine kinases, with which they share weak similarity. Since then, genes with greater sequence and functional similarity to the bacterial protein-histidine kinases have been found: the *SLN1* gene of *S. cerevisiae* encoding the upstream two-component regulator involved in an osmosensing MAP kinase cascade (Hughes, 1994); the *DokA* gene of *D. discoideum*, which is also involved in osmosensing (Schuster and Simon, 1996); the *ETR-1* gene of *Arabidopsis thaliana* which is involved in the ethylene response (Chang *et al.*, 1993); and finally, the *CheA* gene in the archaeon *Halobacterium salinarium*, which is involved in chemotactic and phototactic sensory-response pathways (Rudolph and Oesterhelt, 1995).

These different lines of evidence have led to an unresolved conflict of views on the evolutionary history of protein kinases (Kennelly and Potts, 1996; Leonard *et al.*, 1998). In one scenario, some of these families were present in the common ancestor of the three domains of life.

In the other scenario the acquisition of kinase genes in different lineages within each domain has occurred through horizontal gene transfer.

1.3.3 Homologous kinase domains in the ePK superfamily

A comparison of sequences from multiple ePK family members first uncovered the conserved features in primary structure that are required for kinase activity (Hanks *et al.*, 1988). These comparisons uncovered that sequence similarity among ePKs is essentially restricted to the core catalytic domain, a stretch of ~ 270 amino acid residues with a high degree of conservation even among the more distantly related superfamily members (Hanks and Hunter, 1995). While they may share similarity that extends beyond the core domain, even the closest relatives often have widely divergent sequences outside their core catalytic domain. [The significance of differences lying outside the catalytic domain will be discussed further on in this thesis]. Hanks *et al.* (1988) also performed phylogenetic analysis on the ePK catalytic domain sequences and found that protein kinases with known related function cluster together on phylogenetic trees. Since their 1988 study, Hanks and his collaborators have carried out more comprehensive studies and have further refined their analysis to include the widest possible cross-section of ePK superfamily members (Hanks and Quinn, 1991; Hanks and Hunter, 1995). Not only did their results confirm earlier findings but similar phylogenetic analyses have allowed the prediction of the properties of novel kinases and have become the basis for classifying the members of the ePK superfamily into distinct families (Hanks and Quinn, 1991; Hanks and Hunter, 1995). Similarity in kinase domain sequences has, therefore, proved to be an excellent indicator of structural and functional similarities shared by closely related kinases.

By performing a multiple sequence alignment of the core catalytic domains of known representative ePKs, Hanks and collaborators have found that the catalytic domain is split into twelve conserved subdomains (Hanks *et al.*, 1988; Hanks and Quinn, 1991; Hanks and Hunter, 1995). These blocks of uninterrupted sequence represent localized regions of high conservation and gaps in the sequence alignment define the boundaries of the subdomains. Eight of the subdomains contain positions with residues that are invariant or nearly invariant across all the different members of the ePK superfamily. Additionally, all twelve subdomains contain a characteristic pattern of positions where biochemically similar amino acids are conserved throughout the superfamily (Hanks and Quinn, 1991; Hanks and Hunter, 1995). In addition to the positions that are conserved in all ePKs, sequence conservation within subdomains is highest among close relatives and, thus, ePKs which are more closely related to one another share conserved positions which are not conserved among other ePK superfamily members. The differing levels of sequence conservation found at different positions along the catalytic domain have facilitated the use of the polymerase chain reaction (PCR) and other homology-based molecular techniques for finding novel superfamily members while at the same time allowing for some specificity when searching for members of distinct kinase families or subfamilies.

The presence of twelve invariant residues within the different members of the ePK superfamily immediately suggests their possible involvement in catalysis. Taking into account the conservation of the twelve invariant residues and the overall conservation of the twelve subdomains, it was predicted that the kinase catalytic domains would fold into a similar three-dimensional conformation (Hanks *et al.*, 1988; Hanks and Quinn, 1991). Having deduced, through X-ray crystallography, the three-dimensional structure of the cAMP-activated protein kinase-C α subunit in a binary complex with a pseudosubstrate peptide inhibitor (Knighton *et al.*, 1991) and in

a ternary complex which also included MgATP (Bossmeier *et al.*, 1993; and Zheng *et al.*, 1993), it has been possible to elucidate how the catalytic domain folds into a three-dimensional conformation which allows it to carry out the three separate roles which are necessary for phosphate transfer (Hanks and Hunter, 1995): binding and orientation of the protein substrate, binding and orientation of the nucleotide triphosphate donor (ATP or GTP) as a complex with a divalent cation (usually Mg^{2+} or Mn^{2+}), and transfer of the γ -phosphate from the nucleotide triphosphate donor to the acceptor hydroxyl residue (serine, threonine, or tyrosine) on the protein substrate. The catalytic domain forms a two-lobed structure with a deep cleft running between the two lobes (Knighton *et al.*, 1991): the smaller lobe comprises the N-terminal half of the catalytic domain and is mostly involved in binding and orienting the nucleotide; the larger C-terminal lobe is involved in binding the substrate and transferring the phosphate; and finally, the cleft is the site of catalysis. The invariant residues come together in three-dimensional space and it has been possible to assign a putative functional role to each (Knighton *et al.*, 1991; Bossmeier *et al.*, 1993; and Zheng *et al.*, 1993). Another notable result obtained from the X-ray crystallography data has been that the conserved subdomains determined from the multiple sequence alignments correspond quite well to secondary and tertiary structure motifs of the catalytic domain (Reviewed in Wei *et al.*, 1994; Hanks and Hunter, 1995). The crystal structures of additional ePKs have been reported (reviewed by Goldsmith and Cobb, 1994) and their kinase domains have been found to fold into a two-lobed structure similar to that of the cAMP-activated protein kinase-C α subunit.

1.3.4 Protein kinase cascades and their role in signal transduction.

Every signal transduction pathway studied so far has had, as its central component, a phosphorylation cascade mediated by protein kinases. Thus, in many respects, protein kinase

pathways are synonymous with signal transduction. Three paradigms for mammalian cellular signalling, the transduction pathways which involve adenylate cyclase, inositol phospholipids, and the receptor tyrosine kinases, serve to illustrate this concept very well and have provided a framework with which novel pathways have been studied (Barrit, 1992).

The adenylate cyclase signalling pathway provides an excellent example of the interplay between the various components involved in extracellular and intracellular communication. Intracellular metabolism is regulated by many hormones that modulate their effects on target cells by activating the adenylate cyclase signalling pathway. Epinephrine, a hormone implicated in many pathways including glycogen metabolism, binds to the β -adrenergic receptor of cells after it is released from the adrenal gland. Conformational changes subsequently occur in the receptor, allowing the receptor to interact with the transducer (the G-protein). The G-protein becomes activated following the formation of the receptor-ligand complex and then binds to the effector enzyme, adenylate cyclase. Adenylate cyclase, a transmembrane enzyme, catalyzes the formation of the second messenger cAMP from ATP. cAMP then leaves the membrane surface (on the cytosolic side) and travels into the cytosol where it activates the cAMP-dependent protein kinase A (PKA), a member of the serine-threonine family of kinases. As intracellular levels of cAMP increase due to continued adenylate cyclase activity, activation of protein PKA causes the alteration of cellular metabolism as it catalyzes the phosphorylation of a number of target enzymes (Barrit, 1992).

1.3.5 The phosphoprotein phosphatases.

Protein phosphorylation and its role in cellular signalling have been very well established. However, in order for protein phosphorylation to play a role in signal transduction, phosphorylation must be reversible in nature such that phosphorylation events can act as a molecular switch (Krebs,

1994). Reversible phosphorylation not only requires the action of protein kinases that phosphorylate target proteins at specific sites, but also requires the action of protein phosphatases that remove these phosphates (Hunter, 1995). Thus changes in the phosphorylation state of a particular protein are not the result of static, singular events, but rather the results of shifts in the equilibrium between phosphorylation by the cognate protein kinases and dephosphorylation by protein phosphatases (Hunter, 1995). These shifts in equilibrium can be regulated by changing the activity of the protein kinase, the protein phosphatase, or both (Hunter, 1995).

There are three structurally distinct families of protein phosphatases: the PPP and PPM families, which are both serine/threonine phosphatases, and the protein tyrosine phosphatase family, PTP. (Only the serine/threonine families will be discussed in this review). The PPP family consists of PP1, PP2A and PP2B that all share a common catalytic domain of 280 residues (Barford *et al.*, 1998). PP1 is involved in the regulation of many cell functions, among which are: glycogen metabolism, muscle contraction, neuronal activities, cell cycle progression and RNA-splicing (Barford *et al.*, 1998). PP2A is similar to PP1, although it is known to regulate pathways that include metabolism, cell signalling, the cell cycle and telomerase activity (Wera and Hemmings, 1995). Mayer Jaekel *et al.* (1993) confirmed the role of PP2A in *Drosophila* cell division by mutating the 55-kDa regulatory β -subunit of PP2A, which resulted in a defective anaphase stage in meiosis. PP2B is characterized by its dependence on Ca^{2+} for its activity and it plays a crucial role in the Ca^{2+} signalling cascade of activated T-cells. Antigen presentation to the T-cell receptor causes an increase in T-cell Ca^{2+} concentrations, thereby stimulating PP2B to dephosphorylate the cytosolic subunit of the transcription factor NFAT1 (Jain *et al.*, 1993). The PPM family of serine/threonine phosphatases is found in both eukaryotes and prokaryotes. PP2C, the archetypal PPM phosphatase, is involved in the reversal of stress-induced protein kinase cascades. In the

fission yeast, for example, the negative regulation of the PBS2/HOG1-MAP kinase pathway, which is activated by osmotic and heat shock, is performed by PP2C (Maeda, 1994).

One example very relevant to this thesis serves to illustrate the interplay between kinases and phosphatases. Increases in the activity of the cAMP-dependent protein kinase (PKA) by raised intracellular levels of the allosteric effector cAMP have been associated with increased phosphorylation and concomitant decrease in the activity of the enzyme Acetyl-CoA carboxylase (ACC). The inactivation of ACC was originally thought to be the result of direct phosphorylation by the PKA. It has now been shown that the sites phosphorylated by PKA have no appreciable effect on ACC activity and that the phosphorylation site which affects ACC activity is phosphorylated by the AMP-activated protein kinase and not PKA (Sim and Hardie, 1988). Nevertheless, PKA can indirectly affect the phosphorylation level of this site by phosphorylating and inactivating the protein phosphatase that normally keeps this site dephosphorylated (Reviewed by Cohen and Hardie, 1991).

1.4 The SNF1/AMPK family of protein kinases. A historical overview.

The yeast SNF1 (for **S**ucrose **N**on-**F**ermenting **1**) kinase was initially identified by several groups in a screen for mutants defective in various aspects of carbon metabolism (reviewed by Hardie *et al.*, 1998). The *SNF1* locus was mapped by Celenza and Carlson (1984a) and subsequent cloning and characterization of the *SNF1* gene (Celenza and Carlson, 1984b; 1986) revealed that the Snf1 protein is a serine-threonine protein kinase, the first in what was to become a family of related protein kinases.

In 1991, Alderson *et al.* reported the cloning of *RKIN1*, a putative *SNF1* homologue isolated from a rye endosperm cDNA library. In addition to showing high sequence similarity to *SNF1*, expression of the *RKIN1* sequence in yeast *snf1* mutants restored their ability to grow on glycerol- and ethanol- containing medium, indicating a possible conservation in function. This report was later followed by a series of papers published between late 1993 and 1995, which reported the cloning and molecular characterization of different mammalian AMPK catalytic and non-catalytic subunits (reviewed by Hardie *et al.*, 1998). The high levels of sequence similarity made it apparent that the different components of the AMPK complex were mammalian homologues of components of the yeast SNF1 protein kinase complex. Similarities at the biochemical level were also reported, and these findings have served to define a new subfamily of protein kinases: the AMP-activated/SNF1 kinase subfamily. These protein kinases are the main components of a phosphorylation cascade conserved in all eukaryotic organisms and whose main physiological role appears to be a response in cellular stresses which compromise the energy status in the cell (Hardie *et al.*, 1998). As I will discuss later, the mostly biochemical nature of studies on AMPK combined with the mostly genetic characterization of SNF1 has allowed the development of models and

hypotheses which would have been difficult to develop otherwise. I will discuss the roles of yeast SNF1 and mammalian AMPKs in more detail in the following sections, and I will follow that by discussing their characteristics held in common, along with a discussion of homologues in other organisms.

1.4.1 The role of SNF1 in the yeast glucose repression.

The yeast *SNF1* gene was originally identified from yeast cells unable to grow on fermentable sugars such as sucrose. Celenza *et al.* (1986, 1989) characterized these mutations at the molecular level and revealed them to be mutations in a protein kinase gene. The role of this protein kinase in the glucose repression pathway is discussed below.

1.4.1.1 Glucose repression in yeast metabolism. Glucose repression, a regulatory system which is common in unicellular organisms, allows for the concerted regulation of a large number of genes following exposure to glucose or other rapidly fermentable sugars such as fructose and mannose (Entian, 1986; Ronne, 1995; Gancedo, 1998). *S. cerevisiae* cells which are in stationary phase or which are growing on non-fermentable carbon sources need to make a number of metabolic changes when presented with glucose or other rapidly-fermentable sugars. Much in the way mitogens, hormones, or other primary messengers affect mammalian cells, exposure to glucose leads to a number of wide and drastic effects on yeast cell metabolism (Thevelein, 1994). A number of enzymes and transport proteins are activated or inactivated by post-translational modification and the transcriptional induction or repression of many genes also takes place (Thevelein, 1994; Ronne, 1995). The glucose repression pathway down-regulates genes encoding enzymes that are unique to the glyoxylate shunt and gluconeogenesis, enzymes that are involved in the tricarboxylic acid cycle

(TCA cycle) and oxidative phosphorylation, and the proteins that take up and metabolize carbon sources which are utilized less efficiently than glucose (Thevelein, 1994). Even in the presence of oxygen, *S. cerevisiae* cells rapidly ferment glucose to ethanol while generating a modest amount of ATP through glycolysis. The ethanol generated during fermentative growth is extruded into the extracellular medium; once the glucose supply is exhausted, the ethanol can be transported back into the cell where it undergoes complete oxidation by being channelled back into the TCA cycle and respiration. Yeast cells thus benefit by being able to transform glucose into ethanol, effectively “sequestering” the reduced carbon in a compound that can be both toxic and non-metabolizable to other competing organisms (Ronne, 1995).

Yeast glucose repression has been extensively characterized, revealing the deceptive complexity of the pathway. Many details, such as the nature of the transduced signal, as well as some intermediate steps along the pathway, are still unknown (Gancedo, 1998). Much of the early characterization of the glucose repression pathway has been centered on the isolation and analysis of mutants with defects in glucose repression (for examples see Zimmermann and Scheel, 1977; Neugeborn and Carlson, 1984). This approach, along with the mapping, identification, and molecular characterization of the genes affected in such mutants has proved extremely powerful in helping determine the *in vivo* role of some of the major players in the pathway (Gancedo, 1998). These approaches have uncovered a bewildering number of genes that are involved, at least to some extent, in glucose repression. Despite the fact that a role for a vast majority of these genes is still elusive, it has been possible to establish a working model for glucose repression (Thevelein, 1994; Ronne, 1995; Gancedo, 1998).

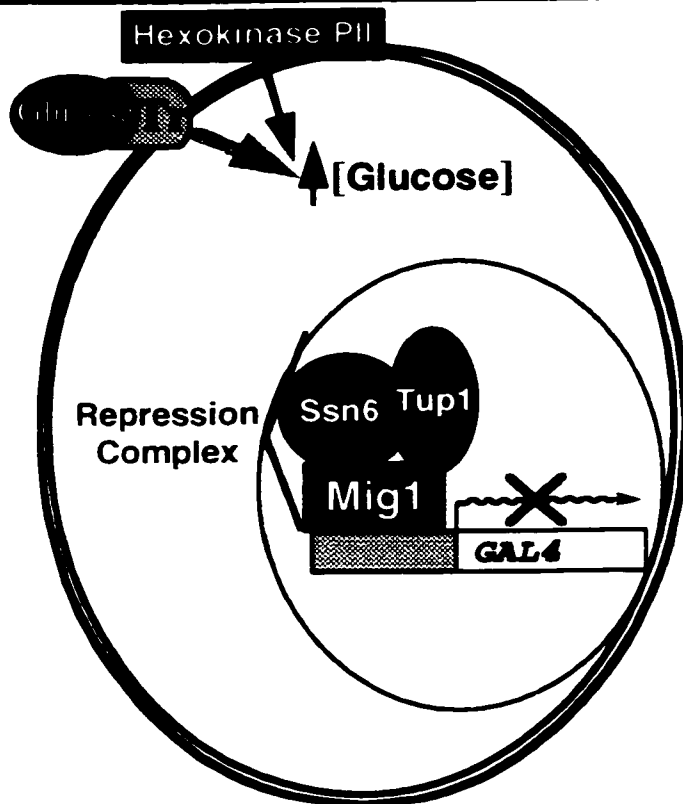
1.4.1.2. The glucose repression pathway: generation of the glucose signal. Glucose repression signalling requires the detection of glucose and the generation of a signal that can be transduced to downstream elements of the pathway. Yeast has a number of different hexose transporter (*HXT*) genes that are necessary for glucose uptake (reviewed in Boles and Hollenberg, 1997). Two of these, encoded by the genes *SNF3* and *RGT2*, have been strongly implicated in generating an intracellular glucose-sensing signal that induces repression (Ozcan *et al.*, 1996; 1998). Snf3 and Rgt2 are necessary for the low glucose- and high glucose-induced expression of the high-capacity glucose transporter genes, respectively (Ozcan *et al.*, 1996). Unlike the other yeast hexose transporters, Snf3 and Rgt2 have long carboxy-terminal tails that extend into the cytoplasm. Ozcan *et al.* (1996) have hypothesized that such an intracellular extension may be involved in interacting with the next component of the signalling pathway in response to glucose binding. In subsequent work they have established that the cytoplasmic tails are essential for transducing the glucose signal (Ozcan *et al.*, 1998), although the downstream events are still poorly understood. One possible target of this interaction is the hexokinase PII (Hxk2) which, in the first step of glycolysis, catalyzes glucose phosphorylation. It is known that *hvk2* mutants in which the Hxk2 still bears considerable catalytic activity are defective in glucose repression despite the presence of two other glucose-phosphorylating enzymes in yeast, hexokinase PI and glucokinase (Entian, 1980). Rose *et al.* (1991) have also found that mutants with minimal glycolytic flux (by reason of phosphoglucosomerase mutations) are still able to carry out glucose repression. It would thus appear that glucose repression signalling does not involve any metabolic step beyond the initial phosphorylation reaction catalyzed by the hexokinase PII (Gancedo, 1998).

1.4.1.3 The glucose repression pathway : the downstream effectors. The downstream effector of the glucose repression signal is encoded by the *MIG1* gene (Multicopy Inhibitor of *GAL* gene expression), which was identified because of its ability to turn off transcription from the *GALI* promoter when present in a multicopy plasmid (Nehlin and Ronne, 1990; Nehlin *et al.*, 1991). The Mig1 protein is a C2-H2 zinc-finger transcription factor related to the mammalian Wilms' tumour (Call *et al.*, 1990) and Egr (early growth response) transcription factors (Joseph *et al.*, 1988) and also has counterparts in *Aspergillus nidulans* (Dowzer and Kelly, 1991) and *Kluyveromyces lactis* (Cassart *et al.*, 1995). Mig1 binds to the promoters of a number of glucose-regulated genes and leads to their repression by forming a complex with the proteins encoded by *TUP1* and *SSN6* (*CYC8*) [refer to Fig I.A], which are also necessary for glucose repression (Vallier and Carlson, 1994; Treitel and Carlson, 1995). Tup1 and Ssn6 carry out the actual repression by coming together to form high-molecular-weight complexes with a stoichiometric ratio of 4:1 (Varanasi *et al.*, 1996). Ssn6/Tup1 are thought to act as general corepressor complexes involved in regulating distinct sets of genes (Keleher, 1992) and their disruption not only abolishes glucose repression but it causes pleiotropic defects in cellular functions which are unrelated to glucose repression, including abnormal cell morphology, flocculence, altered mini-chromosome maintenance, mating and sporulation defects (Johnston and Carlson, 1992). In addition, expression of glucose-repressible genes in *snf1* mutants, which are unable to carry out glucose derepression, is restored by an additional mutation in *SSN6* (Carlson *et al.*, 1984; Wright and Poyton, 1990). Although neither one of these two factors has apparent DNA binding domains, the Tup1/Ssn6 complex is recruited to different promoters by pathway-specific DNA-binding proteins, where two mechanisms are thought to provide repression: the assembly of nucleosome arrays which lead to a repressive chromatin structure (Cooper *et al.*, 1994) and interactions with the general transcription machinery (Redd *et*

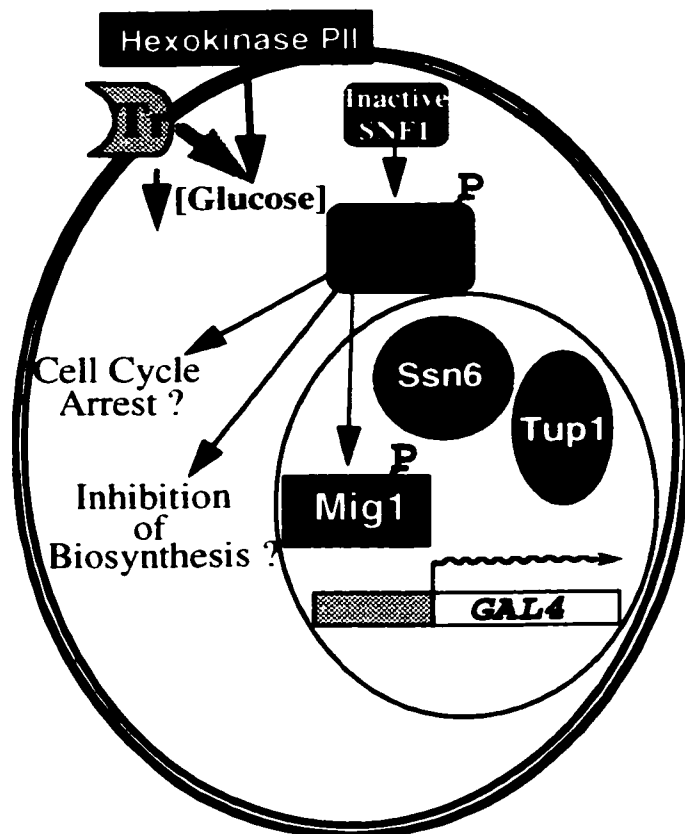
Figure I. An overview of the yeast glucose repression mechanism.

(A) High extracellular and intracellular levels of glucose are detected by a sensing mechanism that may involve the glucose transporters Snf3 and Rgt2 (Tr) and Hexokinase PII. Under high glucose conditions (repressing conditions) a repression complex prevents transcription from glucose repressible promoters: Ssn6 and Tup1 form a general repressor complex that is recruited to glucose repressible genes by the transcription factor Mig1, leading to transcriptional shut-down of these genes. (B) When glucose becomes limiting (derepressing conditions) the SNF1 kinase complex is activated through phosphorylation by an upstream kinase-kinase. The active SNF1 kinase complex phosphorylates Mig1, thereby inactivating the repressor complex and restoring gene expression of glucose repressed genes. In addition, SNF1 is thought to also inactivate some biosynthetic pathways thus conserving ATP and it may also be involved in the cell-cycle arrest that occurs under low-glucose conditions. [Figure based on Thevelein (1994) and Ronne (1995)].

I.A



I.B



al., 1997). Tup1 is sufficient for establishing repression and Edmonson *et al.* (1996) have provided evidence that it interacts with histones H3 and H4. The role of Ssn6 appears to be that of a coupler between Tup1 and the various targeting proteins, including Mig1 (Tzamarias and Struhl, 1995). Two-hybrid studies carried out by Treitel and Carlson (1995) and Tzamarias and Struhl (1995) have confirmed an interaction between Ssn6 and Mig1 that is necessary for glucose repression. Mig1-mediated repression of a reporter gene occurs only in glucose-grown cells and repression is Ssn6- and Tup1-dependent (Treitel and Carlson, 1995). Analysis of various Mig1 deletion-derivatives (Ostling *et al.*, 1996) has shown that as little as 24 residues from the Mig1 carboxy-terminal domain are sufficient to mediate repression and thus, this is all that is required to interact with the corepressor complex.

1.4.1.4 The glucose repression pathway: the role of SNF1 in derepression. When glucose becomes limiting, glucose repressed genes must be derepressed. A signal is generated which leads to the phosphorylation and activation of the SNF1 protein kinase complex [refer to Fig I.B], which is necessary for the derepression of most glucose-repressed genes (Celenza and Carlson, 1986; 1989). The kinase activity of the SNF1 complex is responsible for the derepression of glucose-repressible genes and this is thought to occur through the phosphorylation of Mig1 and inactivation of its ability to recruit the Ssn6/Tup1 corepressor complex (Treitel and Carlson, 1995) [refer to Fig I.B].

Mutants deficient in SNF1 activity are unable to grow efficiently on most carbon sources other than glucose and are generally unhealthy even when grown on glucose (Carlson *et al.*, 1984; Johnston and Carlson, 1992). Mutants at another locus, *snf4*, have defects in glucose repression similar to those of *snf1* mutants (Neugeborn and Carlson, 1984). Snf4 is a necessary activator of the catalytic activity of the full length Snf1 (Celenza and Carlson, 1989; Woods *et al.*, 1994), although deletion of the Snf1 regulatory domain still retains glucose regulation while obviating the need for

Snf4 (Celenza and Carlson, 1989; Jiang and Carlson, 1997). Jiang and Carlson (1996) used two-hybrid analysis to investigate the Snf1/Snf4 interactions in response to the glucose signal and found that this interaction was regulated by glucose and detected only under low glucose conditions. Snf1 and Snf4 could only interact in high glucose if the two-hybrid experiments were carried out in mutant backgrounds with known glucose repression defects upstream of *SNF1* (such as *hxx2*, *glc7-T152K*, *reg1*). Finally, the Snf1/Snf4 two-hybrid interaction was unaffected in mutants with repression defects downstream of *SNF1* (such as *ssn6*). The Snf1/Snf4 interactions observed are consistent not only with the proposed role of Snf4 in activating kinase activity, but also with the interaction of the SNF1 complex with other components of the glucose-signalling pathway.

1.4.1.5 The Snf1 kinase forms part of a heterotrimeric complex. Although early biochemical evidence suggested that Snf1 is associated with Snf4 regardless of glucose availability (Celenza *et al.*, 1989; Woods *et al.*, 1994; Wilson *et al.*, 1996) two-hybrid analysis has established that direct Snf1/Snf4 interaction occurs only in low glucose conditions (Jiang and Carlson, 1994; 1997). Snf1 and Snf4 are thought to remain associated in high-molecular-weight complexes during high glucose conditions through their separate interactions with a family of related proteins encoded by the *SIP1*, *SIP2*, and *GAL83* genes (Yang *et al.*, 1994; Jiang and Carlson, 1997) [refer to Fig II..A]. Genetic evidence suggests that each of the members of this family is an alternate component of SNF1 complexes. Since they individually affect distinct subsets of glucose-repressed genes normally reactivated by the SNF1 complex, their possible role is as targeting subunits that direct the complex to specific downstream targets (Yang *et al.*, 1992; Erickson and Johnston, 1993). The members of this family share two conserved domains each of which is responsible for separate interactions with the Snf1-RD and Snf4, respectively (Jiang and Carlson, 1997), as shown by two-hybrid analysis and *in vitro* binding assays. The functions mediated by the SNF1 complex are largely unaffected in a

sip1.sip2.gal83 triple mutant (Yang *et al.*, 1994) and the SNF1 complex retains full catalytic activity. Jiang and Carlson (1997), however, have shown that the level of detectable Snf1 and Snf4 that copurify together in the triple mutant is only half of that found in wild-type cells. This further supports the idea that one of the roles of the Sip1/Sip2/Gal83 family is in helping stabilize the Snf1-Snf4p interaction (Jiang and Carlson, 1997). As I will discuss later, the different components of the SNF1 complex (Snf1, Snf4, and Sip1/Sip2/Gal83) each have mammalian counterparts.

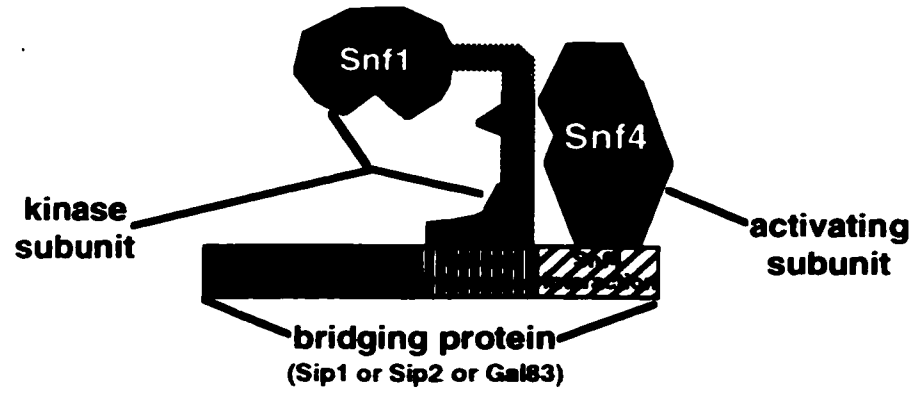
A number of ePKs have a regulatory, or auto-inhibitory, C-terminal domain (RD) which is responsible for interacting with the conserved N-terminal catalytic domain (CD) thereby inhibiting kinase activity (Hunter and Hanks, 1995) [refer to Fig II.B]. The inhibition is the result of an interaction between the catalytic site and a pseudo-substrate motif in the RD domain that resembles the substrate recognition sequence for the kinase (Hunter and Hanks, 1995). In order to map the region of interaction between Snf1 and Snf4, Jiang and Carlson (1997) expressed the Snf1 catalytic domain (Snf1-CD) and Snf1 regulatory domain (Snf1-RD) separately and tested for two-hybrid interactions. Snf4 did not interact with the Snf1-CD, whereas the Snf4/Snf1-RD interaction was readily detected under both low and high glucose conditions, suggesting that Snf4 interacts with Snf1 through its Snf1-CD domain. Under high glucose conditions an Snf4-independent interaction between the Snf1-CD and the Snf1-RD occurs. Mapping of the region of interactions showed that the Snf1-CD and Snf4 each interact with the Snf1-RD by binding to distinct yet overlapping sites with opposite regulation by the glucose signal. The high glucose Snf1-CD/Snf1-RD interaction prevents the Snf1-RD from interacting with Snf4 and, more importantly, it leads to inhibition of kinase activity [refer to Fig II.B]. Under low glucose conditions, the situation is reversed and, by interacting with the Snf1-RD and preventing auto-inhibition, Snf4 activates catalytic activity of the Snf1-CD (Jiang and Carlson, 1997) [refer to Fig II.C]. Since Snf1 catalytic activity increases under

Figure II. Structure of the SNF1 kinase complex and glucose-regulated interactions.

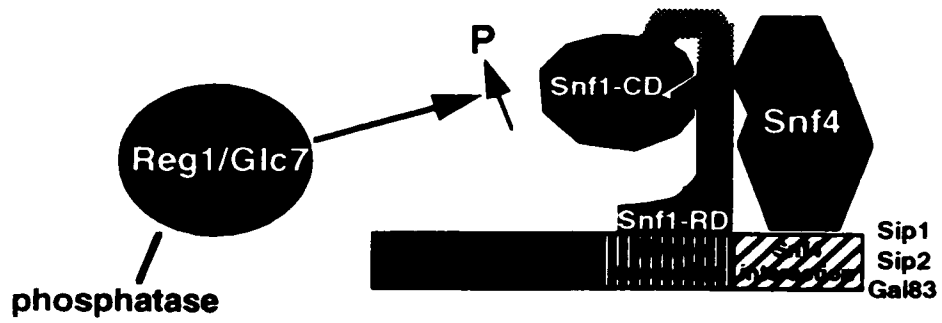
(A) The SNF1 kinase is a ternary complex composed of: Snf1, the kinase subunit; Snf4, the kinase activating domain; a bridging protein which tethers the two together (three different bridging proteins exist in yeast: Sip1, Sip2, and Gal83). The SNF1 kinase complex subunits undergo differential interactions in response to glucose availability and this has an effect on kinase activity.

(B) In high glucose, the Reg1/Glc7 protein phosphatase complex keeps Snf1 dephosphorylated and the Snf1 kinase catalytic domain [Snf1-CD] interacts with the Snf1 (auto) regulatory domain [Snf1-RD], which keeps the SNF1 kinase complex in an inactive state. (C) In low glucose, kinase-kinase [Snf1-K] phosphorylates Snf1, changing its conformation such that the Snf1-CD no longer interacts with the Snf1-RD, which activates the SNF1 kinase complex. The Snf4 subunit is responsible for maintaining the activated state by blocking the Snf1-CD/Snf1-RD region of interaction. [Adapted from Hardie *et al.*, 1998]

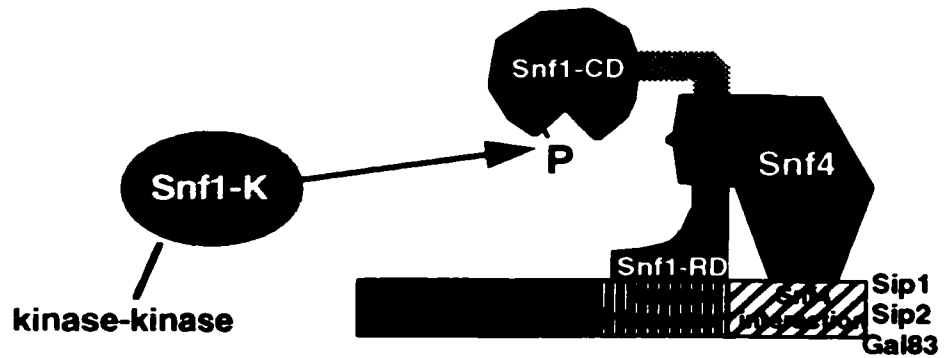
II.A



II.B



II.C



glucose-limited conditions and since the Snf1-CD is sufficient for glucose-regulated catalytic activity, the catalytic domain is presumably one of the targets of the glucose signal (Jiang and Carlson, 1997). Therefore while Snf4 is required to disrupt the auto-inhibitory effect that the Snf1 regulatory domain has on the kinase catalytic domain the glucose signal can regulate Snf1 activity independently of Snf4.

1.4.1.6 Glucose repression: targets of the SNF1 kinase complex. The kinase activity of the SNF1 complex is greatly increased after the switch from high to low glucose conditions (Woods *et al.*, 1994; Wilson *et al.*, 1996). Most *in vivo* targets remain unidentified and the SNF1 complex may phosphorylate proteins other than those directly related to glucose repression. Genetic evidence, however, suggests that *MIG1* and *SNF1* act in antagonistic fashion and recent studies suggest that Mig1 is under the negative control of Snf1 (Treitel and Carlson, 1995; Ostling *et al.*, 1996; Treitel *et al.*, 1998).

Mig1 is differentially phosphorylated in response to glucose levels (Treitel and Carlson, 1995; DeVit *et al.*, 1997). DeVit *et al.* found that glucose removal leads to Mig1 translocation from the nucleus to the cytoplasm and that *snf1* mutants have constitutive nuclear localization of Mig1. In a pivotal paper published last year, Treitel *et al.* (1998) describe a series of experiments designed to cement the roles of *MIG1* and *SNF1* in glucose repression and derepression, respectively, and also designed to test the involvement of other genes implicated in the pathway through previous genetic studies. Treitel *et al.* (1998) created Mig1 and Ssn6 fusion proteins with the heterologous DNA-binding domain LexA and monitored the expression from a non-glucose-repressible promoter bearing *lexA* operators in a number of different mutants. While a LexA-Ssn6 fusion confers *TUP1*-dependent and *MIG1*-independent repression on the *lexA*-containing promoter regardless of glucose status, the *TUP1/SSN6*-dependent repression conferred by a LexA-Mig1 fusion is glucose-

dependent. Repression mediated by the latter is no longer alleviated by a shift to glucose-limiting conditions in an *snf1* mutant, indicating a direct effect of Snf1 on Mig1 function. Treitel and Carlson (1995) had previously shown that Mig1 undergoes phosphorylation under low glucose conditions. Treitel *et al.* (1998) determined that Snf1 is required for this phosphorylation by showing that it is abolished in *snf1* mutants. Conversely since, as discussed previously, *reg1* and *hxx2* mutations both function upstream of *SNF1* and result in an inability to carry out glucose repression, Snf1 remains active under high glucose conditions in *reg1* and *hxx2* mutants. Consistent with this, Treitel *et al.* (1998) found that Mig1 remained phosphorylated under high glucose conditions in *reg1* and *hxx2* mutants. To further establish the Snf1-dependent phosphorylation of Mig1, Treitel *et al.* (1998) also verified direct Mig1/Snf1 interaction in two-hybrid and co-immunoprecipitation studies, and showed that Mig1 is phosphorylated *in vitro* in an Snf1-dependent reaction that can be reduced if potential Snf1 phosphorylation sites on Mig1 are removed by serine to alanine mutations.

1.4.1.7 The role of protein phosphatases in glucose repression. Two other mutations have also been shown to affect glucose repression by relieving glucose-repressed genes in *snf1* mutants that would be otherwise unable to derepress them. Molecular characterization of these two mutants, *GLC7* and *REG1*, has shown that they encode a multisubstrate type-1 protein phosphatase (PP1) (Ohkura *et al.*, 1989; Feng *et al.*, 1991) and one of its regulatory subunits (Niederacher and Entian, 1991), respectively. PP1 regulates a number of cellular processes and several mutations that do not impair its catalytic function nonetheless affect some of its regulatory processes while leaving others unaffected. Tu and Carlson (1994) established that one known mutation, *glc7-T152K*, partially relieves glucose repression of invertase without affecting glycogen synthesis, whereas another

mutation, *glc7-1*, impairs glycogen synthesis without affecting glucose repression of invertase. Tu and Carlson (1994) suggested that these mutations might affect Glc7 function by impairing its interaction with regulatory subunits that direct it to distinct subsets of targets. Although previously cloned by Niederacher and Entian (1991), the actual role of Reg1 remained elusive because its deduced amino acid sequence did not hint at potential functional features. On the basis of similar phenotypes shown by *reg1* and *glc7-T152K* (i.e. partial relief of invertase glucose repression and suppression by *snf1*), Tu and Carlson (1995) searched for a possible Reg1/Glc7 interaction.

Genetic evidence for the interaction was obtained when overexpression of *REG1* was found to counteract the glucose repression defects of the *glc7-T152K* mutant (Tu and Carlson, 1995). Two-hybrid analysis and co-immunoprecipitation studies confirmed the actual physical interaction and, as expected, Tu and Carlson also found that the *glc7-T152K* mutation markedly decreased the Reg1/Glc7 two-hybrid interaction. Genetic evidence suggests that the SNF1 and Glc7 complexes play antagonistic roles in glucose repression. It is still unknown whether the Glc7 complex acts by dephosphorylating Snf1 [as proposed by Tu and Carlson, (1995)], by dephosphorylating regulators of Snf1, and/or dephosphorylating targets of Snf1 [perhaps Mig1, as proposed by DeVit *et al.* (1997)]. There is evidence that the Reg1 targeting subunit prevents the interaction between the Snf1-RD and Snf4 in low glucose, which “frees up” the catalytic domain thereby activating the kinase. *REG1* mutations relieve the inhibitory effect of high glucose on Snf1/Snf4 two-hybrid interactions (Jiang and Carlson, 1996) and lead to constitutive Snf4-dependent Snf1 kinase activity (Ludin *et al.*, 1998). Ludin *et al.* (1998) have also reported that Reg1 interacts with the Snf1-CD in two-hybrid assays under low glucose conditions, when the SNF1 complex is expected to be active. This interaction is prevented by a mutation in the putative phosphorylation site of the kinase

activation loop [the T-loop, as described by Hunter and Hanks (1995)], a mutation that abolishes Snf1 function *in vivo* and *in vitro* (Estruch *et al.*, 1992). This mutation also weakens the Snf1/Snf4 interaction (Ludin *et al.*, 1998), which would suggest that phosphorylation in the Snf1 T-loop is required for both the Snf1-CD/Reg1 and Snf1-RD/Snf4 interactions. In combination, these two effects would be expected to activate the kinase by preventing auto-inhibition resulting from the Snf1-CD/Snf1-RD interaction.

In the current model of SNF1 complex regulation by Reg1/Glc7 proposed by Ludin *et al.* (1998), after SNF1 is activated by the glucose signal through phosphorylation by an SNF1 kinase-kinase [refer to Fig II.C], Reg1 binds to the Snf1-CD and recruits the Glc7 phosphatase to the kinase complex. The phosphatase can then return the kinase complex to the inactive form by dephosphorylating Snf1 or other components of the complex [Refer to Fig II.B]. One interesting result obtained by Ludin *et al.* (1998) was that *reg1Δsnf1Δ* double mutants transformed with a plasmid encoding the Snf1 catalytic domain showed glucose-repressible expression of *SUC2* and *GALI-lacZ* while the same mutant transformed with full-length Snf1 expressed these genes constitutively. These results would imply that the Snf1-CD receives the glucose signal independently of Reg1 and Snf4.

As a result of the amenability of yeast genetic studies, the bulk of the characterization of the glucose repression pathway has been in the form of *in vivo* genetic studies seeking to establish the relationship between different members of the pathway (Gancedo, 1998). Although many details remain to be elucidated, the observations from the various studies cited here provide compelling evidence for the *in vivo* interaction between the different signalling, repression, and derepression components involved in glucose repression. In particular, they have established the importance of the Mig1/Snf1 interplay in repression and derepression, respectively.

1.4.2 The role of the AMPK in mammalian lipid metabolism.

Biochemical characterization of the AMP-activated protein kinase (AMPK) can be traced back to observations made by Carlson *et al.* (1973) and Beg *et al.* (1973; 1978) and which can now be ascribed to the activity of AMPK. Carlson *et al.* (1973) had observed that their crude preparations of Acetyl-CoA carboxylase (ACC) were inactivated by a protein kinase contaminant, the ACC kinase, in a process which was later found to be stimulated by 5'-adenosine monophosphate (AMP) (Yeh *et al.*, 1980). Meanwhile, Beg *et al.* (1973) had also observed that a crude protein fraction inactivated their microsomal preparations of 3-hydroxy-3-methylglutaryl-CoA reductase (HMG-CoA reductase or HMGR) in a time-dependent process that required ATP. Later studies showed that this inactivation was brought about by phosphorylation, presumably through the actions of a protein kinase that was dubbed the HMGR kinase (Beg *et al.*, 1978; Keith *et al.*, 1979). It soon became apparent that the HMGR kinase was itself regulated through phosphorylation (Ingebritsen *et al.*, 1978) by an upstream kinase kinase (Beg *et al.*, 1979; Ingebritsen *et al.*, 1981) in what, at the time, represented only the second protein kinase cascade to be reported. It was later found that the rat liver HMGR kinase was also stimulated by AMP (Ferrer *et al.*, 1985).

Although a number of groups continued the biochemical characterization of the ACC kinase and HMGR kinase by using partially purified preparations, the two kinases were to remain uncharacterized at the molecular level for more than a decade after their initial discovery. At the time Hardie and co-workers had been studying ACC regulation and characterizing several different potential ACC kinase activities. One of their kinases, ACCK 3, was found to be a cAMP-independent/AMP-dependent kinase purified from rat liver (Carling *et al.*, 1987). Munday *et al.* (1988) discovered that ACCK 3 was the only kinase activity that could inactivate ACC at saturating

concentrations of citrate, an allosteric activator of ACC. This partially purified kinase was later found to also phosphorylate and inactivate HMGR. Both activities were stimulated by AMP and regulated by phosphorylation and thus, based on these shared characteristics, Carling *et al.* (1987) first proposed that ACC and HMGR were inactivated by the same kinase. This hypothesis was later confirmed when the two protein kinase activities were copurified to near-homogeneity from rat liver (Carling *et al.*, 1989). The two kinase activities were found in the same fraction after a 6 step/4800-fold purification. The preparation contained polypeptides of apparent molecular masses of 63-, 38-, and 35-kDa but only the 63-kDa band was detected by affinity labelling using the reactive ATP analogue [¹⁴C]fluorosulphonylbenzoyl adenosine (Carling *et al.*, 1989). Incubation with the reactive analogue also inactivated both ACC and HMGR kinase activities with identical kinetic parameters (Carling *et al.*, 1989). Because this new kinase appeared to have two physiological substrates, Hardie *et al.* (1989) proposed the name of 5'-AMP-activated protein kinase (AMPK), in keeping with the naming of other multi-substrate protein kinases by their allosteric activator (e.g. cAMP and cGMP activated kinases). Two different groups subsequently purified the AMPK to homogeneity by using affinity column chromatography. Davies *et al.* (1994) purified porcine AMPK by performing binding in an ATP- γ -Sepharose column followed by elution with AMP, while Mitchelhill *et al.* (1994) purified rat AMPK by using a peptide-substrate affinity column. As had been reported previously by Carling *et al.* (1989), both groups detected the presence of nearly equimolar amounts of two polypeptides of apparent molecular mass of 38- and 35-kDa that copurified along with the 63-kDa catalytic polypeptide. Based on the co-migration of all three polypeptides on native gel electrophoresis and on gel filtration data, Davies *et al.* (1994) suggested that the three proteins represented subunits in a 1:1:1 heterotrimeric complex.

A significant development in the study of the AMPK was the development of a new and

specific assay for its activity that could overcome the significant shortcomings of previous assays. Assays based on the inactivation of HMGR in crude rat liver microsomes or after partial purification (Beg *et al.*, 1979; Ingebritsen *et al.*, 1978; Ferrer *et al.*, 1984) were insensitive, time-consuming and subject to interference by HMG-CoA lyase and mevalonate kinase (Ness *et al.*, 1980). Carling *et al.* (1987; 1989) assayed AMPK activity by measuring the incorporation of radioactivity from [γ - 32 P]ATP into purified ACC and, although more convenient and accurate, it required large amounts of purified ACC and was still subject to interference by the other endogenous ACC kinase activities (Hardie *et al.*, 1989). Hardie and co-workers developed a new assay involving the phosphorylation the SAMS peptide, a synthetic peptide with the sequence HMRSSAMSGLHLVKRR-amide (Davies *et al.*, 1989). This peptide is based on the sequence from His-73 to Lys-85 in rat ACC and which surrounds the AMPK phosphorylation site known to inactivate ACC activity [Ser-79] (Munday *et al.*, 1988). By changing the serine (Ser-77) in the original sequence to an alanine, they eliminated the site phosphorylated by the cAMP-dependent protein kinase (Munday *et al.*, 1988); adding the two extra arginines facilitated purification of the peptide (Davies *et al.*, 1989). Incorporation of radioactivity into the SAMS peptide has provided a rapid, highly sensitive, and very specific assay for AMPK activity (Davies *et al.*, 1989) which, as I will discuss in section 1.4.6, has been exploited for detecting AMPK-like activities from a number of different sources, including plants.

1.4.3 Molecular characterization of the mammalian AMPKs

The first report of molecular characterization of AMPK was in the form of preliminary amino acid sequence data obtained for the porcine AMPK 63-kDa catalytic subunit by Mitchelhill *et al.* (1994) who sequenced different peptides obtained from a highly purified porcine AMPK

preparation. Concurrently, Carling *et al.* (1994) independently obtained partial amino acid sequence data from the rat AMPK catalytic subunit and subsequently isolated clones from a rat liver cDNA library by screening with probe amplified by PCR using degenerate primers. The rat Ampk shares 59% identity with the yeast Snf1 and rye Rkin1 within the catalytic domain and, although the extent of the conservation beyond the catalytic domain is less extensive, conservation extends across the entire protein sequence (46% identity) (Carling *et al.*, 1994).

In view of the emergent similarities between the mammalian AMPKs and the yeast SNF1, Stapleton *et al.* (1994) investigated whether the two polypeptides which copurified with the AMPK catalytic subunit were related to proteins known to interact with Snf1. Using a combination of peptide sequencing and degenerate-primed PCR they obtained partial peptide and cDNA sequence data for both proteins from porcine and rat AMPK preparations and found that the porcine and rat 38-kDa polypeptides were homologous to Snf4 while the 35-kDa polypeptides were homologous to the Sip1/Sip2/Gal83 family of proteins. Stapleton *et al.* (1994) also performed cross-linking experiments and their results suggested that the 63-, 38-, and 35-kDa polypeptides form complexes, most likely as heterotrimers, which is in agreement with previous data by Davies *et al.* (1994). Stapleton *et al.* also proposed naming them α -, β -, and γ -subunits of the 5'-AMP-activated protein kinase, respectively. Subsequent cloning of the full-length cDNAs for the rat AMPK- β and human AMPK- γ non-catalytic subunits confirmed the homology to Sip1/Sip2/Gal83 and Snf4, respectively (Gao *et al.*, 1996). Woods *et al.* (1996) independently cloned and characterized cDNAs for the rat AMPK- β and AMPK- γ subunits and investigated interactions between the three subunits *in vitro*. Their two-hybrid assay results suggest that although the α - and γ -subunits do not interact with each other, the β -subunit is able to form a stable interaction with each of them. *In vitro* binding data confirmed the β/α and β/γ interactions and also showed that when all three subunits are co-

translated, an $\alpha/\beta/\gamma$ ternary complex is formed (Woods *et al.*, 1996). Dyck *et al.* (1996) investigated the role of the catalytic subunits by performing co-transfection studies in COS7 cells and monitoring the resulting AMPK activity. When compared to expression of catalytic AMPK- α subunit alone, co-expression of the α - with either β - or γ - subunits leads to modest increases in kinase activity, 1.5- and 2.5-fold respectively. Co-expression of all three subunits leads to a synergistic 50- to 110-fold activation in kinase activity, suggesting that full activation of AMPK- α requires all three subunits to come together in a ternary complex (Dyck *et al.*, 1996).

Based on subtle differences between the published cDNA sequence for the rat AMPK- α (Carling *et al.*, 1994; Gao *et al.*, 1994) and the sequence of various peptides obtained during a more thorough protein sequencing of the purified rat AMPK- α polypeptide, Stapleton *et al.* (1996) sought out and cloned the cDNA for a second α -subunit isoform. They called this isoform AMPK- α 1 to reflect the fact that its deduced ORF was identical in sequence to the various peptides derived from the original purified AMPK preparations. AMPK- α 1 and (the previously characterized) AMPK- α 2 are 90% identical within their catalytic domains while sharing 61% identity elsewhere. Although they show different expression patterns (Stapleton *et al.*, 1996) and small differences in peptide substrate specificity (Woods *et al.*, 1996), both isoforms contribute significantly to liver AMPK activity (Woods *et al.*, 1996) and have similar specific activity (Dyck *et al.*, 1996). Multiple isoforms of the β - and γ -subunits are also known to exist. Gao *et al.* (1996) have reported the presence of several distinct human cDNA sequences at GenBank's database of expressed sequence tags that are homologous to the AMPK- β and AMPK- γ sequences. These sequences have been isolated from a variety of tissues including heart, breast, placenta, and brain and are different from the published full-length β - and γ -subunit sequences isolated from liver by Gao *et al.* The full length

cDNA for a second β isoform, termed $\beta 2$, has recently been reported and shown to be highly expressed in skeletal muscle and lowly expressed in liver, the opposite pattern than is observed for $\beta 1$ (Thornton *et al.*, 1998). The presence of multiple α -, β -, and γ -subunit genes provides a pool of different isoforms which appear to assemble in heterogeneous ternary complexes (Thornton *et al.*, 1998) and functional differences between any of the components may thus provide an assortment of different *in vivo* roles (Hardie *et al.*, 1998). Stapleton *et al.* (1997) have used protein sequencing to show that $\alpha 2$ associates with $\beta 1$ and $\gamma 1$ in liver, whereas complexes purified from skeletal muscle contain $\alpha 2$, $\beta 2$, and $\gamma 1$. Vavvas *et al.* (1997) have reported that AMPK- $\alpha 2$, and not AMPK- $\alpha 1$, is activated in response to increased contraction in rat skeletal muscle. In addition to showing differences in the level of allosteric and AMPKK-mediated activation in response to AMP, and differences in their resistance to protein phosphatase PP2A treatment, Salt *et al.* (1998) have also recently shown that the two isoforms appear to show differences in subcellular localization.

1.4.4 Regulation of the mammalian AMPKs

Two mechanisms involved in regulating AMPK activity were first noted in the early studies of the (then known) ACC and HMGR kinases. Yeh *et al.* (1980) and Ferrer *et al.* (1985) first noted the stimulatory effects of AMP on the kinase-dependent inactivation of ACC and HMGR, respectively. Likewise, the involvement of an upstream activating kinase, the AMPKK, was first noted in some of the early studies of the HMGR kinase (Ingebritsen *et al.*, 1978; 1981; Beg *et al.*, 1979). Hawley *et al.* (1996) have subsequently shown that the AMPK is activated 50- to 100-fold by phosphorylation by the AMPKK and that the major site phosphorylated by AMPKK, Thr-172, is located within the activation segment where many protein kinases require phosphorylation for their activation (Hanks and Hunter, 1995). Although the allosteric stimulation of AMP on AMPK

activity is modest, the stimulatory effect of AMP on the activity of AMPK occurs through a more complex mechanism than had originally been thought. As it turns out, AMP activates the AMPK through four distinct effects (Carling *et al.*, 1989; Corton *et al.*, 1995; Davies *et al.*, 1995; Hawley *et al.*, 1995): (1) direct allosteric activation of the AMPK; (2) direct allosteric activation of the upstream AMPKK; (3) binding to AMPK thus making it a better substrate for the AMPKK; (4) binding to AMPK thus making it a worse substrate for the inactivating protein phosphatase 2C. High ATP levels inhibit at least three, if not all four, of the activating effects of AMP such that the system is activated by conditions of high AMP and low ATP (Corton *et al.*, 1995; Davies *et al.*, 1995; Hawley *et al.*, 1996). The involvement of AMP at several levels makes the system exceedingly sensitive to small changes in AMP concentration and recent modelling of the system by Hardie *et al.* (1999) has shown that the system displays an ultrasensitive response: only a 6-fold increase in activating nucleotide levels is required in order to increase AMPK activity from 10% to 90% of the maximal activity.

Under optimal conditions, eukaryotic cells maintain a high ATP:ADP ratio (10:1) owing to high levels of oxidative phosphorylation. In turn, high ATP levels lead to low AMP levels because the enzyme adenylate kinase catalyzes the reaction: $2\text{ADP} \leftrightarrow \text{ATP} + \text{AMP}$ and the reaction is maintained at near equilibrium with an equilibrium constant that is close to unity ($K_{\text{eq}} \approx 1 \approx \frac{[\text{ATP}][\text{AMP}]}{[\text{ADP}]^2}$). The latter equation can be reworked into the following equality: $(\frac{[\text{AMP}]}{[\text{ATP}]} \approx (\frac{[\text{ADP}]}{[\text{ATP}]})^2$) and thus, the AMP:ATP ratio changes roughly as the square of ADP:ATP ratio, making it a sensitive indicator of changes in ATP levels (Hardie *et al.*, 1994; 1998). If cells suffer conditions which interfere with ATP production, the ADP:ATP increase leads to an even larger increase in AMP:ATP ultimately activating the AMPK (Hardie *et al.*, 1994; 1998). Hardie and co-workers have hypothesized that the pathway acts as a cellular fuel gauge that enacts ATP conserving steps when cellular stresses deplete ATP. The active AMPK protects the cell by

inactivating rate-limiting enzymes in ATP-consuming anabolic pathways such as fatty acid synthesis (acetyl-CoA carboxylase) and sterol synthesis (HMG-CoA reductase) (Corton *et al.*, 1994; Henin *et al.*, 1995). At the same time, AMPK may have a role in activating alternative pathways for ATP generation (Hardie and Carling, 1997; Velasco *et al.*, 1997; Makinde *et al.*, 1997; Kemp *et al.*, 1999).

1.4.5 The physiological role of mammalian AMPKs

Studies on the possible physiological role of the AMPK have been prompted by observations of the *in vitro* phosphorylation of potential target proteins. Nevertheless, a protein that is phosphorylated *in vitro* may not be a substrate for a protein kinase *in vivo*; similarly, not every site phosphorylated on a protein will necessarily affect its activity. Therefore, as studies detailing possible new substrates and regulatory roles for the AMPK continue to surface, it has become of crucial importance to assess the true *in vivo* relevance of the AMPK in various metabolic pathways. The *in vitro* phosphorylation and inactivation of the two best known AMPK targets, acetyl-CoA carboxylase and HMG-CoA reductase, was well established through the early years of AMPK biochemical characterization (reviewed by Hardie, 1992). The *in vivo* role of the AMPK in regulating fatty acid and sterol synthesis has been established through several lines of evidence: (1) treatments which deplete cellular ATP levels [heat shock (Corton *et al.*, 1994), high fructose (Moore *et al.*, 1991), 2-deoxyglucose (Sato *et al.*, 1992), uncouplers of oxidative phosphorylation such as antimycin A, azide, or dinitrophenol (Witters *et al.*, 1991), and the TCA cycle inhibitor arsenite (Corton *et al.*, 1994)] have been shown to lead to increased AMPK levels, increased phosphorylation and inactivation of ACC and/or HMGR, and inhibition of fatty acid and/or sterol synthesis in cultured cells; (2) Epinephrine treatment leads to AMPK activation and increased ACC phosphorylation (Haystead *et al.*, 1990); (3) addition of insulin, a known activator of ACC and fatty

acid synthesis, to Fao hepatoma cells leads to inhibition of AMPK activity prior to the onset of ACC activation (Witters *et al.*, 1992); (4) incubation with AICAR (5-amino-4-imidazole-carboxamide ribonucleoside), a cell-permeable compound that is metabolized into the AMP analogue ZMP (5-amino-4-imidazole-carboxamide ribotide), activates AMPK and leads to phosphorylation and inhibition of ACC in primary rat adipocytes (Sullivan *et al.*, 1994; Corton *et al.*, 1995) and HMGR in rat hepatocytes (Corton *et al.*, 1995; Henin *et al.*, 1995).

The inactivation of ACC counteracts decreases in cellular ATP levels through a secondary effect distinct from that of preventing fatty acid biosynthesis. The product of ACC is malonyl-CoA, which inhibits β -oxidation of long-chain fatty acids through its potent allosteric inhibition of carnitine palmitoyltransferase I, the enzyme responsible for fatty acid uptake into the mitochondrial matrix (McGarry *et al.*, 1978; 1983). Decreases in malonyl-CoA levels due to inactivation of ACC can ultimately lead to a concomitant stimulation of fatty acid oxidation and ATP generation. In muscle tissue, which does not perform *de novo* fatty acid synthesis, the primary role of ACC inactivation could be activation of fatty acid oxidation (Hardie *et al.*, 1997). High rates of fatty acid oxidation are observed in ischemic hearts following reperfusion, and Kudo *et al.* (1995; 1996) have shown that these are the result of depressed malonyl-CoA levels resulting from the AMPK-dependent inactivation of ACC following a rise in AMP:ATP ratio. Similarly, Makinde *et al.* (1997) have shown that the increased fatty acid oxidation observed in the heart tissue of newborn rabbits is associated with increases in AMPK activity. Decreases in circulating insulin levels are observed after birth and lead to AMPK activation and inactivation of ACC, ultimately resulting in reduced malonyl-CoA levels and stimulation of fatty acid oxidation (Makinde *et al.*, 1997). A similar chain of events is observed in quadriceps muscle isolated from rats after prolonged running on a treadmill (Winder and Hardie, 1996). *In situ* electrical stimulation of rat skeletal muscle tissue has also been shown to lead to increases in the AMP:ATP ratio that are concomitant with increased AMPK levels

(Hutber *et al.*, 1997). Cells undergoing increased β -oxidation of fatty acids in response to decreased ATP levels are likely to decrease incorporation of fatty acids into complex lipids such as triacylglycerol. Consistent with this prediction, incubation of hepatocyte and myocyte cultures with AICAR leads to decreased [^{14}C]oleate and [^3H]glycerol incorporation into diacylglycerol, triacylglycerol, and phospholipid (Muoio *et al.*, 1999). It also leads to an AMPK-dependent inactivation of the mitochondrial isoform of *sn*-glycerol-3-phosphate acyltransferase (GPAT), the enzyme that catalyzes the first committed step in glycerolipid biosynthesis. Muoio *et al.* (1999) have proposed that by inhibiting the mitochondrial form of GPAT and alleviating the malonyl-CoA inhibition of carnitine palmitoyltransferase I, AMPK activation channels fatty acids towards mitochondrial β -oxidation and away from glycerolipid biosynthesis.

AMPK has recently been reported to phosphorylate and inhibit creatine kinase (CK) *in vitro* (Ponticos *et al.*, 1998). CK catalyses the conversion of creatine (Cr) into phosphocreatine (PCr) in the following reaction: $\text{Cr} + \text{ATP} \leftrightarrow \text{PCr} + \text{ADP}$. In skeletal muscle, PCr represents a temporary energy reserve that serves as a buffer providing immediate fuel to meet the rapid burst in ATP demand during muscle contraction (Walliman, 1994). In addition to showing that AICAR treatment of differentiated muscle cells leads to a reduction in CK activity, Ponticos *et al.* (1998) have shown that the AMPK is itself inhibited by PCr, but that this inhibition is relieved by Cr. In resting muscle, PCr:Cr and ATP:AMP are high and the AMPK is kept in an inactive form. During muscle contraction, the PCr:Cr ratio drops and the AMPK is activated. The subsequent inhibition of CK is thought to prevent a further depletion of cellular ATP levels in “charging” Cr into PCr, at least in the initial stages of muscle contraction. Phosphorylation and regulation of CK still remains to be established *in vivo*.

A fourth known physiological target for the AMPK is the hormone-sensitive lipase (HSL), an enzyme that catalyzes the hydrolysis of triacylglycerols and cholesterol esters (Garton *et al.*,

1988; Yeaman, 1990). The HSL is phosphorylated and activated by the cAMP-dependent protein kinase (PKA) (Garton *et al.*, 1988). *In vitro*, the AMPK phosphorylates the HSL at Ser-565 but this does not directly affect its activity. It does, however, prevent the phosphorylation by the PKA (at Ser-563) and thus the AMPK has been hypothesized to play an anti-lipolytic role (Garton *et al.*, 1989). Consistent with this, incubation of isolated hepatocytes with AICAR not only activates the AMPK but it prevents activation of lipolysis in response to agents known to raise intracellular cAMP levels (Sullivan *et al.*, 1994; Corton *et al.*, 1995). Hardie and Carling (1997) have argued that the antilipolytic role of the AMPK represents a mechanism through which the rate of hydrolysis of triglycerides and cholesterol esters does not outstrip the capacity of the cell to utilize the free fatty acids. The release of excessive free fatty acids can lead to an ATP-consuming futile cycle in which free fatty acids are recycled into glycerolipids and cholesterol esters through re-esterification.

AMPK has been shown to phosphorylate two further proteins *in vitro*: glycogen synthase (Carling and Hardie, 1989) and the Raf-1 kinase (Sprenkle *et al.*, 1997; Hardie *et al.*, 1998). Glycogen synthase is the key regulatory enzyme in glycogen synthesis whereas Raf-1 forms part of the mitogen-activated protein (MAP) kinase pathway. Hardie *et al.* (1998) have argued that both of these activities are consistent with the role of AMPK during stress [i.e. if energy status becomes compromised, AMPK could inhibit glycogen synthesis (which consumes UTP) and cell proliferation (which is energetically expensive)]. So far there is no evidence to suggest that either of these two proteins are phosphorylated and regulated by the AMPK *in vivo*, although studies are currently addressing this issue. It is important to recognize that the AMPK need not be the only regulatory mechanism acting on any individual enzyme. For example, in addition to AMPK-dependent regulation HMGR is also regulated at the transcriptional and translational levels (Goldstein and Brown, 1990) and by turnover rates (Correll and Edwards, 1994).

1.4.6 Similarities between SNF1 and AMPK protein kinases.

In addition to reporting on the high level of similarity with yeast Snf1 and rye Rkin1 primary sequences, Mitchelhill *et al.* (1994) showed that SNF1 and AMPK have similar substrate specificities because SNF1 can phosphorylate the SAMS peptide and, more importantly, because both the porcine AMPK and yeast SNF1 inactivate purified yeast ACC *in vitro*. In a subsequent paper, Carling and co-workers studied possible functional homology between SNF1 and AMPK by testing whether, like its mammalian counterpart, SNF1 could regulate ACC activity *in vivo*. By using the SAMS peptide phosphorylation assay, they found that SNF1 activity increases in a time-dependent manner in response to low glucose conditions and that ACC activity changes inversely with SNF1 activity (Woods *et al.*, 1994). The decreased ACC activity in low glucose could not be detected in *snf1Δ* mutants, indicating that SNF1 is responsible for regulating ACC activity *in vivo*. Taken together, the results of Mitchelhill *et al.* and Woods *et al.* suggested a novel function for SNF1: the regulation of *de novo* fatty acid synthesis, at least during glucose derepressing conditions. Conversely, as speculated by the authors of both studies, the finding also raised the possibility that AMPK may be involved in regulating cellular processes that are known to be regulated by SNF1 in yeast cells. For example, Salt *et al.* (1998) have observed that AMPK- α 2 can localize to the nucleus raising the possibility that, like the yeast SNF1 complex, AMPK- α 2 complexes may be involved in regulating gene expression (Salt *et al.*, 1998). AICAR, used to increase AMPK activity, has been recently shown to inhibit glucose-dependent expression of fatty acid synthase in cultured rat hepatocytes through a transcriptional mechanism (Foretz *et al.*, 1998).

Cloning and characterization of the AMPK non-catalytic subunits have revealed them to be homologous to the SNF1 non-catalytic subunits (Gao *et al.*, 1996): Sip1/Sip2/Gal83 are homologous to the β -subunit and Snf4 is homologous to the γ -subunit. Unlike AMPK, where full activation of AMPK requires all three subunits to come together in a ternary complex (Dyck *et al.*,

1996), Snf1 kinase activation only requires Snf4 (Celenza *et al.*, 1989; Celenza and Carlson, 1989; Woods *et al.*, 1994). Nonetheless, Sip1/Sip2/Gal83 help stabilize the Snf1/Snf4 interaction (Yang *et al.*, 1994). In high-glucose conditions Snf1 and Snf4 do not interact directly with each other yet remain associated in high molecular mass complexes. Sip1/Sip2/Gal83 interact with both Snf1 and Snf4 (Jiang and Carlson, 1996; 1997) and this mirrors the interactions observed between the AMPK subunits: although the α - and γ -subunits do not interact with each other, the β -subunit is able to form a stable interaction with each of them (Woods *et al.*, 1996).

Interestingly, despite the obvious conservation in function between the SNF1 complex and the AMPK complex, some differences exist (Hardie *et al.*, 1998). Unlike some plant SNF1 homologues, mammalian AMPKs have failed to rescue SNF1 function in *snf1* Δ mutants (Carling *et al.*, 1994). Also, despite many attempts to directly implicate it in the activation of the SNF1 complex, AMP does not activate SNF1 *in vitro* (Wilson *et al.*, 1996). Neither other nucleotides nor numerous sugar phosphates tested have an allosteric effect on SNF1 (Wilson *et al.*, 1996) and thus the mechanism through which changes in glucose levels trigger SNF1 activation is still unknown (Hardie *et al.*, 1998). Wilson *et al.* (1996) have demonstrated that a significant correlation exists between glucose levels, the yeast intracellular AMP:ATP ratio, and SNF1 activity. Specifically, although SNF1 is not allosterically affected by AMP, conditions which activate the SNF1 complex are accompanied by increases in the *in vivo* AMP:ATP ratio. Interestingly, Wilson *et al.* (1996) have also shown that, like the AMPK, SNF1 is inactivated *in vitro* by protein phosphatase 2A treatment and can be reactivated by using a partially purified preparation of the AMPK-kinase (AMPKK) (Woods *et al.*, 1994; Wilson *et al.*, 1996). They have also found an endogenous activity similar to the AMPKK in yeast extracts.

Unlike yeast cells, the majority of mammalian cells do not encounter extreme fluctuations in extracellular glucose levels owing to both the many homeostatic mechanisms which ensure stable

plasma glucose levels and to the glycogen stores present in many cell types. The AMPK should not be expected to be affected by changes in extracellular glucose because most mammalian cells do not respond to changes in glucose supply. The pancreatic β -cells that are responsible for insulin secretion are an exception to this rule and are thus extremely sensitive to changes in extracellular glucose levels. In these cells the AMPK, like SNF1, has been shown to respond to changes in extracellular glucose (Salt *et al.*, 1998). In pancreatic β -cells, changes in extracellular glucose levels lead to concomitant changes in the AMP:ATP ratio and to changes in the phosphorylation state and activity of AMPK (Salt *et al.*, 1998).

1.4.7 Phylogenetic distribution of the SNF1/AMPK protein kinases.

The SNF1/AMPK family of protein kinases is widely distributed among the different eukaryotic lineages, pointing to an ancient origin for both the family and the cascade they mediate. Genomic clones, cDNAs, and PCR products sharing high levels of sequence identity with yeast *SNF1* have been obtained from a number of different mammalian species (reviewed by Hardie *et al.*, 1998), plants (reviewed by Halford, 1994; Halford and Hardie, 1998), *Xenopus laevis* (Rhogi *et al.*, unpublished) and *C. elegans* (Wilson, R., unpublished). Genes representing the mammalian AMP-activated protein kinases have been cloned from several species including human, rat, and pig (Beri *et al.*, 1994; Carling *et al.*, 1994; Woods *et al.*, 1994; Gao *et al.*, 1995). Many different plant *SNF1*-like protein kinases, the SnRK1 subfamily (for Snf1-Related protein Kinases), have been cloned and these include rye *RKIN1* (Alderson *et al.*, 1991), *Arabidopsis thaliana* *AKIN10* (LeGuen *et al.*, 1992), barley *BKIN12* (Halford *et al.*, 1992), tobacco *NPK5* (Muranaka *et al.*, 1994), potato *PKIN1* (Man *et al.*, 1997), and sugar beet *SBKIN154* (Monger *et al.*, 1997), among others. In addition, kinase activities that are able to phosphorylate the SAMS peptide and/or mammalian HMG-CoA reductase have been partially purified from cauliflower, barley, and potato (Ball *et al.*,

1994; Barker *et al.*, 1996; MacKintosh *et al.*, 1992). Although not activated by AMP, these kinase activities share other biochemical characteristics with AMPK including inactivation by phosphatase 2A and reactivation by the AMPKK (MacKintosh *et al.*, 1992; Ball *et al.*, 1994; Dale *et al.*, 1995). The plant kinases encoded by the rye *RKIN1* and tobacco *NPK5* genes have been shown to complement yeast *snf1Δ* mutants (Alderson *et al.*, 1991; Muranaka *et al.*, 1994) and appear to represent functional *SNF1* homologues. In fact *NPK5* has been shown to interact with Snf4 in two-hybrid assays (Jiang and Carlson, 1996). Although the other plant kinases mentioned have not been well characterized, they are thought to represent *RKIN1* and *NPK5* homologues based on sequence similarity and clustering on phylogenetic trees (Halford and Hardie, 1998).

The mammalian and plant kinases mentioned above represent the closest relatives of yeast *SNF1*. A number of protein kinases showing slightly lower levels of similarity to *SNF1* have also been described although none have been characterized in any detail. These include kinases from the plant SnRK2 and SnRK3 families (Halford and Hardie, 1998), and kinases such as the human *KP78* (Maheshwari *et al.*, unpublished), human *SNRK* (Becker *et al.*, 1996), mouse *EMK* (Inglis *et al.*, 1993), mouse *MSK* (Ruiz *et al.*, 1994), and *PARI* from *C. elegans* (Guo and Kemphues, 1995). Although to date no phylogenetic analysis has been carried out which includes all plant and animal SNF1/AMPK kinases and their close relatives, different subsets of SNF1/AMPK-related kinases have been analyzed (Hanks and Hunter, 1995, Becker *et al.*, 1996; Halford and Hardie, 1998). The SNF1/AMPK kinases and close relatives form a separate branch within the larger group of animal calmodulin-dependent protein kinases and plant calmodulin domain protein kinases defined by Hanks and Hunter (1995). Yeast *SNF1* and its closest relatives [the mammalian *AMPKs*, *C. elegans* *PAR2.3*, and the SnRK1 group of plant kinases (which includes rye *RKIN1* and tobacco *NPK5*)] all cluster together as a single group in phylogenetic trees (Becker *et al.*, 1996). The plant SnRK2 and SnRK3 groups are closely related to the plant SnRK1 group, though less so than *SNF1* and *AMPK*

(Halford and Hardie, 1998). The group of animal kinases that includes *KP78*, *SNRK*, *EMK*, *MSK*, and *PAR1* also shares some sequence conservation with the SNF1/AMPK kinases (Becker *et al.*, 1996; Halford and Hardie, 1998). Little is known about these last three groups of kinases.

Based on similarities between AMPK, SNF1, and their closest relatives, Hardie *et al.* (1998) have proposed a unifying hypothesis for the role of this family of protein kinases. They propose that the kinase cascades mediated by the SNF1/AMPK kinases are involved in detecting changes in cellular energy status and enacting changes in metabolism, gene expression, and perhaps cell proliferation in response to stresses that compromise the energy status of the cell.

1.5 The glucose repression system in *Drosophila*

In addition to temporal, spatial, and developmental regulation, the α -amylase gene of *Drosophila melanogaster* is repressed by dietary glucose (Benkel and Hickey, 1986a,b). We have been studying the *Drosophila* α -amylase gene-enzyme system as a model for gene regulation because it provides a rare example of glucose repression in a higher eukaryote (Hickey *et al.*, 1989). *Drosophila* produces α -amylase in order to utilize non-preferred carbohydrate sources (Haj-Ahmad and Hickey, 1982). Alpha-amylase is a digestive enzyme that catalyzes the hydrolysis of the internal glycosidic bonds of α -(1,4)-linked glucose polymers such as starch or amylose. The *Drosophila Amy* locus contains two closely linked genes (Gemmill *et al.*, 1986; Boer and Hickey, 1986; Benkel *et al.*, 1987) that are divergently transcribed. A number of different isozymes exist for each gene copy and many strains produce two different α -amylase variants. Amylase activity is restricted to the larval and adult midgut and discrete patterns of activity are observed during the different developmental stages of *Drosophila* (Doane *et al.*, 1983; Thompson *et al.*, 1992). A *trans*-acting regulatory locus (*mapP*) has been shown to exert tissue-specific control of amylase activity, positively regulating amylase levels in the posterior region of the midgut (Doane *et al.*, 1983; Thompson *et al.*, 1992). Dietary glucose has been shown to depress the levels of α -amylase activity in both larvae and adult flies (Hickey and Benkel, 1982) and most naturally occurring α -amylase variants show glucose repression (Benkel and Hickey, 1986a; b). Repression occurs at the transcriptional level and results in large decreases in α -amylase mRNA levels, along with concomitant decreases in protein levels, following exposure to glucose (Benkel and Hickey, 1987; Hickey *et al.*, 1989). It has been shown previously that elements within the α -amylase promoter start site are responsible for mediating both overall transcriptional activity as well as glucose repression (Benkel and Hickey, 1987; Magoulas *et al.*, 1993). Deletion constructs with as little as

109 bp of the sequence immediately upstream of the α -amylase coding sequence show both an appropriate level of expression as well as an appropriate response to dietary glucose (Magoulas *et al.*, 1993). Attempts to define the minimal *cis*-acting elements responsible for glucose regulation of expression within the *Drosophila* α -amylase promoter have shown that the mechanism of glucose repression likely involves several functionally redundant elements (Magoulas *et al.*, 1992; Yoshida, 1998).

Previous work in our lab has provided evidence that the mechanism of glucose repression acting on the *Drosophila* α -amylase gene is similar to that found in *S. cerevisiae* (Hickey *et al.*, 1994). Yeast cells transformed with shuttle vectors bearing the wild-type *D. melanogaster* α -amylase gene, including both coding and non-coding regions, produced and secreted enzymatically active fly amylase in a glucose repressible fashion (Hickey *et al.*, 1994). Constructs in which the promoter region of *Drosophila* α -amylase gene was fused to the firefly luciferase coding region also expressed this reporter gene in a glucose responsive manner and thus the fly α -amylase promoter is fully glucose repressible when introduced into yeast cells. That this *Drosophila* promoter is recognized by the transcriptional machinery involved in effecting glucose repression in yeast is highly suggestive of evolutionary conservation between the *Drosophila* and *S. cerevisiae* glucose repression pathways (Hickey *et al.*, 1994).

1.6 Thesis project : description and goals.

Molecular evidence mentioned in the previous section suggests evolutionary conservation between the yeast and *Drosophila* glucose repression pathways. The working hypothesis underlying this thesis is that the glucose repression pathways of *D. melanogaster* and *S. cerevisiae* evolved from the same original cascade and that, as such, important components of the yeast glucose repression pathway have been conserved in *Drosophila*. In yeast, the SNF1 protein kinase complex mediates derepression. SNF1 forms a part of a larger family of kinases involved in regulating metabolism in response to nutritional stress and we reasoned that similar protein kinases would also exist in *Drosophila*.

The family of SNF1 related protein kinases have retained a conservation in function across widely divergent taxa and this predicts that putative *Drosophila* SNF1 homologues will also have similar roles. In Section 2, I describe the methodology used in both the search for putative *Drosophila* SNF1 homologues and in exploring the molecular evolution of the SNF1/AMPK protein kinase family. Section 3 is devoted to describing the results obtained in the various experiments and analyses.

In the first series of experiments (Section 3.1), human and *Drosophila* protein kinases were identified by using the polymerase chain reaction (PCR) along with degenerate oligonucleotides based on conserved motifs of the SNF1/AMPK protein kinases. In the second set of experiments (Section 3.2), I describe the subsequent cloning and characterization of the novel protein kinases and also describe results obtained from the analysis of the molecular data. Lastly, in Section 3.3, I describe the sequences and phylogenetic analysis of the SNF1/AMPK protein kinase family.

2. MATERIALS AND METHODS

2.1 Degenerate PCR primer design.

2.1.1 SNF1/AMPK biased primers.

Degenerate oligonucleotides were designed based on general conserved features of the eukaryotic protein kinase catalytic domain and were biased towards the amplification of SNF1- and AMPK-like sequences by taking into account residues conserved between yeast SNF1 and rat AMPK- α 2 catalytic domain. The following three peptide sequences were selected for oligonucleotide design (See Fig 1.A for location along the catalytic domain): as sense primers: a) K929: G(S/T)FGKVK and b) K932: VHRDLKPEN, and as the antisense primer: c) K930: Y(A/V)GPEVD(I/L/V)W. Degenerate primers corresponding to these peptides are as follows:

K929 5'-GGI-ICI-TT(TC)-GGI-AA(GA)-GT(IG)-AA-3'

K930 5'-CCA-IAI-(AG)TC-IAC-(TC)TC-IGG-ICC-I(AG)C-(AG)TA-3'

K932 5'-GTI-CA(CT)-(AC)GI-GA(CT)-(TC)TI-AA(GA)-CCN-GA(GA)-AA-3'.

2.1.2. Primers biased towards SNF1 and its plant homologues

In order to bias amplification of kinases sharing higher similarity with SNF1 and functional plant homologues, sets of degenerate primers were also designed based on sequences shared between Snf1, rye Rkin1 (Alderson *et al.*, 1991) and tobacco Npk5 (Muranaka *et al.*, 1994). Several primer pairs were initially designed although only one primer pair yielded consistent results. The

sense/antisense primer pair K1184 and K1187 was based on peptide sequences T(R/G)(H/Q)KVA and HPHII(R/K) respectively. The fully degenerate primers corresponding to those peptides are as follows:

K1184: 5'-(CT)(TG)(TGA)-AT(TGA)-AT(GA)-TG(ATGC)-GG(GA)-TG-3'

K1187: 5'-ACN-(ACG)G(ATGC)-CA(ATGC)-AA(AG)-GT(ATGC)-GC-3'.

The K1184/K1187 primer pair was successfully used to amplify sequences from *Drosophila* genomic DNA and from yeast genomic DNA. Additionally, the sense primer K929 was also successfully used with K1187. All primers were synthesized in an ABI381A oligonucleotide synthesizer according to manufacturer instructions. Inosine (I) residues were incorporated at some sites to minimize degeneracy.

2.2 PCR cloning of SNF1-like sequences.

Products were amplified using both K929-K930 and K932-K930 primer combinations. 100 pmol of each primer were used along with 100 to 400 ng of *Drosophila melanogaster* (Ore-R) genomic DNA template in 50 µl reaction volumes [1 to 2.5 mM MgCl₂; 50 mM KCl; 10 mM Tris-HCl (pH 9.0); 0.01% Triton X-100; 100 pmol of each primer; 0.2 U Taq polymerase]. Amplification was performed for a total of 35 cycles with a 30 s denaturation at 95°C, 45 s of annealing (as described below) and a 45 s extension at 72°C. Annealing temperatures were based on a 10 cycle touchdown-PCR profile (initial annealing temperature of 50°C, and a 1°C decrease in each subsequent cycle down to a final annealing temperature of 40°C), followed by 25 cycles with

annealing at 40°C and a final extension of 6 minutes. Amplification products were visualized by ethidium bromide staining after agarose gel electrophoresis and processed according to the scheme shown in Appendix I. Products of the expected sizes were excised and purified using the Qiaex II gel purification kit (Qiagen), followed by cloning into the pCRTMII vector (TA cloning kit, Invitrogen). Individual clones were picked and insert sizes were ascertained directly by PCR on colonies using vector primers flanking the cloning site. Plasmid DNA was prepared from positive clones selected at random (plasmid purification kit, Qiagen), and sequenced using dye-deoxide chemistry (Applied Biosystems) in an ABI373A automated sequencer. Sequences obtained were then searched against the GenBank database.

The K930-K932 and K929-K932 primer combinations were also used for PCR using a number of different templates including: human intestinal, skeletal muscle and liver cDNA; mouse genomic DNA; yeast genomic DNA. Human cDNAs were synthesized by reverse transcription of 1 µg of poly A⁻ RNA (Clontech) using 200 U of Superscript II (Gibco BRL) using the following conditions: 500 µM of each dNTP; 50 mM Tris-HCl (pH 8.3); 75 mM KCl; 3 mM MgCl₂; 10 mM DTT; 50 µg/ml oligo (dT)₁₂₋₁₈; 50 U/ml RNasin. Reaction times are as follows: 10 min denaturation at 70°C, annealing and extension at 42°C for 55 min, heat inactivation at 70°C for 15 min, followed by a 20 min treatment with RNaseH. Approximately 50 ng of cDNA were used per PCR in a 50 µl reaction as described above. Yeast genomic DNA (strain AH22) was prepared from spheroplasts (cells were harvested from liquid culture by centrifugation and were incubated for 1 hour at 37°C in zymolyase solution [0.9 M Sorbitol, 0.1 M Tris-HCl (pH 8.0), 0.1 M EDTA (pH 8.0), 0.03 g/ml zymolyase, 30 mM β-mercaptoethanol]). Spheroplasts were lysed in 50 mM Tris (pH 8.0), 20 mM EDTA (pH 8.0), 1% SDS. The lysate was then applied to an anion exchange chromatography column (Genomic tip, Qiagen) and DNA was purified according to the manufacturer's instructions.

Mouse genomic DNA was courtesy of Dr. B. Benkel. 100 ng of yeast genomic DNA and 500 ng of mouse genomic DNA were used per PCR reaction as described above.

2.3 Probe construction and *Drosophila* library screening.

Approximately 2.5×10^5 pfu of a *Drosophila* λ -Fix II genomic library (Stratagene) were screened using each of the two distinct 192 bp insert bearing clones that were obtained using the K932/K930 primer pair. Phages were plated at a density of 5×10^4 pfu/plate and plaques were double-lifted onto Hybond-N membranes (Amersham) according to manufacturer's instructions. Following 4 min denaturation in 0.2 M NaOH/1.5 M NaCl and 2 min neutralization in 3 M sodium acetate, membranes were air dried and DNA was UV cross-linked.

Probe templates were prepared by excising the 192 bp inserts from pCRTMII using *Eco*RI followed by gel purification. Approximately 50 ng were then labelled to high specific activity with [α -³²P]dCTP (3000 mCi/mmol, Amersham) by random priming (Oligolabelling Kit, Pharmacia) and unincorporated label was removed by filtration in Sephadex G-50 columns (Boehringer Mannheim). Hybridizations were carried out in 50% Formamide, 10 X Denhardt's solution [100 X Denhardt's: 2% Ficoll 400, 2% polyvinylpyrrolidone, 2% bovine serum albumin (Fraction V)], 1M NaCl, 50 mM Tris-HCl (pH 8.0), and 100 mg/ml of sheared salmon sperm DNA at 42 °C for 12-16 h. The membranes were then briefly washed at room temperature in 2 X SSC/0.1% SDS (2 X 5 min), followed by two 20 min room temperature washes in 0.2 X SSC/0.1% SDS [20 X SSC: 3 M NaCl, 0.3 M sodium citrate, pH 7.0]. Autoradiography was then performed for 1-2 days at -80°C. Positive plaques were identified, resuspended in SM buffer [100 mM NaCl, 8 mM MgSO₄, 50 mM Tris-HCl

(pH 7.5), 0.01% gelatin] and re-plated at lower densities for each subsequent round of re-screening.

Single well-isolated positive plaques were isolated and resuspended in SM buffer. Resuspended plaques were used to prepare confluent plates that were then used to obtain a high titer supernatant. The latter was used to infect liquid cultures at low multiplicity. Phage DNA was then purified from the resulting liquid lysates (λ DNA purification kit, Qiagen).

2.4 Subcloning and DNA sequencing of genomic clones.

Southern hybridizations (Southern, 1975) were performed in order to determine subfragments of the λ clones hybridizing to the PCR probes. DM_SNF1A and DM_SNF1B λ genomic clones were restricted in single and double digests using *NotI* and *EcoRI*. The restriction digests were visualized by ethidium bromide staining following agarose gel electrophoresis. DNA was then transferred to Hybond-N membranes by alkaline capillary transfer as described by the manufacturer and UV crosslinked to the membrane. Probe preparation and hybridizations were carried out as described in section 2.3.

The *NotI*, *NotI/EcoRI* and *EcoRI* subfragments hybridizing to the probes were gel-purified and then cloned into the pBluescript (SK+). Resulting plasmid clones were sequenced using dye-deoxy chemistry (Applied Biosystems) in an ABI373A sequencer through a combination of subcloning and primer walking. A 6.5 Kb *NotI/EcoRI* subclone bearing two *DM_SNF1B* gene copies was completely sequenced by using a combination of unidirectional nested Exonuclease III deletions (Erase-a-base system, Promega) [See Fig 5] and primer walking. Sets of nested deletion clones were generated from both ends of insert in order to generate sequence from both strands.

2.5 Northern and genomic Southern analysis.

Total whole-body RNA was isolated from 0.5-1.0 g of Ore-R adult flies or larvae grown in either glucose supplemented or glucose deficient medium by the method of Chomczynski and Sacchi (1987). Flies or larvae (0.5 to 1.0 g) were homogenized in 5 ml of denaturing solution [4 M guanidium thiocyanate, 25 mM sodium citrate (pH 7.0), 100 mM β -mercaptoethanol, 0.5% sarcosyl]. To the resulting homogenate was added 0.5 ml of 2 M sodium acetate (pH 4.0) and 5 ml of water-saturated phenol, followed by thorough mixing. One ml of 49:1 chloroform/isoamyl alcohol was then added to the mixture, followed by a 15 min incubation on ice. Following a 20 min centrifugation (10000 x g) at 4°C, the aqueous phase was removed and an equal volume of 100% isopropanol was added and the mixture was incubated for 30 min at -20°C to precipitate the RNA. Following a 10 min centrifugation (10000 x g), the RNA pellet was dissolved in 1.5 ml of denaturing solution and the RNA was re-precipitated with 1 volume of 100% isopropanol for 30 min at -20°C. Following a 20 min centrifugation (10000 x g) at 4°C, the RNA pellet was resuspended in 75% ethanol with vortexing at room temperature to remove residual traces of guanidium. The mixture was then centrifuged for 5 minutes at 10000 x g. The RNA pellet was dried and re-suspended in DEPC-treated water. RNA yield and quality were ascertained by gel fractionation in denaturing gels [1.0 % agarose, 2 M formaldehyde] and by spectrophotometric quantitation. For Northern blotting experiments, 10 μ g of each total RNA sample were run on a denaturing gel followed by capillary transfer in 20 X SSC to a Hybond-N membrane (Amersham) according to the manufacturer's instructions. Poly-A⁺ RNA was also isolated from the total RNAs obtained above using poly-A⁺ isolation kits (Qiagen and Invitrogen) according to the manufacturer's instructions. Approximately 1 μ g of each poly-A⁺ RNA were then run on denaturing gels and transferred as described above.

Genomic DNA was isolated by homogenizing 1-2 g of Ore-R adult flies in lysis buffer [100 mM Tris-HCl (pH 8.0), 50 mM NaCl, 50 mM EDTA, 1% SDS, 0.15 mM spermine, 0.5 mM spermidine] on ice, followed by proteinase K digestion (100 µg/ml, 40°C, overnight incubation) with simultaneous RNase A (300 µg/ml) treatment. Supernatants were cleared by centrifugation and DNA was purified by anion exchange chromatography (Genomic-tip 500, Qiagen). Yield and quality were ascertained by gel fractionation and by spectrophotometric quantitation. For Southern blots, 5 µg of genomic DNA was used per restriction digest, run on a 0.9% agarose gel, followed by alkaline capillary transfer to a Hybond-N membrane as described by the manufacturer's instructions. DNA probes were labelled by random priming and hybridization was carried out in 5 X Denhardt's, 50% formamide, 0.2 M Tris-HCl (pH 7.5), 0.86 M NaCl and 0.1 mg/ml salmon sperm DNA at 42 °C for 12-16 h. Membranes were briefly washed in 2 X SSC/0.1% SDS at room temperature (2 X 5 min), followed by two ten minute washes in 0.2 X SSC/0.1% SDS at 65°C. Some hybridizations were carried out under more relaxed stringency conditions, Medium stringency hybridizations were carried out exactly as described above but membranes were briefly washed at room temperature in 2 X SSC/0.1% SDS (2 X 5 min), followed two 10 min washes in 0.2 X SSC/0.1% SDS at 50°C followed by autoradiography.

2.6 Cloning of a cDNA clone of *DM SNFIA*.

A *Drosophila* adult λ -Dash II cDNA library (Stratagene) was screened in order to obtain either full length cDNA clones for *DM_SNFIA* or clones bearing the region that includes the translational start site and the exon 1-exon 2 junction. Screenings were carried out exactly as described for the genomic library. Once single plaques were isolated in tertiary screening,

phagemids were rescued by *in vivo* excision (Stratagene), inserts were sized and then characterized by PCR using a combination of vector-derived primers and primers designed based on previously obtained genomic sequence. Twelve positive clones were brought to third screening although none of the clones obtained included the putative start of the coding sequence as ascertained by PCR using vector primers flanking the cloning site. The cDNA was therefore obtained by a RT-PCR. Two μg of total larval RNA (isolated as described above) were reverse transcribed using 400 U of Superscript II (Gibco BRL) [500 μM of each dNTP; 50 mM Tris-HCl (pH 8.3); 75 mM KCl; 3 mM MgCl_2 ; 10 mM DTT; 50 $\mu\text{g/ml}$ oligo (dT)₁₂₋₁₈; 50 U/ml RNasin]. Reaction times are as follows: 10 min denaturation at 70°C, annealing and extension at 42°C for 55 min, heat inactivation at 70°C for 15 minutes, followed by a 20 min treatment with RNase H. One tenth of the resulting cDNA was then used for PCR using the forward primer (K1162: 5'-GGA GCT CCA TGG TCG AGA TGC CCC AGA TGA G – 3') and the reverse primer (K1163: 5'-CAC CTC AAG CTT AGC GAG CCA GTT GAA TG-3'). K1162 is expected to anneal in the region of the ATG start codon of *DM_SNF1A*. K1163 is expected to anneal in the region of the UGA stop codon of *DM_SNF1A* (ie. TCA represents the reverse complement of the UGA stop codon). The K1162-K1163 primer pair thus generates a PCR product that extends from the putative start codon to the putative stop codon of *DM_SNF1A*. The following PCR conditions were used: 2.0 mM MgCl_2 ; 50 mM KCl; 10 mM Tris-HCl (pH 9.0); 0.01% Triton X-100; 100 pmol of each primer; 0.2 U Taq polymerase. Cycling conditions were as follows: 30 cycles of 30 s denaturation at 94°C, 45 s annealing at 55°C, 150 s extension at 72°C, using a high fidelity Taq polymerase (TAKARA Biomedicals). The resulting ~ 1.7 Kb amplicon was cloned into pCRTMII (TA cloning kit, Invitrogen) and the identity of individual clones was then ascertained by dye-deoxy sequencing (Applied biosystems).

2.7 Sequence analysis and sequence comparisons.

DNA and deduced amino acid sequences were analyzed and manipulated using the Unix-based program GDE 2.2 (Smith *et al.*, 1994) and the Windows NT/95 program SequEdit 0.5 (Drouin *et al.*, 1999). Database searches were conducted on ENTREZ and BLAST WWW servers at ncbi.nlm.nih.gov. Sequence alignments were performed using CLUSTALW 1.74 (Thompson *et al.*, 1994) using the program defaults; minor adjustments were then made to resulting alignments. Sequence shading was performed using BOXSHADE 2.7 (Kay Hoffman, unpublished) and CLUSTALX 1.8 (Thompson *et al.*, 1997) [a Windows 95-based graphical interface which uses the CLUSTALW alignment algorithm]. Phylogenetic analysis was carried out with the PHYLIP 3.52c suite of programs (Felsenstein, 1993): trees were inferred using maximum parsimony and protein distance (Dayhoff model) methods using PROTPARS and PROTDIST, respectively. When using PROTPARS, sequences were jumbled 10 times in order to avoid bias resulting from input order of the sequences; global rearrangements were also allowed. The Neighbour-joining algorithm was used to derive trees from the distance matrix data generated by PROTDIST by using the program NEIGHBOR. In order to obtain statistical support for different branches, bootstrapping was used to resample the alignment 1000 times using SEQBOOT; the input order of sequences was randomized. In order to derive consensus trees from the multiple data sets generated by bootstrapping the program CONSENSE was used. In order to generate distance trees with bootstrap values assigned to individual branches, a Dayhoff distance matrix was generated from the original data set using PROTDIST, and a Neighbour-Joining tree was then inferred from the resulting matrix. Bootstrap values were then added to this tree when both consensus bootstrapped trees and distance trees had identical topologies. To avoid biases resulting from input order of sequences, ten separate distance

trees were generated by jumbling the sequences using different seed numbers and topologies were compared. Trees shown in the results section are unaffected by sequence input unless otherwise stated.

A third inference method was used in addition to distance and parsimony methods: PUZZLE 4.0.2 (Strimmer and von Haeseler, 1996 and 1997) was used to infer phylogenetic trees based on the quartet puzzling and maximum-likelihood algorithms. The parameters used included 10000 puzzling steps, exact parameter estimates, quartet sampling, neighbour-joining tree-fitting, the computation of non-clock like branch lengths and use of a mixed model of substitution rate heterogeneity which includes an invariable rate and eight gamma rates. The JTT (Jones et al., 1992) model of amino acid substitution was used. The percentage of puzzling steps supporting individual branches was then added to maximum-likelihood distance trees.

2.8 Molecular sequence accession numbers.

GenBank accession numbers for the sequences reported here are shown in Appendix I. Also included are accession numbers for all of the sequences used in the phylogenetic analyses.

3. RESULTS

3.1 PCR identification of novel kinase domains

Degenerate oligonucleotides were designed based on general conserved features of the kinase core catalytic domain that is common to all eukaryotic protein kinases. The consensus peptide sequences were chosen to reflect both consensus sequences common to all serine/threonine protein kinases and residues conserved only in the yeast SNF1 and human AMPK sequences (in order to favour the amplification of SNF1-like sequences) (Fig 1.A). Two sense and one antisense primers were initially designed (Fig 1.A and 1.B): the sense primer K929 (G-S/T-F-G-K-V-K) is based on the consensus motif GXGXXGXV found in Subdomain I of the majority of ePKs, including SNF1. The sense primer K932 (V-H-R-D-L-K-P-E-N) is based on the consensus sequence HRDLKXXN of subdomain VIb. The antisense primer K930 (Y-A/V-G-P-E-V-D-I/L/V-W) is based on the sequence around an invariant aspartic acid residue in subdomain IX (D220 in PKA-C α). Excluding possible introns, partially overlapping PCR products of 192 bp and 523 bp are expected from reactions with K930-K932 and K929-K932, respectively (Fig 1.A). Because each sense primer was used with the same antisense primer, the shorter PCR product is fully contained within the larger amplicon.

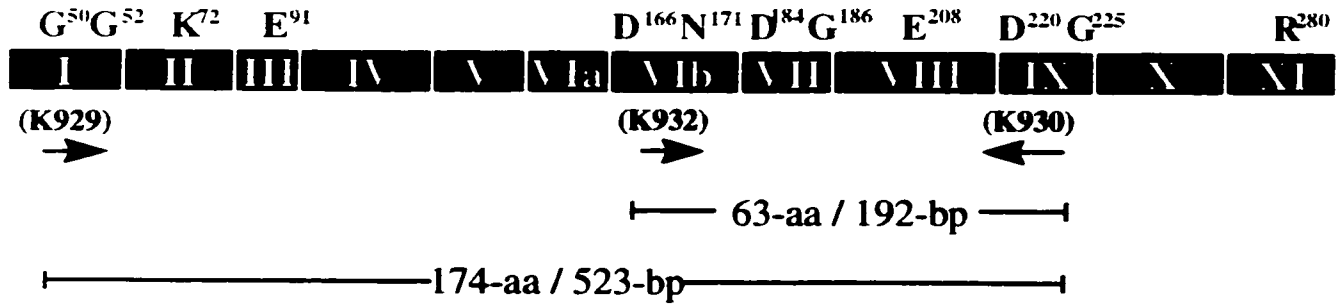
3.1.1 Identification of novel *Drosophila* SNF1-like kinase domains.

Although both primer pairs were used, only the K930-K932 primer pair generated appropriately sized PCR products from *Drosophila* genomic DNA (Fig 2A). These products were cloned and twenty-five clones were then selected at random for sequencing.

Figure 1. Identification of SNF1/AMPK-like sequences from *Drosophila melanogaster*.

(A) Position of the three different SNF1/AMPK-biased primers relative to the conserved ePK catalytic domain. The different subdomains are shown with roman numerals, invariant residues (red) are numbered with respect to human cAMP-activated protein kinase-C α subunit. (B) Design of degenerate PCR primers biased towards the amplification of SNF1- and AMPK-like sequences (the residues in red are invariant in the majority of ePKs; residues in blue are conserved between the protein kinases encoded by yeast *SNF1* and human *AMPK- α 2* genes but not in most other ePKs).

1.A



1.B

Forward Primer (K929)	G	S T	F	G	K	V	K		
Forward Primer (K932)	V	H	R	D	L	K	P	E	N
Reverse Primer (K930)	Y	A V	G	P	E	V	D	I L V	W

Sequencing results revealed that the cloned PCR product represents a mixture of two different amplified sequences, which we named DM_SNF1A and DM_SNF1B (Fig 2.B). BLAST searches performed on DM_SNF1A yielded as top-scoring hits the mammalian AMPK- α 2 subunit (alignment shown in Fig 2.B), yeast SNF1, and the putative mouse kinase *ink76*. Similar analysis performed on DM_SNF1B revealed it to be most similar to the putative human kinase p78 (KP78) (alignment shown in Fig 2.B), a relative of the mammalian AMPKs.

3.1.2 Identification of novel human SNF1-like kinase domains.

The mouse *ink76* kinase was previously identified as a small PCR product amplified from cDNA from the cells of the intestinal lining during an RT-PCR based survey of kinases expressed in the murine intestine (Siyanova *et al.*, 1994). A close inspection of the conceptual translation of the *ink76* sequence revealed differences at amino acid sites that are identical between the two mammalian AMPK- α sequences that had been cloned at the time [rat AMPK- α 2 (Carling *et al.*, 1994) and human AMPK- α 2 (Gao *et al.*, 1995)]. These differences raised the possibility that *ink76* represented a novel mammalian AMPK- α isoform. We thus tested the K930-K932 and K929-K932 primer pairs on cDNA from human intestinal tissue. Both primer pairs yielded products of appropriate sizes, ~190 and ~520 bp respectively (see Fig 3A), which were then cloned. Analysis of multiple clones revealed that each PCR product represented a single homogeneous amplified sequence. Comparison of the two types of cDNA sequences obtained also revealed that both products are identical within the region of overlap, suggesting a common type of template molecule. The longer K929-K932 RT-PCR product was compared to the sequences of rat and human α -subunits reported by Carling *et al.* (1994) and Gao *et al.* (1995).

Figure 2. PCR amplification of SNF1/AMPK-like sequences from *Drosophila melanogaster*.

(A) Primers K932-K930 (Fig 1.B) were used on *Drosophila* genomic DNA template. A PCR product of \approx 190 bp was obtained and cloned. (B) Sequence alignment of the conceptual translation of DM_SNF1A and DM_SNF1B PCR products to high scoring hits obtained on GenBank searches. HUM_AMPK: human AMPK- α 1 (Yano, K., unpublished; accession # BAA36547.1). HUM_KP78: human kinase p78 (Maheshwari *et al.*, unpublished; accession # S27966). Similar and identical residues are boxed in red.

2.A

226
123



~190 bp

2.B

```
DM_SNF1A 1 VHRDLKPENLLLDHNMHVKIADFGLSNMM DGEFLRTSCGSPNYAAPEV I SGKLYAGPEVDIW 63
VHRDLKPEN+LLD +M+ KIADFGLSNMM DGEFLRTSCGSPNYAAPEV I SG+LYAGPEVDIW
HUM_AMPK 36 VHRDLKPENVLLDAHMNAKIADFGLSNMMS DGEFLRTSCGSPNYAAPEV I SGRLYAGPEVDIW 198
```

```
DM_SNF1B 1 VHRDLKAENLLLDQH MNIKISDFGFENTSDPNA QLETFCGSPPYAAPELH MGRNYA GPEVDLAW 63
VHRDLKAENLLLD MNIKI+DFGF N +L+TFCGSPPYAAPELH G+ M GPEVDL W
HUM_KP78 175 VHRDLKAENLLLDAD MNIKIADDFGFSNEFTVGGKLDTECGSPPYAAPELH QGKMD GPEVDLW 237
```

Figure 3. PCR identification of a novel human AMPK isoform.

(A) Primers K929/K930 and K932/K930 (Fig 1.B) were used on human intestinal cDNA template. PCR products of ≈ 510 and ≈ 190 bp were obtained and cloned. (B) The nucleotide sequence of the novel human AMPK sequence (accession # HSU22456) in an alignment to the human AMPK- α 2 cDNA sequence (accession # U06454) to highlight differences at the DNA level. (C) Also shown is the conceptual translation of the novel sequence and alignment to the AMPK- α 2 amino acid sequence.

3.A



492
369
246
123

3.B

AMPK_A2
HSU22456 GCCACCTTCGGCAAAGTGAA GATTGAGAACATCAATTA CAGCCATAAAGTGGCAGTT AAAATCTTAAAFAGACAGAA GATTCGCAGTTTAGATGTTG
.GG.G.....G.G.....G.....CA.....G.....G.....G.....A..T..G ..G..AC.C...C.....G...CC.T.....G.....G.....

AMPK_A2
HSU22456 TTGGAATAAATAAACGAGAA ATTCAAATCTAAAATCTTT TCGTCATCCCTCATATTATCA AACTATACCAGGTGATCAGC ACTCCAACAGATTTTTTTAT
.A.....CCGCA.....G..C..C..G..T.. CA.G.....A..T.. ..G.....C.....T ..A...T..T...A.....C.....

AMPK_A2
HSU22456 GGTAAATGGAATATGTGTCTG GAGGTGAATATTGACTAC ATCTGTAAGCATGGACGGGT TGAAGAGATGGAAGCCAGGC GGCTCTTTCAGCAGATTTCTG
.G.....C..A.A..GC.....T..TA.....A..C..G..T..A.AA...AGTC... .T..G..C..A.....C..T

AMPK_A2
HSU22456 TCTGCTGTGGATTACTGTCA TAGGCATATGGTTGTCATC GAGACCTGAAACCAGAGAAAT GTCCCTGTTGGATGCACACAT GAATGCCAAGATAGCCGATT
.G.....T..... C.....G..C..ATT.....T..A... ..C..T.....A.....T.....

AMPK_A2
HSU22456 TCGGATTATCTAATATGATG TCAGATGGTGAATTTCTGAG AACTAGTTCGGATCTCCA ATTATACAGCACCTGAAGTC ATCTCAGGCAGATTTGATGC
.T..TC.T..A..C.....T..A... ..A.....T..C..A..C.. .C...G.T.....A.....A ..T.....A.....C..T

AMPK_A2
HSU22456 AGGTCCTGAAGTTGATATCT GGA
C...C...C.....C..CC.....

3.C

HSU22456 1 GAFGKVKVKGHELTGHKAVKILNRQKIRSLDVVGKIRREIQNLKLFRRPHIHKLYQVIS 180
G FGKVK+G+H+LTGHKAVKILNRQKIRSLDVVGKI+REIQNLKLFRRPHIHKLYQVIS

AMPK_A2 25 GTFGKVKIGEHLTGHKAVKILNRQKIRSLDVVGKIKREIQNLKLFRRPHIHKLYQVIS 84

HSU22456 181 TPSDIFVMVEYVSGGELFDYICKNGRLDEKESRRLFQQILSGVDYCHRRHMVVRDLKPEN 360
TP+D FVMVEYVSGGELFDYICK+GR++E E+RRLFQQILS VDYCHRRHMVVRDLKPEN

AMPK_A2 85 TPTDFVMVEYVSGGELFDYICKHGRVEEMEARRLFQQILSAVDYCHRRHMVVRDLKPEN 144

HSU22456 361 VLLDAHNAKIADFGLSNMMSDGEFLRTSCGSPNYAAPEVISGRLYVGPVDLW 522
VLLDAHNAKIADFGLSNMMSDGEFLRTSCGSPNY APEVISGRLY GPEVD+W

AMPK_A2 145 VLLDAHNAKIADFGLSNMMSDGEFLRTSCGSPNYTAPEVISGRLYAGPEVDIW 198

This comparison revealed that the sequence we amplified from human intestinal cDNA was different from the previously characterized mammalian AMPK- α 2 sequences (including rat, pig and human). Our RT-PCR fragment was distinct enough at the DNA level [\sim 80% identity] from the published human AMPK- α 2 sequence of Gao *et al.* (1995) to represent a novel gene copy rather than a variant allele (Fig 3.B). The sequence we obtained from human intestinal cDNA shows only 91% identity to the human AMPK- α 2 sequence at the amino acid level (Fig 3.C). In contrast, a second rat AMPK- α isoform subsequently reported by Carling *et al.* in 1996 is 100% identical to our sequence at the amino acid level over the 175 residues spanned by our PCR product. Stapleton *et al.* (1996) subsequently reported cloning of the full-length gene for this second rat AMPK- α isoform (now known as the AMPK- α 1 isoform). The rat AMPK- α 1 sequence reported by Stapleton *et al.* (1996) is identical to the conceptual translation of our partial human cDNA sequence (small differences are noted over the region coded by the primers, although these are likely the result of the inosines included at selected third codon positions). Cloning of the full-length copy of the human counterpart was beyond the scope of this thesis and it was therefore not pursued, however, the full-length cDNA for the human AMPK- α 1 isoform was recently obtained from mammary gland tissue by Yano, K. (1999; unpublished; Accession # AB022017) and also agrees with the sequence reported here.

The region from which the murine *ink76* sequence is derived is very conserved among mammalian AMPK isoforms: rat and human α 1 and α 2 pairs are 100% identical over the stretch of 66 amino acids represented by *ink76*. Comparing α 1 to α 2 isoforms reveals that 64/66 amino acids are identical. Interestingly, *ink76* is distinct from both mammalian α 1 and α 2 isoforms at the amino acid level (61/66 and 62/66 similar residues, respectively). If *ink76* were to represent the murine homologue of either mammalian α 1 or α 2 isoforms, it might be expected to be 100% identical to either of these sequences at the amino acid level. Instead, similarity levels are in the range of what

might be expected for a separate mammalian isoform and it thus appears distinct enough from the other cloned mammalian AMPK- α genes to be ruled out as the murine homologue of either $\alpha 1$ or $\alpha 2$ isoforms. When searching the GenBank database with *ink76* using the BLAST algorithm, a number of fungal and plant SNF1 homologues are returned as the highest scoring hits, followed closely by the mammalian AMPK- $\alpha 1$ and AMPK- $\alpha 2$ sequences. Thus, although its full-length gene remains uncharacterized, the mouse *ink76* gene may either represent a third mammalian AMPK- α isoform or it may represent another type of SNF1-related kinase. Confirmation of this hypothesis would require either the cloning of a longer PCR product or of the full-length murine *ink76* gene, the determination of its presence in the mouse genome through Southern hybridization, and determining if a similar isoform is present in other mammalian species. Attempts to obtain a longer *ink76* fragment through inverse PCR on mouse genomic DNA proved unfruitful (B. Benkel, personal communication).

3.1.2.1 Expression studies on novel Human SNF1-like kinase domains. Northern blots of liver, intestinal and skeletal muscle poly-A⁺ RNA probed with the human AMPK- $\alpha 1$ PCR product did not reveal any transcripts in any of the tissues tested (not shown). When RT-PCR using nested $\alpha 1$ -specific primers was used to detect possible transcripts in each of the different tissues tested by Northern blotting, we were unable to amplify the $\alpha 1$ kinase from either liver or skeletal muscle cDNA despite using conditions that could be used to re-amplify AMPK- $\alpha 1$ from intestinal cDNA (not shown). Stapleton *et al.* (1996) have since shown that the human $\alpha 1$ isoform is expressed at very low levels over a wide range of tissues. However, in their experiments Stapleton *et al.* (1996) use 4 times more poly-A⁺ RNA than was used in the Northern blot experiments described here.

3.2 Cloning of novel protein kinases from *Drosophila*.

Both unique 192 bp SNF1-related sequences obtained from *Drosophila* were used separately to screen a genomic λ -Fix II library. Approximately 2.5×10^5 pfu's were initially screened with each the DM_SNF1A and DM_SNF1B probe, yielding a large number of primary positive signals. Eight clones hybridizing to the DM_SNF1A probe and four clones hybridizing to the DM_SNF1B probe were carried to third screening. Lambda DNA was prepared from these clones for further characterization through subcloning and sequencing.

3.2.1 Restriction mapping and Southern analysis of λ clones.

As a first step in their characterization, positive λ genomic clones were first restricted in single and double digests with *NotI* and *EcoRI*. These digests were used to construct a basic restriction map (Fig 4.A) and were subsequently blotted onto nylon membranes for Southern hybridization analysis using the same probes that were used during library screening. This allowed us to determine smaller restriction fragments hybridizing to the probes that could be subcloned into plasmids for further restriction mapping and sequencing (not shown).

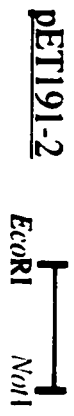
3.2.2 Characterization of *DM_SNF1B-1* and *DM_SNF1B-2*.

The lambda clones λ 179-A4 and λ 179-D1 yielded *EcoRI/NotI* fragments (of 6.5 and 0.7 Kb, respectively) that hybridized to the DM_SNF1B probe and were thus subcloned. Sequencing of the small 0.7 Kb *EcoRI/NotI* subclone (pET191-2, seen in Fig 4.A) allowed us to confirm identity with the DM_SNF1B probe. We then sequenced the larger 6.5 Kb *EcoRI/NotI* subclone [pET193-2, seen in Fig 4.A] and were able to determine that pET191-2 is fully contained within pET193-2. We thus set out to sequence the longer subclone in order to obtain sequence that spanned the complete gene.

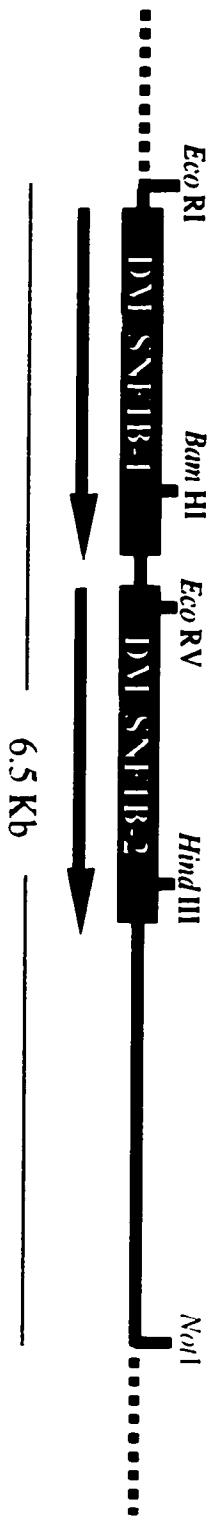
Figure 4. Restriction mapping and subcloning of the *Drosophila* DM_SNF1B gene cluster.

(A) Lambda clones obtained from genomic screening using the DM_SNF1B probe were digested with *EcoRI* and *NotI* and Southern blotted onto nylon membranes and subjected to Southern hybridization with the DM_SNF1B probe. The plasmid subclones pET193-2 and pET191-2 were identified as *EcoRI/NotI* restriction fragments hybridizing to the DM_SNF1B probe [the region of overlap between the probe and the two fragments is shown in grey. The *NotI* sites highlighted in red represent artificial sites resulting from the library construction process]. Shown below is the organization of the 6.5 Kb insert of the pET193-2 subclone as determined by restriction mapping and sequencing of both strands. (B) Southern blot of *Drosophila* genomic DNA probed with DM_SNF1B (from l. to r., *HindIII*, *HindIII/BamHI*, *BamHI*, *HindIII/KpnI*, *KpnI*, *HindIII/EcoRV*, *EcoRV*).

4.A



4.B



1 kb

Figure 5. Sequencing strategy used on subclone pET193-2.

Two sets of nested deletion clones were generated by Exonuclease III digestion. Sequencing was performed on a panel of representative clones representing progressive deletions of the 6.5 Kb insert. Shown in (A) is the linearized plasmid and the direction of deletions performed in (B). (B) fragments resulting from a progressive digestion of the *EcoRI* end. (C) A set of representative clones with unidirectional deletions spanning the entire insert. Both strands were sequenced by reversing the direction of Exonuclease III deletions and sequencing of a second panel of clones.

3.2.2.1 Restriction mapping and sequencing of the *DM_SNF1B-1 / DM_SNF1B-2* gene cluster. We first sequenced subclone pET191-2 because of its small size. Since it overlaps with the DM_SNF1B probe we could expect to find relevant sequences immediately. This enabled us to quickly confirm identity with the probe. The sequence surrounding the *EcoRI* site of pET191-2 is homologous to the amino-terminal region of the human KP78 catalytic domain and partially overlaps with the DM_SNF1B probe (Fig 4). The 0.7 Kb region contained within subclone pET191-2 bears sequences corresponding from -45 bp to +647 bp with respect to the putative start ATG codon of Dm_snf1b. The *NotI* site of pET191-2 truncates the sequence at position G217 along the Dm_snf1b ORF.

Because of the larger size of pET193-2 (6.5 Kb), we first had to establish the location of *DM_SNF1B* within the insert. We initially took advantage of naturally occurring *HindIII*, *EcoRV*, and *BamHI* sites (refer to Fig 4) to subdivide and further subclone the 6.5 Kb insert in order to determine the regions of interest by quickly obtaining sequences upstream and downstream of all three restriction sites using standard vector primers. The sequence surrounding the *EcoRI* site of pET193-2 is homologous to the amino-terminal region of the human KP78 catalytic domain, just as is found in the case of pET191-2. The sequence obtained also revealed that pET191-2 and pET193-2 share the *EcoRI* site in common (Fig 4). This implied that pET191-2 is fully contained within pET193-2 and that the *NotI* site of pET191-2 is an artificial site resulting from the library construction process (Fig 4). These results also implied that pET193-2 had ~ 5.8 Kb of sequence downstream of the pET191-2 *NotI* site and that it should be possible to obtain the complete gene sequence by extending the sequence through primer walking.

Unexpectedly, the sequence surrounding the *EcoRV* site of pET193-2 was also found to share homology to the amino-terminal region of human KP78. This is despite the fact that the two

restriction sites are separated by approximately 2 Kb (refer to Fig 4). Close inspection of the sequence surrounding the *EcoRI* and *EcoRV* sites revealed sequences that are similar but not identical to each other. Further sequencing obtained by primer walking revealed what appeared to be two tandem copies of the *Drosophila SNF1B* gene.

We used unidirectional Exonuclease III deletions in order to fully sequence both strands of the complete pET193-2 insert (Fig 5 A, B, C). The resulting sequence data (Fig 6) reveals two tandem copies of the *DM_SNF1B* gene separated by a short 189 bp spacer found between the TAG stop codon of *DM_SNF1B-1* and the ATG start codon of *DM_SNF1B-2*. *DM_SNF1B-1* and *DM_SNF1B-2* are expected to code for polypeptides of 662 and 559 amino acids respectively (Fig 7). Although highly conserved in the core catalytic domain region, the two gene copies are quite divergent at their carboxy-terminal ends. Both gene copies share high similarity (approximately 83% within the catalytic domain) with the putative human protein kinase gene KP78, a kinase that appears to form part of a distinct sub-branch within the SNF1 protein kinase family.

The presence of two or more gene copies in close proximity to each other is not unknown in *Drosophila*. As mentioned in section 1.5, two α -amylase genes are present in the *Drosophila* genome and are separated by a 4 Kb spacer region (Gemmill *et al.*, 1986; Boer and Hickey, 1986; Benkel *et al.*, 1987). An even more interesting example is observed with genes encoding the serine protease trypsin. Davis *et al.* (1985) and Wang (1995) have uncovered a total of eight trypsin genes clustered within a 12 Kb region of the *Drosophila* genome. With this precedent, it was important to investigate the possibility of other *DM_SNF1B* genes in the *Drosophila* genome. Two lines of evidence seem to suggest that, other than the two-gene cluster, there are no additional *DM_SNF1B*

Figure 6. Sequence of a 6.5 Kb region bearing the *DM_SNF1B-1/DM_SNF1B-2* gene cluster.

Complete sequence of this genomic region is shown for one strand (*Eco*RI to *Not*I direction). Restriction sites included in are highlighted in green. The coding regions for the genes *DM_SNF1B-1* and *DM_SNF1B-2* are shown along with their conceptual translations. Putative start (ATG) and stop (TAG) codons are boxed in blue and red respectively. Note the small spacer between the TAA codon of *DM_SNF1B-1* and the ATG codon of *DM_SNF1B-2*. The GenBank accession # for this sequence is AF023484

1640 L A I H L F T J P E V A C K L I D K T A L N T I A P Q K L Y P E V N I M K K L N 116
1641 R E N I V P L L L V I E D E P T L Y L V M E Y V S G D E L F N Y L V K N J P M F 159
1642 E R D A R V L F P L L V G A I E Y I H S K D I V H R D L F A E N L L L D L I M K 198
1643 L Y I A D F J F L T T F E P F A P L E T F T J E R P Y A A P E L P K J K F Y L J 219
1644 R E V L L W L L G V V L Y T L V D J D L R F D G T N L F E L R D R V L F J F Y F 278
1645 V R Y Y V J I E T E D L V R K F L V L N F T L R T D L C A Y M A D R W I N M J Y 312
1646 E L I N I L R F F L E F R M D L H D V N R L D L L E N M J H K R R D V E L L L F 359
1647 N L Y F D D I Y C A Y M L L D V A E R F P L T A T L E F C G C L F F E T F T A M F 396
1648 T L J R I F V P I A A P N Y T I J L V T F A L D K D T P N R F J A T J I R P M A 434
1649 F R L A D A L D F L E L T F R F K Y I C D C A G F A A N R R F D E F D C I P L 478
1650 C M E F F D V G D P V D V F T T L L C A L P F L A V N H F L T C A D H L I R D 518
1651 _____
1652 L A D E T T R D F F T R F E M L D J T L D F L K L L Y L M A L N L 561
1653 J T C R M T M H R T R V I R F D K L D A G V R R D C M F T F F C L J R L 604
1654 _____
1655 _____
1656 _____
1657 _____
1658 _____
1659 _____
1660 _____
1661 _____
1662 _____
1663 _____
1664 _____
1665 _____
1666 _____
1667 _____
1668 _____
1669 _____
1670 _____
1671 _____
1672 _____
1673 _____
1674 _____
1675 _____
1676 _____
1677 _____
1678 _____
1679 _____
1680 _____
1681 _____
1682 _____
1683 _____
1684 _____
1685 _____
1686 _____
1687 _____
1688 _____
1689 _____
1690 _____
1691 _____
1692 _____
1693 _____
1694 _____
1695 _____
1696 _____
1697 _____
1698 _____
1699 _____
1700 _____

Figure 7. Protein sequence alignment of the proteins encoded by *DM_SNF1B-1* and *DM_SNF1B-2* and the human kinase Kp78.

Sequences were aligned with CLUSTALW 1.74 (Thompson *et al.*, 1994) and formatted to highlight similarities and identities using BOXSHADE 2.7 (Kay Hoffman, unpublished) [biochemically similar residues are shaded in grey; identical residues are shaded in black]. The invariant residues of the core catalytic domain are boxed in red. Both *Drosophila* kinases have approximately %83 similarity to the human Kp78 within the catalytic domain. High conservation extends roughly 55 residues past the C-terminal boundary of the catalytic domain. The rest of the non-catalytic domain has much lower similarity values (yielding an overall similarity level of %39) and is characterized by several insertions/deletions between the sequences. Note the region boxed in blue, which represents an alignable loop that is shared between Kp78 and Snf1b-1, but is missing in Snf1b-2. The *Drosophila* sequences are based on conceptual translations from the nucleotide data shown in Figure 6. The human Kp78 sequence has the GenBank accession # S27966.

genes in the *Drosophila* genome. Genomic Southern blots fail to reveal the extra signals that might be expected from additional gene copies hybridizing to the *DM_SNF1B* probe (Fig 4.B). Additionally, *DM_SNF1B-1* and *DM_SNF1B-2* specific PCR primers derived from the pET193-2 sequence were used to re-screen by PCR the three remaining positive λ clones which had been obtained during the initial library screening. In each case, we were able to obtain amplification products of the expected size from either one or both of the gene-specific primer pairs (results not shown). Thus all four lambda clones appear to be related as judged by our PCR results.

3.2.3 Characterization of *DM_SNF1A*.

Lambda genomic clones λ 178-J1 and λ 178-J2 yielded overlapping *NotI* fragments of 2.3 and 8.2 Kb, respectively hybridizing to the *DM_SNF1A* probe and which were subcloned into pBluescript (SK+) for sequencing of the full length *DM_SNF1A* gene (Fig 8.A).

3.2.3.1 Restriction mapping and sequencing of the *DM_SNF1A* gene. Initial sequencing of the 2.3 Kb *NotI* insert of subclone pET190-2 immediately revealed a sequence which was homologous to the mammalian AMPK- α genes and we proceeded to sequence the full clone by primer walking (refer to Fig 9). Subclone pET190-2 contains sequence that includes the amino-terminal end of the kinase catalytic domain and a 500-bp intron that interrupts the putative ORF. It also extends to 44 bp downstream of the TAA stop codon. Although the clone pET190-2 codes for an ORF that includes 27 residues lying upstream of the amino-terminus of the kinase catalytic domain, the *DM_SNF1A* sequence is 5'-truncated at one of the *NotI* sites and lacks an in-frame start codon (Fig 8.A).

Sequencing of the subclone pET190-7 revealed that it contains the same naturally occurring *NotI* site that truncates the 5' end of the *DM_SNF1A* gene in clone pET190-2. This implied that

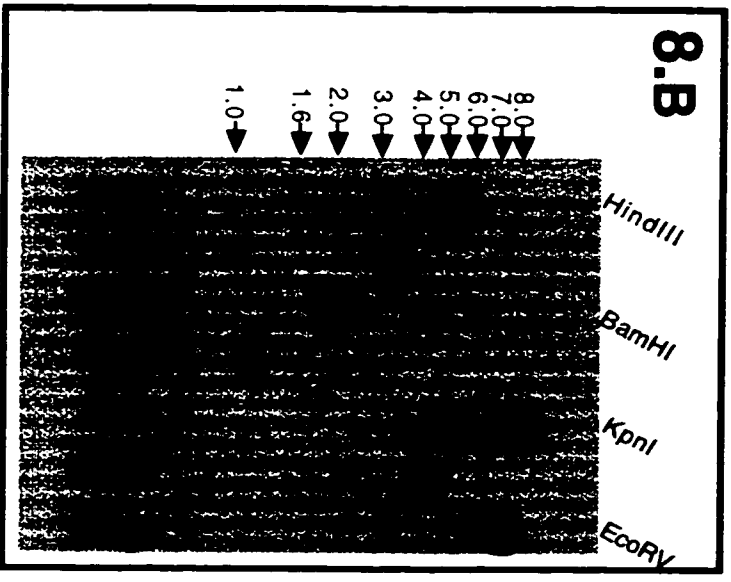
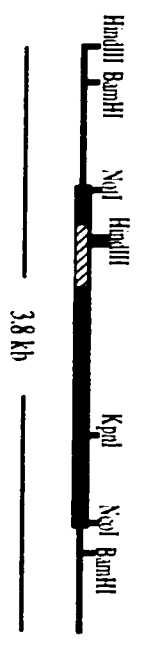
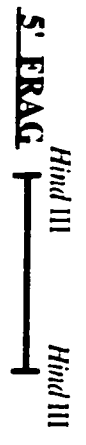
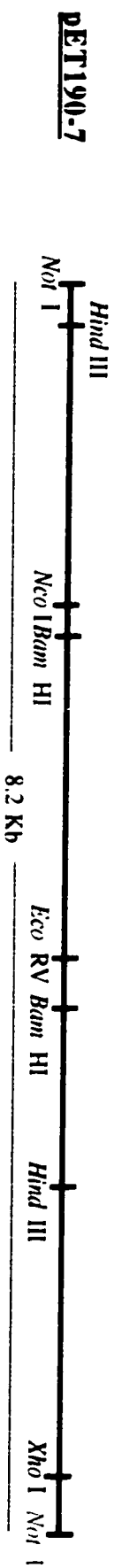
pET190-2 is fully contained within pET190-7 (Fig 8.A). These results also implied that pET190-7 has ~ 6 Kb of extra sequence downstream of the *DM_SNF1A* stop codon compared to pET190-2. We obtained sequence up to 0.85 Kb downstream of the putative TAA stop codon by primer walking (refer to Fig 9) and also obtained sequence surrounding naturally occurring restriction sites (*EcoRV*, *BamHI* and *HindIII*) within the pET190-7 insert through further subcloning. Due to the lack of other exceptional features downstream of the *DM_SNF1A* gene, full sequencing of the 6 Kb downstream of *DM_SNF1A* was not pursued.

In order to obtain the 5' end of the *DM_SNF1A* gene, the 5' *NotI/HindIII* fragment (300 bp) on pET190-7 was used as a probe on a *HindIII* digest of the λ 178-J1 genomic clone. An overlapping 1.2 Kb *HindIII* fragment was obtained which enabled us to obtain sequence for approximately 0.85 Kb upstream of the *NotI* site. The various sequences obtained represent a 3.8 Kb contig containing the entire *DM_SNF1A* coding region (including an 0.5 Kb intron), 0.85 Kb upstream of the putative ATG start codon and 0.85 Kb downstream of the stop codon (Fig 10). Two in-frame ATG codons are found a mere 21 bp upstream of the *NotI* site shared by pET190-2 and pET190-7 (Fig 10). The presence of several in-frame termination codons upstream of these potential start codons suggests that either of the ATGs may represent the functional start codon. Kozak (1987) has described a consensus sequence for efficient initiation of translation in vertebrate nuclear genes. A similar analysis has been performed for *Drosophila* nuclear genes (Cavener, 1987), enabling the prediction of the likely translational start site of a gene based on sequence surrounding putative in-frame ATG codons. In the case of *DM_SNF1A* the sequence CACTTCATG, surrounding the first of the two in-frame start sites, comes closest to the *Drosophila* consensus sequence of CAAAACATG described by Cavener (1987). The putative *DM_SNF1A* promoter contains a TATA-box 201 bp upstream of the putative start codon, which places the

Figure 8. Restriction mapping and subcloning of the region containing the *DM_SNF1A* gene.

(A) The plasmid subclones pET190-7 and pET190-2 were identified as *NotI* restriction fragments hybridizing to the *DM_SNF1A* probe. The region of overlap between the probe and the two fragments is shown in grey; the *NotI* sites (highlighted in red) represent artificial sites resulting from the library construction process. Shown below pET190-7 and pET-190-2 is the organization of a contiguous 3.8 Kb sequence which includes the full-length coding region, a 500 bp intron, approximately 0.85 Kb sequence upstream of the start codon and 0.85 Kb downstream of the stop codon. The cDNA clone shown was independently obtained through RT-PCR (Refer to Sections 2.6 and 3.2.3.2). (B) Southern blot of *Drosophila* genomic DNA probed with *DM_SNF1A* (from l. to r., *HindIII*, *HindIII/BamHI*, *BamHI*, *HindIII/KpnI*, *KpnI*, *HindIII/EcoRV*, *EcoRV*).

8.

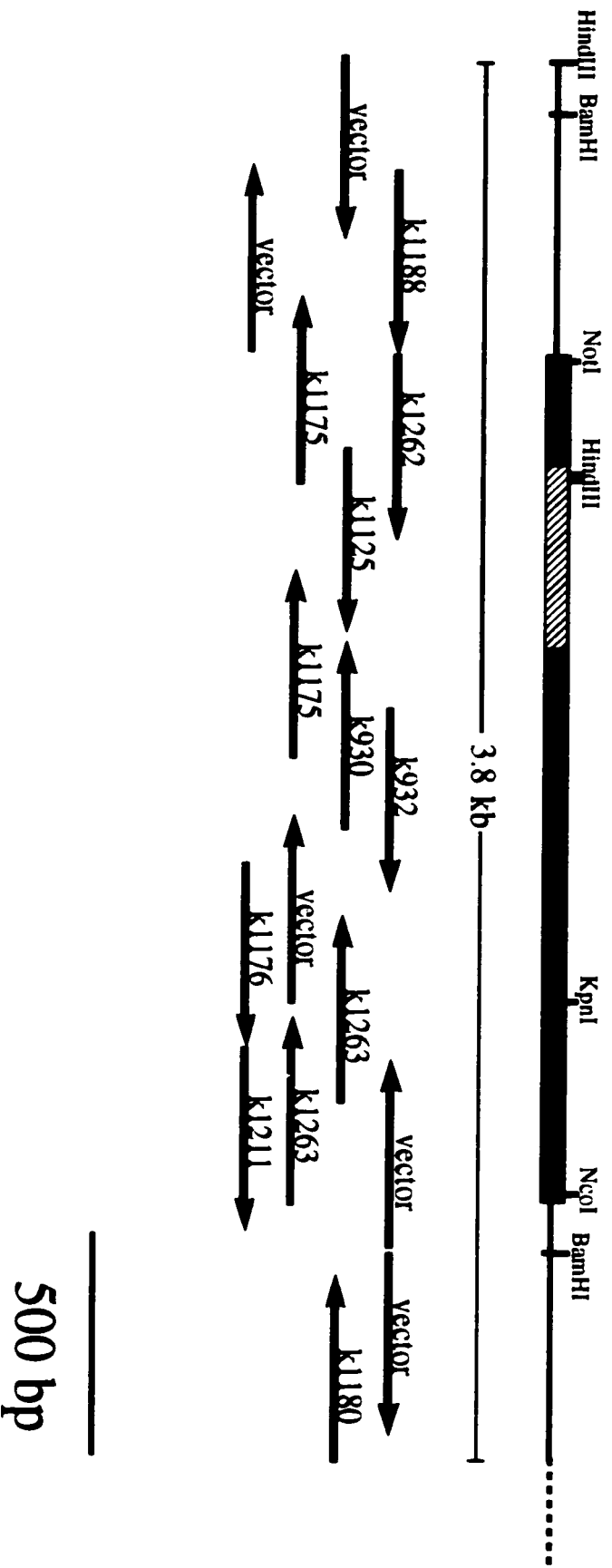


1Kb

Figure 9. Sequencing strategy used to obtain the *DM_SNF1A* gene.

The region containing the full length *DM_SNF1A* gene was obtained by a combination of subcloning, which allowed for sequence to be quickly obtained around naturally occurring restriction sites, and primer walking. The sequence was obtained by sequencing overlapping *Hind*III and a *Nor*I fragments (only one *Nor*I site is shown). Arrows represent individual sequencing runs obtained with either primers designed on previously obtained sequence (e.g. K1188) or vector primers used on the initial sequencing of subclones. Sequence representing the majority of both strands was obtained by organizing the various sequencing runs into a single contig.

9.



putative transcription start site for *DM_SNF1A* at 172 bp upstream of translational start site. The open reading frame contained within *DM_SNF1A* is predicted to code for a 559 amino acid polypeptide. Although the protein encoded by *DM_SNF1A* shares high similarity with the yeast *SNF1* kinase (61% identical and 77% similar within the catalytic domain; 40% identical and 60% similar overall), it shares an even higher level of sequence conservation with the mammalian *AMPKs*. The deduced Dm_Snf1a and human Ampk- α 1 and Ampk- α 2 amino acid sequences share high levels of sequence conservation within the core catalytic domain (identity: 81%[α 1] and 82.5%[α 2]; similarity: 92%[α 1] and 92%[α 2]) although the similarity extends beyond the core catalytic domain and near co-linearity is maintained which extends to the carboxy-terminal region (overall identities are 57% [α 1] and 60% [α 2]; overall similarities are 69% [α 1] and 70% [α 2]) (refer to Fig 11A).

3.2.3.2 Cloning of a cDNA copy of *DM_SNF1A*. The genomic copy of *DM_SNF1A* is interrupted by a 0.5 Kb intron (Fig 10). In view of plans to use *DM_SNF1A* in complementation assays on yeast *snf1* mutants, we wanted to obtain a full-length cDNA sequence to subsequently insert into an expression vector. Although initial attempts to get the full length gene by RT-PCR and screening of a cDNA library failed to yield the 5' end of the coding sequence, RT-PCR with a high fidelity Taq polymerase finally allowed us to obtain a cDNA clone representing the uninterrupted *DM_SNF1A* coding sequence from putative start codon to putative stop codon. The predicted intron position (see Fig 10) based on the well-known eukaryotic splice consensus sequences of Shapiro and Senapathy (1987) fits with the observed AG/GT sequence in the putative splice junction of the cDNA clone.

Figure 10. Sequence of a 3.8 Kb region bearing the complete *DM_SNF1A* gene.

Complete sequence of this genomic region is shown for one strand (*Hind*III to *Bgl*III direction). Restriction sites included in Figure 5 are shown in green. The coding region for the genes *DM_SNF1A* is shown along with the translated protein sequence. The putative start (ATG) and stop (TAG) codons are boxed in blue and red, respectively. Highlighted in yellow is an intron that splits the coding sequence (consensus splice sites are shown in bold underlined). The position of the intron was verified by sequencing of a cDNA clone. The GenBank accession number for the genomic sequence is AF020309; the accession number for the cDNA sequence is AF020310.

HindIII

AAATTTGACATATACACGCATATTTGTGAACAACTGAATTTTACCATACATCTTTGAGAATTTTAGGACAGTATTCTTGATAATCTATAT 90
 BamHI
 91 TGAATGTTGATCTACACTTAACGTTGCACAAAAGAGGTATCTTTCTGACCTTCTATGCGCGTGGAGTGCATCCAAAAATCAACTTGGTG 180
 181 TAATAAATGTGTATTTAAATCACAACCGGGCATCTTTCCCTTTCAATTGGACATCAGAGAGTTTCGACACCGCACACCTACTAGTTTT 270
 271 TTTGCGAGCTTAACATTTTCAAATGTGGTTGAGTAAAGCCATTAATATTGGCTTTTTATAAATTCAGCATAAGTGGCAAGATGGTTTTAAA 360
 361 TTTGCGCGCACTGCGGAAATACCTGGCATCACTGGCCGCACTTATCGATAGCCACTGGAAGGCGAGAGAGAGAGAAAGAGAGGGCGC 450
 451 GAGTCAGCTGCTGGAACGAGTGTGTGGAAGGCATGTTGTTGGAAGTGGCAGCGCCTTCCAATGTCAATTAATAAGCAAAACCGGAAG 540
 541 GAAACCCAAAGCGTAAAGGAGAGTCTCGTTGCCGCGCTCGAATAGAAATCCAGGAAATCAGTATCAATTCAACCGCCCTTGACACATCCAC 630
 631 AACCAATTTGACATCCGAATCCGTATCCGGGCGAGGTGGCAGCGCTGGCAGGAGGAAGTAAAGTGAAGTGGCTGCTTTATTTTATTTTAT 720
 721 TTTGTTGATTACGCGGAGTACAATACGCAATTACCGATAACGACCTTACCACGGTTCCCTTTCCGCGCCATCCGCGAGCGGAGCAGTGG 810

NotI

311 AGCACCCCCAGTCCAACAAGGAGCTACCTGCTCGAG[REDACTED]CCCCAGATGAGGGCTGGCGCCGCTGAGGGCGGTGGCCCGCGGCAGCGCCAAT 900
 DM_SNF1A → M P Q M R A A A A E A V A A G S A N 18

901 GGCCAGCGCTGGTCAAGATTTGGGCACTACCTACTGGGCGCAACCTGGGAACGGGAACCTTCGGCAAGGTGAGATCGGGGAGCACCAG 990
 S Q P L V K I G H Y L L G A T L G T G T F G K V K I G E H Q 48

991 ATCACACGCGTCAAGGTGGCCGTCAGATACTCAATCGTCAGAAGATCAAAAGCCTGGACGTCGTGGGCAAGATCCGGCGTGAGATCCAG 1080
 I T R V K V A V K I L N R Q K I K S L D V V G K I R R E I Q 78

HindIII

HindIII

1081 AACCTAAGCTCTTCCGCCATCCGCACATCATCAAGTTGTACCAAGTGGAGTTTCATGGTTCCAAAGCTTAGACTTTCTAAGTCCAAGCTTTTC 1170
 N L K L F R H P H I I K L Y Q 94
 1171 CAGGAAACTGGGTTACTGGTTTCTCTTCCAATTACGCTGGCCAGCTACTAATAACAAACAATCAAGCCGCGAGTTTAAATAACAAAAGG 1260
 1261 CAGGATATGCCCGAAGGGGAATAGCTTTTCTATAACAATAATTTGTTTTATTTTTGTTTATCTATAAAGTACACGGTCATGCTAGTCA 1350
 1351 TCCGAATTTTGTGCGGAAGGCAATGCATATCTAGAACTGGGCCACTGGCTAGACATTCCGCAACAGACATAGAAGCCCAACTTAAATG 1440
 1441 AGCAAAAAGTTTGAAGGAAGTTTATAAAAGGAATATGTACGGGGTCTGCTTTGGTTTACAAAGCACATTTCCCTCGATTTTCGGATT 1530
 1531 GGATCAGTTACCAATTAAGTTAGGCAGGGCTTTGGGGAAACCTAAGGACGGCTGCAAGTCATCAGCTGATATAAACCCTGCTTGCTAACCTT 1620
 1621 TCCGTAACCTCCGAAAGCTCATATCGACCGCGCTCGGACATCTTCATGATCATGGAGTACGCTGAGCGCGGGGAGCTGTTTGATTATATTG 1710
 V I S T P S D I F M I M E Y V S G G E L F D Y I V 119

1711 TGAAGCACGGCAAGCTGCAGGAGCACCAAGCGCGCGCTTTTTTTCAGCAGATTATCTCTGGCGTGGACTATTGCCACCGGCACATGATCG 1800
 K H G K L Q E H Q A R R F F Q Q I I S G V D Y C H R H M I V 149

1801 TGCATAGAGACCTGAAGCCAGAGAACCTGTGCTCGACCACAACATGCACGTGAAGATCGCCGACTTTGGGCTCTCGAACATGATGCTGG 1890
 H R D L K P E N L L L D H N M H V K I A D F G L S N M M L D 179

1891 ACGGCGAGTTCTTGGCACCTCGTGGCGCTCGCCCAACTACGCGCGCCCGAGGTGATTTCCGGTAAGCTGTACGCCGCTCCCGAGGTGG 1980
 G E F L R T S C G S P N Y A A P E V I S G K L Y A G P E V D 209

1981 ACATCTGGTCTGCGCGCTCATCTGTACGCCCTGCTGTGCGGGACGCTGCCCTTTGACGACGAGCATGTGCCTACGCTGTTTCGCAAAA 2070
 I W S C G V I L Y A L L C G T L P F D D E H V P T L F R K I 239

2071 TCAAGTCCGGTATTTTCCCCATTCCCAGTACCTCAACAACAAGTGGTCAATCTGGTCTGTCTCAGATGCTGCAAGTGGACCCCGCTCAAGC 2160
 K S G I F P I P E Y L N K Q V V N L V C Q M L Q V D P L K R 269

2161 GGGCCAAACATCGAGGAAATCAAGAAGCACGAGTGGTTCCAAAAGGACCTCCCGGCTACCTCTTCCCCTCGTCCATCGAGCAGGACTCCA 2250
 A N I E E I K K H E W F Q K D L P A Y L F P S S I E Q D S N 299

2251 ATGTGATCGACACCTACGCTGTGGCAGAAGTCTGCACCAAGTTCGGCGTCAAGGAGACCGAGGTGCACAACCTCGCTGCTCAGCGCGGATC 2340
 V I D T Y A V A E V C T K F G V K E T E V H N S L L S G D P 329

2341 CGCACGATCAGCTGGCCATCGCTTACCATCTGATCATCGACAACAACGCTTCCCGACGACGCGGCCAACCAATAACAGATCAACA 2430
 H D Q L A I A Y H L I I D N K R F A D D A A N Q I N E I N N 359

2431 ACTTCTTTGTGGCCGGCTCTCCACCACCGCGCGCTCCGCCGCTTCCCCAATCATCGATGGACCACCAAGCCCTCTGGCCACCGTGA 2520
 F F V A G S P P P P P P P V P Q S S M D H Q A P L A T V T 389

KpnI

2521 CCCTTGGCGGAATTAATTCAGCCAGCTCGGGCACCGCCACACAGTGCACAGTGGCTGGAGGCAGCCCTAGCAGCACCATACCCATTTC 2610
 V G G G T S A S S G T A T P V P P V A G G T P S S T I P I R 419

2611 GACCGCATCCGGAGCGGATTGCGCCGATGCGGGACCGTCAACTGGCCATGTCCGTGCAGACATCTGGCGGGGAGCCCTTCCGGGAGAAAA 2700
 P H P E R I A P M R D R Q L A M S V Q T S G G G A F P E K T 449

2701 CCGCACGTTGGCGGCACTCCCATCAAGCGCGCAAGTGGCATCTGGGCATCCGGTTCGAGAGCAAGCCCAACGACATAATGCTCGAGGTAT 2790
 A R G G T P I K R A K W H L G I R S Q S K P N D I M L E V Y 479

2791	ACCGAGCGATGAAGGCACTCAGCTACGAGTGAAGATCATCAATCCGTACCACGTTCCGGGTCCGCAGGCAGAACGTGAAGACCGGCAAGT	2880
	R A M K A L S Y E W K I I N P Y H V R V R R Q N V K T G K F	509
2881	TCTCGAAGATGTCACTGCAGCTGTATCAGGTGGACGCCAAGAGCTATCTGCTCGACTTCAAGTCGCTGACCAACGACGAGGTCCGAGCAGG	2970
	S K M S L Q L Y Q V D A K S Y L L D F K S L T N D E V E Q G	539
2971	GCGACGACGTCATCATGGAGAGCCTTACTCCGCCGCCGCTTAGCGTGTCCGGGGTGATGCCCGCTGCAGCCGACCGGCCACCACACCATGG	3060
	D D V I M E S L T P P P L S V S G V M P L Q P T G H H T M E	569
3061	AGTTTTTCGAAATGTGCGCCGCCCTGATCATTCAACTGGCTCGC [REDACTED] GGATGAGGTGACCGCCATCGGCCCGCAGCGGACCGGATGGGATC	3150
	F F E M C A A L I I Q L A R *	584
3151	TCTCGGCTGCGGATCCGCACTCGGACCGTCTGAGGATTGACCATFACCGGAAGACACAGTCCCACAAAGGAAAGCTGGGGGGAGTTT	3240
	<i>Bam</i> HI	
3241	CGGCTCCGC [REDACTED] CCGGAGTCTGGCCTTAACCACATGACTAAGTTCCTTGCAATTCGTCCCAAACCCTTTTGATTTTAATCGGAAACCCA	3330
3331	ATGTGCAATTATCGGTAGCAGGATTTGGGAATAGATTTTAAAAATGGCCTGCTACTAGACGACGAACAAGAGCTTGGGACTCCGACTCCA	3420
3431	GCTTTTCTTTGTGCGGCTGTGTAATCCACATAAAATGTTATTGAACTCGGACACCGCCAACCGCGTTGCCGCTACGCTACCGTTACCCG	3510
3511	CTTTGCCGATCGAATTTCCGCTTCCGGTTAGTCTTAAGTCTGTAGTTCCGCCCGGTCCCGCGCTTCGATCCGGTTCCGTTCTGTTCGGT	3600
3601	TCCAATTCGAAATTAGTTTAATAAGCCAAGTTAAATTATTATCCAACCGCATGTGCACTTGAAGTTATTGGTTTAAGCCTAACAAATGTTA	3690
	<i>Bgl</i> III	
3691	ATGGACCTCCTCTGGAACCTTTGCAAGACACCAAAATGAACATTTTAGATCTGTTTCTGACCACCACCATAAAGACCTTATTCAACATTGG	3780
3781	AATTTATAATCCCATAAATTTTTTTT	3870

Figure 11. Protein sequence alignments featuring *Drosophila* DM_Snf1a and the human and yeast homologues.

Sequences were aligned with CLUSTALW 1.74 (Thompson *et al.*, 1994) and formatted to highlight similarities and identities using BOXSHADE 2.7 (Kay Hoffman, unpublished) [similar residues are shaded in grey; identical residues are shaded in black, the invariant residues of the kinase catalytic domain are boxed in red]. The *Drosophila* sequence is based on a conceptual translation of the cDNA (GenBank accession # AF020310), the sequences of the α -1 and α -2 human AMPK isoforms correspond to GenBank accession # BAA36547.1 and P54645, respectively, and the yeast SNF1 sequence corresponds to GenBank accession # P06782. (A) Three-way alignment between DM_Snf1a and both human Ampk- α isoforms. DM_Snf1a shares extensive similarity to both human isoforms within the catalytic domain (identity: 81% [α 1] and 82% [α 2]; similarity: 92.2% [α 1] and 92.2% [α 2]). It also shows significant levels of conservation in the non-catalytic domain (identity: 38.8% [α 1] and 42.3% [α 2]; similarity: 50.9% [α 1] and 53.47% [α 2]) such that it is possible to align all three sequences with the addition of a few indels [boxed in blue]. (B) Four-way alignment between DM_Snf1a, human Ampk α -1, human Ampk- α 2, and yeast Snf1. Regions of the alignment that show indels held in common between *Drosophila* and human sequences are boxed in green. Regions of the alignment that show an indel held in common between *Drosophila* and yeast sequences are boxed in blue.

11.B

HUM_AMPK1	15	EGHYLDDEIVLLEKMKELH
HUM_AMPK2	13	EGHYLDDEIVLLEKMKELH
DM_SNF1A	25	EGHYLDDEIVLLEKMKELH
YSC_SNF1	52	ENYGVKDEIVLLEKMKELH
HUM_AMPK1	75	HPHLEKIVLLEKMKELH
HUM_AMPK2	73	HPHLEKIVLLEKMKELH
DM_SNF1A	85	HPHLEKIVLLEKMKELH
YSC_SNF1	112	HPHLEKIVLLEKMKELH
HUM_AMPK1	135	IKVHRLDEKIVLLEKMKELH
HUM_AMPK2	133	IKVHRLDEKIVLLEKMKELH
DM_SNF1A	145	IKVHRLDEKIVLLEKMKELH
YSC_SNF1	171	IKVHRLDEKIVLLEKMKELH
HUM_AMPK1	195	PLADIVSIVLLEKMKELH
HUM_AMPK2	193	PLADIVSIVLLEKMKELH
DM_SNF1A	205	PLADIVSIVLLEKMKELH
YSC_SNF1	231	PLADIVSIVLLEKMKELH
HUM_AMPK1	255	PKRAIKIVLLEKMKELH
HUM_AMPK2	253	PKRAIKIVLLEKMKELH
DM_SNF1A	265	PKRAIKIVLLEKMKELH
YSC_SNF1	291	PKRAIKIVLLEKMKELH
HUM_AMPK1	297	IVKIECSIEVLSQYNRN
HUM_AMPK2	295	IVKIECSIEVLSQYNRN
DM_SNF1A	307	IVKIECSIEVLSQYNRN
YSC_SNF1	351	IVKIECSIEVLSQYNRN
HUM_AMPK1	344	INTSPD-----SFD-----HHT-----
HUM_AMPK2	342	INTSPD-----SFD-----HHT-----
DM_SNF1A	361	INTSPD-----SFD-----HHT-----
YSC_SNF1	410	INTSPD-----SFD-----HHT-----
HUM_AMPK1	361	PIPERPFAE--TPRRIITLDE
HUM_AMPK2	365	PIPERPFAE--TPRRIITLDE
DM_SNF1A	419	PIPERPFAE--TPRRIITLDE
YSC_SNF1	469	PIPERPFAE--TPRRIITLDE
HUM_AMPK1	417	RAKIDIVLLEKMKELH
HUM_AMPK2	421	RAKIDIVLLEKMKELH
DM_SNF1A	479	RAKIDIVLLEKMKELH
YSC_SNF1	524	RAKIDIVLLEKMKELH
HUM_AMPK1	465	IKSDDIVVQRSLSSTPQRS
HUM_AMPK2	469	IKSDDIVVQRSLSSTPQRS
DM_SNF1A	527	IKSDDIVVQRSLSSTPQRS
YSC_SNF1	584	IKSDDIVVQRSLSSTPQRS
HUM_AMPK1	525	DLTPRPSIIIVLLEKMKELH
HUM_AMPK2	527	DLTPRPSIIIVLLEKMKELH
DM_SNF1A	563	DLTPRPSIIIVLLEKMKELH
YSC_SNF1	620	DLTPRPSIIIVLLEKMKELH

3.2.3.3. Analysis of the sequence divergence between *DM_SNF1A* and its human counterparts. In comparing the protein encoded by *DM_SNF1A* to the two human isoforms, one notable observation can be made. *DM_Snf1a* has a subtle yet noticeably higher level of sequence identity to the α -2 isoform (Table 1). The effect is seen in the catalytic domain, where although sequence similarity with the two isoforms is identical, there are four extra identities with α -2. The effect is more pronounced within the non-catalytic domain, where another eleven extra identities are observed. Thus the *DM_Snf1a* shares a slightly higher level of overall sequence identity with the human α -2 isoform and this may have important functional implications. In order to investigate the divergence rates between these sequences, protein distances were calculated using the Dayhoff model of amino-acid substitution. Because the kinase catalytic domain is evolving at a slower rate than the rest of the sequence, distances were calculated separately for the catalytic and non-catalytic domains. Results are shown in Table 2. The protein distances calculated corroborate the observations made based on sequence similarity levels alone. The distances observed between *DM_Snf1a* and *Ampk- α 2* are smaller than those observed against *Ampk- α 1*, although the effect is more pronounced when the analysis is restricted to the non-catalytic domains. Similar results are obtained when the analysis is repeated by doing pairwise comparisons between *DM_Snf1a* and the rat *Ampk- α* isoforms.

3.2.3.4 The *DM_SNF1A* gene appears to be in single copy in the *Drosophila* genome. The two copies of *DM_SNF1B* raised the possibility that there may be other copies of *DM_SNF1A* in the *Drosophila* genome. We addressed this possibility through medium-stringency hybridization on genomic Southern blots and through the use of new PCR primers to attempt amplification of novel *DM_SNF1A*-like sequences. Medium-stringency hybridizations using the *DM_SNF1A* PCR-derived probe failed to reveal novel hybridization signals even on prolonged exposure (Fig. 8.B).

Table 1. Percentage of sequence identity and sequence similarity between *DM_SNF1A* and the human AMPK isoforms α -1 and α -2

	IDENTITY ^A			SIMILARITY ^B		
	Observed identities	total ^C sites	% identity	Observed similarities	total ^C sites	% similarity
<u>Overall :</u>						
<i>DM_SNF1A</i> vs AMPK- α 1	338	590	57.28	407	590	68.98
<i>DM_SNF1A</i> vs AMPK- α 2	353	589	59.93	415	589	70.45
<u>Catalytic Domain :</u>						
<i>DM_SNF1A</i> vs AMPK- α 1	209	258	81.00	238	258	92.24
<i>DM_SNF1A</i> vs AMPK- α 2	213	258	82.55	238	258	92.24
<u>Non-Catalytic Domain :</u>						
<i>DM_SNF1A</i> vs AMPK- α 1	129	332	38.85	169	332	50.90
<i>DM_SNF1A</i> vs AMPK- α 2	140	331	42.29	177	331	53.47

Legend :

^A - number of identical residues

^B - sum of (identical residues) + (biochemically similar residues)

^C - number of sites considered in the sequence alignment, including gaps.

Table 2. Dayhoff distance matrix for the catalytic and non-catalytic domains of seven different protein kinases.

Kinases were aligned and Dayhoff distances were calculated for the catalytic domain (top half) and the non-catalytic domain (bottom half). HUM_BARK1 and HUM_BARK2 represent the sequences of human β -adrenergic receptor kinases, which were used to calculate the divergence levels of an unrelated pair of kinases. Pairwise distances between DM_Snf1a and each of the human Ampk isoforms are coded in yellow. Pairwise distances between Dm_Snf1a and rat Ampk isoforms are coded in green. As reference points, distances were also calculated between human β -adrenergic receptor 1 and the human Ampk- α isoforms (coded in pink) and between human β -adrenergic receptor 2 and the human Ampk- α isoforms (coded in light blue).

	RAT_AMPK1	HUM_AMPK1	RAT_AMPK2	HUM_AMPK2	DM_SNF1A	HUM_BARK2	HUM_BARK1
RAT_AMPK1	-	0.00388	0.11471	0.10970	0.23990	1.74409	1.85638
HUM_AMPK1	0.01688	-	0.11908	0.11404			
RAT_AMPK2	0.46815	0.45634	-	0.01625	0.21825	1.84952	1.88249
HUM_AMPK2	0.47927	0.47196	0.03424	-			
DM_SNF1A		0.77354		0.643950	-	1.82661	1.85623
HUM_BARK2	4.06677		4.17843		4.38323	-	0.08605
HUM_BARK1	4.33212		4.44877		4.55868	0.26689	-

New PCR primers (K1184 and K1187, Section 2.1.2) were designed based on conserved features between yeast SNF1 and two plant kinases that have been shown to complement the *snf1* mutation (*RKIN1* and *NPK5*). The original primers K930 and K932 (Fig 1.B) that were used to obtain DM_SNF1A and DM_SNF1B probes were not entirely specific to *AMPK- α* and *SNF1*, since they could amplify a lesser SNF1 relatives such as *DM_SNF1B-1*. K1184 and K1187 are based on peptides that are conserved only among the closest SNF1 relatives and would thus be expected to be more discriminating than the original K930-K932 primer pair. When these primers were used on *Drosophila* genomic DNA, we could only amplify *DM_SNF1A*. This primer pair is expected to amplify each of the sixteen close relatives of the SNF1/AMPK kinases included in the sequence alignment shown in Figure 14. Any extra *DM_SNF1A* gene copies in the *Drosophila* genome should therefore represent amplifiable targets, especially given the full degeneracy of the primers.

3.2.4 Expression studies on the novel *Drosophila* kinases.

Expression of the *DM_SNF1A* and *DM_SNF1B* genes was studied by generating gene-specific probes by PCR and using the resulting probes in Northern hybridization experiments. Northern blots of whole-body total RNA from larvae and adults grown in glucose-supplemented and glucose-deficient medium were hybridized to PCR-derived gene specific probes. None of the probes yielded hybridization signals despite the fact that genomic Southern blots performed in parallel yielded strong hybridization signals. Integrity of the RNA or problems during the nucleic acid transfer would not appear to have been a problem as hybridization to a control α -amylase probe yielded a strong signal of expected size \approx 1.5 Kb (results not shown). The experiments were repeated using poly-A⁺ enriched RNA preparations although once again hybridization signals could not be detected despite a working α -amylase control. It is worth noting that α -amylase is a highly

expressed gene in *Drosophila*. The lack of other hybridization signals could result from either very low levels of expression, expression in a small subset of cells or a combination of both. These problems are compounded by the fact that RNA was prepared from whole body tissue, which would have a diluting effect on potential hybridization signals.

3.3 Phylogenetic survey of the SNF1/AMPK protein kinase family

3.3.1 DM_SNF1B: relationship to other related kinases.

Based on sequence similarity, the proteins encoded by the *Drosophila* *DM_SNF1B-1* and *DM_SNF1B-2* genes are most closely related to the putative human kinase *KP78* and sequences related to *KP78*. Although *KP78* has been considered a member of the SNF1/AMPK family of kinases (see for example Becker *et al.*, 1996), it has lower similarity levels to *SNF1* and its closest relatives. Several *KP78*-related protein kinase sequences have been submitted to GenBank over the last few years, although their physiological role remains obscure.

3.3.1.1 Protein sequence alignments. As can be seen in Figure 12, the region aligns very well in all of the sequences analyzed. All of the invariant residues of the catalytic domain are fully conserved in all the sequences. This is not unexpected since the catalytic domain is responsible for the phosphate transfer (Hanks and Hunter, 1995) and, as such, there are many constraints on primary, secondary and tertiary catalytic domain structure. Becker *et al.* (1996) have previously defined a small region flanking the C-terminus of the catalytic domain, the SNH domain (for **SNF1 Homology Domain**), which appears to define a subfamily of kinases that includes SNF1, AMPK and KP78, among others. An interesting observation is the conservation of the threonine residue within the kinase "activation loop" (T173 in human Ampk- α 1), across all of the kinases included in Figure 12. Phosphorylation at this site is required for many kinases and allows for an open conformation that allows the peptide substrate to enter the catalytic cleft (Goldsmith and Cobb, 1994). This would predict the requirement of phosphorylation for kinase activity, and more importantly, predicts the presence of an upstream kinase-kinase activity in each of the organisms included in the alignment.

Figure 12. Protein sequence alignment of a group of kinases similar to *DM_SNFIB-1* and *SNFIB-2*.

Sequences were aligned with the program CLUSTALW 1.74 (Thompson *et al.*, 1994) and shaded to highlight amino acid similarities using BOXSHADE 2.7 (Kay Hoffman, unpublished). Identities are boxed in red while similarities are boxed in blue. Names of the sequences include the descriptive name of the protein (when available, the accession number is used otherwise) and the organism (HUM: human, MUS: mouse, RAT: rat, DM: *D. melanogaster*, CEL:*C. elegans*, YSC: yeast, Xen: *Xenopus laevis*). HUM_PKA-CA represents the sequence for the human cAMP-activated protein kinase C- α subunit. HUM_CAMKI represents the human calcium/calmodulin-dependent protein kinase. These two sequences were used to provide an outgroup for phylogenetic trees (Fig 13). Information regarding each sequence analyzed is shown in Appendix I.

HUM_AMPK1	17	I	D	T	K	HK	N	OK	R	LDV	G	
HUM_AMPK2	15	V	GD	V	E	Q	TT	HK	R	LDV	G	K
YSC_SNF1	54	Q		R	S	Y	R	OK	R	AKS	DMOG	E
HUM_KP78	55	R		R	Z	R	R	RE	Q	NPT	S	Q
MUS_EMK	52	R		R	Z	R	R	KE	Q	N	S	S
CEL_PARI	169	K		K	Z	K	I	HE	Q	NPS	S	Q
DM_SNF1B-1	75	K		K	Z	R	I	RE	Q	N	S	ARQ
DM_SNF1B-2	62	K		K	Z	R	I	RE	Q	N	I	ARQ
MUS_MSK	72	D	E		V	R	R	K	R	D	S	N
RAT_SNRK	15	D	D	R	H	V	R	F	R	D	L	ATGH
XEN_P69EG3	12	E	HE	R	T	G	S	L	F	EK	G	KT
HUM_CAMKI	19	DFRDV		T	A	S	E	I	EDKR	OKL	C	A
HUM_PKA-CA	43	ERI		T	S	M	V	K	KE	NHY	C	A

HUM_AMPK1	67	Q		F	H	Q	S	PSD		C	K	N	D
HUM_AMPK2	65	Q		FR	H	H	S	PTDF		V		N	D
YSC_SNF1	104	S	Y	R	H	Q	S	KDE	I				M
HUM_KP78	104	R				E		OKT				K	Q
MUS_EMK	101	R				E		EKT					K
CEL_PARI	218	K		Q	D	Q		EQT					K
DM_SNF1B-1	124	K		Q	D	Q		EQT		R			K
DM_SNF1B-2	111	N		K		L	Q	ERT		V		N	R
MUS_MSK	121	Q				L	Q	KDM		T	KN		N
RAT_SNRK	64	R	C			L	Q	OTK		L	G	D	N
XEN_P69EG3	60	D	A	N	S	O	H	C	F	P	K	K	N
HUM_CAMKI	67	A	S	H	K	K	A	D	D	Y	G	G	H
HUM_PKA-CA	93	R	I	Q	A	N	F	F	E	S	F	K	D

HUM_AMPK1	116	S	R	L	Q	D	R	H	M		A	H	A	L
HUM_AMPK2	114	R	L	Q	D	R	H	M			A	H	A	L
YSC_SNF1	152	R	F	Q	D	R	H	K			A	H	A	L
HUM_KP78	153	S	K	R	Q	R	H	K		A	A	A	A	L
MUS_EMK	150	A	K	R	R	L	H	Q	K	F				L
CEL_PARI	267	A	K	R	R	L	H	Q	K	F				L
DM_SNF1B-1	173	V	I	R	R	Q	L	S	K	N				L
DM_SNF1B-2	160	V	L	R	R	Q	L	S	K	F				L
MUS_MSK	170	O	K	V	A	E	S	N	H	H				L
RAT_SNRK	114	L	K	Y	A	H	S	K	L	H		V	F	L
XEN_P69EG3	109	V	F	R	R	A	I	S	Q	G	Y	A		L
HUM_CAMKI	116	S	R	L	I	F	D	K	L	D	L	G		L
HUM_PKA-CA	142	H	F	Y	A	A	L	T	F	E	L	S	L	A

HUM_AMPK1	163	M	M	S	D	E	-	F	R	S	N	S	L	A	S	A	A
HUM_AMPK2	161	M	M	S	D	E	-	F	R	S	N	S	L	A	S	A	A
YSC_SNF1	199	I	M	T	D	N	-	F	K	S	N	S	L	A	S	A	A
HUM_KP78	200	E	F	T	V	G	-	K	D	F	P	P	P	P	P	P	S
MUS_EMK	197	E	F	T	F	N	-	K	D	F	P	P	P	P	P	P	S
CEL_PARI	314	T	F	S	L	N	-	K	D	F	P	P	P	P	P	P	S
DM_SNF1B-1	220	T	S	D	P	N	A	-	Q	E	F	F	P	P	P	P	S
DM_SNF1B-2	207	T	T	F	E	P	K	A	-	P	E	F	F	P	P	P	S
MUS_MSK	217	F	Y	K	P	E	-	P	S	C	V						S
RAT_SNRK	162	K	F	O	P	K	-	K	T	S	C		L	A	S		S
XEN_P69EG3	156	A	K	P	K	G	L	D	Y	H	M	C	A	G	V		S
HUM_CAMKI	166	K	M	E	D	P	S	-	V	S	C	A	L	A	G	V	S
HUM_PKA-CA	189	K	R	V	K	G	R	T	-	-	-	-	V	L	E		A

HUM_AMPK1	211DHPTFKCIIYI.....QNPVIS.....MLLN
HUM_AMPK2	209DEHPTFKRGVYI.....ENRSVATMHQ.....
YSC_SNF1	247	RR.....DESFPVFKNSNVT.....KPGAAG.....
HUM_KP78	248Q.....KE.....R.....F.....T.....EN.....
MUS_EMK	245Q.....KE.....R.....F.....T.....EN.....
CEL_PARI	362Q.....KE.....R.....F.....T.....EN.....
DM_SNF1B-1	268GT.....KE.....R.....Y.....M.....EN.....
DM_SNF1B-2	255T.....KE.....R.....Y.....I.....ES.....
MUS_MSK	265P.....PT.....RQ.....E.....T.....
RAT_SNRK	210	QP.....OEA.....DSETLTM.....DC.....T.....PR.....AGRD.....TR.....VQR.....
XEN_P69EG3	206	Y.....DD.....MV.....YK.....E.....T.....K.....PGSVL.....SQ.....Q.....
HUM_CAMKI	213	YP.....YDE.....DAK.....F.....C.....K.....E.....E.....F.....DSPYWDD.....DSAKDF.....H.....EK.....
HUM_PKA-CA	234	YP.....FAD.....PIQ.....YE.....S.....V.....F.....S.....HF.....SDLKD.....NL.....Q.....LT



HUM_AMPK1	257	A.....KD.....E.....ODLPK.....Y.....F.....ED.....PSYSS.....
HUM_AMPK2	255	A.....KDRE.....E.....FKO.....LPS.....Y.....F.....ED.....PSYDA.....
YSC_SNF1	293	I.....HE.....ODD.....FKVDLPE.....Y.....LPDLKPHPEEENENNDS
HUM_KP78	294	G.....EQ.....KD.....A.....HEE.....DE.....K.....FV.....EPEL.....
MUS_EMK	291	G.....EQ.....KD.....V.....HED.....DE.....K.....YV.....EPLL.....
CEL_PARI	408	S.....DN.....KD.....V.....YED.....DE.....K.....FI.....EPPK.....
DM_SNF1B-1	314	T.....SA.....SD.....L.....HDES.....DR.....R.....FR.....EKPM.....
DM_SNF1B-2	301	T.....SA.....AD.....M.....YEQG.....NG.....R.....FR.....EKPM.....
MUS_MSK	311	I.....AQ.....RQ.....ADPTLLQ.....QDDPAFDMQ.....GYTSN.....
RAT_SNRK	256	A.....EE.....ES.....P.....QGVDPSP.....ATKYNI.....LVS.....YKN.....
XEN_P69EG3	252	I.....KH.....N.....P.....MH.....YSC.....EWQS.....KYPLG.....
HUM_CAMKI	263	F.....CEQALQ.....P.....IAGDTAL.....DKNIHQSVSEQIKKN.....
HUM_PKA-CA	280	FGNLKNG.....ND.....KN.....K.....FAT.....TDW.....AIYQR.....

HUM_AMPK1	287TMI.....EA.....KE.....CE.....F.....ECS.....E.....LSC.....YNRN.....H.....P
HUM_AMPK2	285NVI.....EA.....KE.....CEKF.....ECTES.....M.....YSGD.....PO.....Q
YSC_SNF1	332	KKDGSSPD.....NDEI.....NL.....NI.....SSTMG.....YE.....D.....Y.....ESSEDTPA.....NE
HUM_KP78	324DIS.....OKR.....DI.....VG.....G.....YSQE.....Q.....SKMK.....NE
MUS_EMK	321TTGPRDR.....DG.....NG.....HTE.....Q.....VGQR.....NE
CEL_PARI	438DOI.....QR.....EK.....IQ.....FQ.....LGFN.....AA.....L.....EKEK.....D
DM_SNF1B-1	345ELO.....AARF.....DL.....MS.....G.....HK.....R.....EQ.....KGQL.....D
DM_SNF1B-2	332DLH.....VNR.....SL.....SN.....G.....HKPR.....EQ.....KNQK.....D
MUS_MSK	346LGDYN.....QV.....GI.....QA.....A.....ID.....ORTI.....QNSS.....NH
RAT_SNRK	289LS.....EHNSI.....QR.....VLGDI.....AD.....DA.....VA.....ETNR.....NH
XEN_P69EG3	282YI.....DC.....TE.....SVFY.....KYS.....TSTTRL.....SEWS.....Y.....H
HUM_CAMKI	297FAKS.....KWKQAF.....NATAVVR.....HMRKLQ.....GT.....QEGQG.....Q
HUM_PKA-CA	309KV.....APF.....PKFKGPG.....DTSNFDDYEEEE.....IR

HUM_AMPK1	321AVA.....I.....N.....
HUM_AMPK2	319AVA.....H.....I.....N.....
YSC_SNF1	378RDA.....K.....N.....
HUM_KP78	356TAT.....GRKS.....
MUS_EMK	351MAT.....GYKS.....
CEL_PARI	473HAT.....GR.....
DM_SNF1B-1	377YCT.....GVA.....
DM_SNF1B-2	364YCA.....DVA.....
MUS_MSK	379FAAI.....Y.....L.....RL.....
RAT_SNRK	324TAT.....F.....A.....RI.....
XEN_P69EG3	314TAS.....L.....HSK.....
HUM_CAMKI	330TAS.....HGE.....TPVAGG
HUM_PKA-CA	337SI.....NEKCGK.....FS.....

In fact, in the case of AMPK, Hawley *et al.* (1996) have shown that this residue is phosphorylated by the AMPKK. The corresponding site in SNF1 is also necessary for activity (Estruch *et al.*, 1992).

3.3.1.2 Phylogenetic analysis of sequences closely related to *DM_SNF1B-1* and *DM_SNF1B-2*.

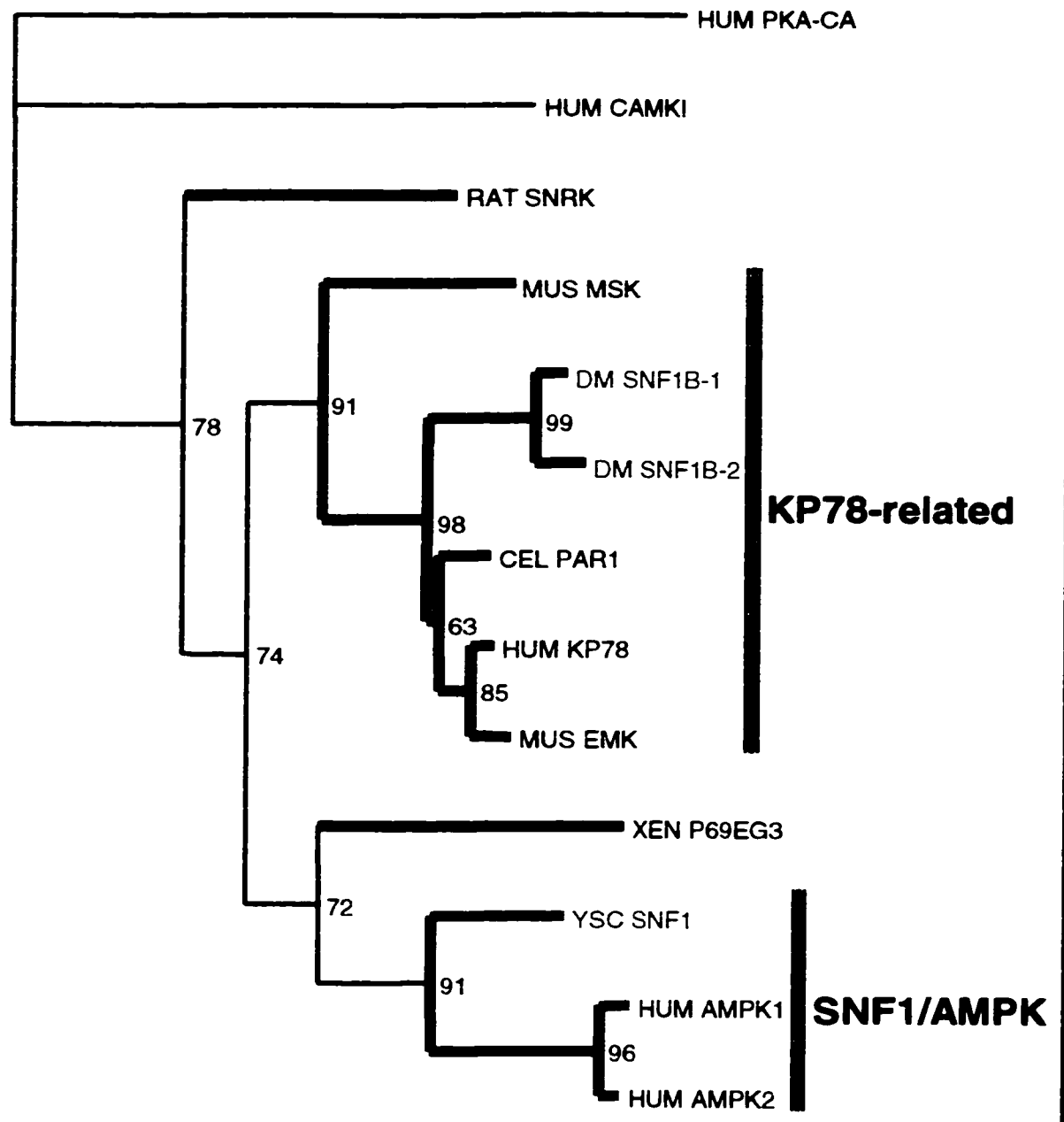
Previous analyses have shown that human *KP78* (Maheshwari *et al.*, unpublished) and a handful of other sequences [human *SNRK* (Becker *et al.*, 1996); mouse *EMK* (Inglis *et al.*, 1993); mouse *MSK* (Ruiz *et al.*, 1994); *PAR1* from *C. elegans* (Guo and Kemphues, 1995)] show moderately significant similarity levels to *SNF1* and its closest relatives [I will use the term “KP78-related kinases” when discussing these sequences]. Becker *et al.* (1996) have suggested that the KP78-related kinases share conservation of a small region flanking the C-terminus of the catalytic domain that is only found among “SNF1 related kinases” (the SNH domain, defined in the previous section). Becker *et al.* (1996) have also suggested that the SNH domain may represent a new type of regulatory domain of protein kinases.

DM_SNF1B-1 and *DM_SNF1B-2*, by virtue of their similarity to *KP78*, form part of the group of KP78-related kinases. We were thus interested in investigating the phylogenetic relationship between the different members of this group of kinases. As can be seen from Figure 13, not all of the different KP78-related kinases form part of a well-supported group. They also fail to group with the three AMPK/SNF1 kinases analysis (Fig 13, highlighted in green). The two *DM_SNF1B* genes cluster into a well-supported branch along with human *KP78*, mouse *EMK*, and *C. elegans PAR1* (Fig 13, highlighted in purple). Although the mouse *MSK* gene forms a sister branch to the KP78 branch (Fig 13, also in purple), the support for this relationship depends on the phylogenetic inference method used. The maximum-likelihood approach yields a moderate branch support of 89% (Fig 13), a protein distance method gives 99.8% support and maximum-parsimony gives 84.5% support [results not shown]. Interestingly, the rat *SNRK* (Fig 13, branch highlighted in

Figure 13. Maximum-likelihood distance tree derived from an alignment of kinases sharing similarity with *DM_SNF1B-1* and *DM_SNF1B-2*.

The protein sequence alignment shown in Figure 12 was used to generate a maximum-likelihood tree with branch lengths using the program PUZZLE 4.0.2 (Strimmer and von Haeseler, 1996 and 1997). Branch support is based on 10000 quartet-puzzling steps. Names of the sequences include the name of the protein kinase (when available, otherwise the GenBank accession number is given) and organism (HUM: human, MUS: mouse, RAT: rat, DM: *D. melanogaster*, CEL: *C. elegans*, YSC: yeast, Xen: *Xenopus laevis*). The HUM_PKA-CA represents the sequence for the human cAMP-activated protein kinase C- α subunit. HUM_CAMKI represents the human calcium/calmodulin-dependent protein kinase. These two sequences were used to provide an outgroup for the phylogenetic trees. Information regarding each sequence analyzed is shown in Appendix I, Table 3. Well-supported branches (KP78-related and SNF1/AMPK groups) and single taxa not belonging to well-supported branches (rat *SNRK* and *Xenopus P69EG3*) have been highlighted.

13.



black), on the basis of which Becker *et al.* (1996) have defined the SNH domain, does not form part of the main KP78 branch since support for the grouping is poor [63% using maximum-likelihood (Fig 13); 76.4% using protein distance (not shown); 23.3% using maximum-parsimony (not shown)]. *Xenopus P69EG3*, another “SNF1-related kinase” as defined by Becker *et al.* (1996), also fails to group with either the main KP78 branch and although it branches with the SNF1/AMPK kinases, it does so with weak support [72% using maximum-likelihood (Fig 13); 48.6% using protein distance (not shown); 17.8% using maximum-parsimony (not shown)].

3.3.2 DM_SNF1A: relationship to SNF1 and other close relatives.

DM_SNF1A shows extensive similarity to yeast *SNF1* and to the mammalian *AMPK- α* subunits. The approximately 92% similarity it shares with the *AMPKs* within the catalytic domain makes it a very likely homologue of the mammalian *AMPKs*. We wanted to gauge the conservation of other close *SNF1* relatives and, in particular, the two plant kinases that have been shown to be functional homologues of *SNF1*.

3.3.2.1 Protein sequence alignments. The sequence alignment shown in Figure 14 includes *SNF1*, *AMPK* and all their closest relatives as judged by overall sequence similarity. One kinase, the human cAMP-activated PKA-C α , was used as an outgroup for all others. Four other sequences (human, rat, *Drosophila* and *S. cerevisiae* Calcium/Calmodulin-dependent protein kinase I) were used to provide an example of a group of kinases for which conserved cellular role and regulation has been shown. Additionally, previous kinase phylogenies (Hanks and Hunter, 1995) have shown that the Calcium/Calmodulin-dependent protein kinases form a deeply branched sister group to the branch that leads to *SNF1* and related kinases.

As can be seen in Figure 14, exceptional levels of sequence similarity are observed among *SNF1*, *AMPK*, and closely related kinases included in the alignment, particularly within the

catalytic domain (position R258 in human Ampk- α 1). All of the sequences show conservation of the invariant catalytic domain residues and also share the conserved T residue in the activation loop (T173 in human Ampk- α 1). The bulk of differences among kinases occur beyond the catalytic domain and this is very evident from Figure 14. Nonetheless, it is still possible to see that despite the many insertions and deletions in their carboxy-terminal halves, there are many “islands” of conservation between the different kinases in the alignment and this makes it possible to align the sequences reasonably well across their entire length.

3.3.2.2 Phylogenetic analysis of closely related SNF1/AMPK sequences. An inspection of the sequence alignment shown in Figure 14 points to a great deal of sequence conservation among *SNF1*, *AMPK*, and closely related kinases. It is possible to align the sequences beyond the catalytic domain despite the fact that the non-catalytic domain of protein kinases is least conserved feature. Using the exact same data set, phylogenetic analysis was performed to elucidate the relationship between these different kinases.

As can be seen from the maximum-likelihood tree in Figure 15, the sixteen close relatives of *SNF1* kinases can be grouped into three major groups that cluster together with strong branch support (Fig 15). The yeast *SNF1* kinases from *S. cerevisiae* and *Candida albicans*, the animal kinases (*D. melanogaster*, *C. elegans*, rat and human) and the plant kinases each form well-supported groups (99%, 95%, and 95% branch support respectively). Although the clustering of all three major groups within this branch of kinases (Fig 15. shaded green) is well supported [85% by maximum likelihood (Fig 15), 100% by maximum-parsimony (not shown), and 99.9% by protein distance (not shown)], the relationship between the individual groups is less clear. Whereas maximum parsimony gives strong support to the yeast/animal grouping (94.1%, not shown), both distance and maximum-likelihood methods show a trifurcating node.

Figure 14. Protein sequence alignment of a group of sixteen highly related SNF1/AMPK-like protein kinases

All sixteen sequences were aligned along their full length in order to display sequence conservation beyond the core catalytic domain (R258 on human Ampk- α 1, green arrowhead). Sequences were aligned with the program CLUSTALW 1.74 (Thompson *et al.*, 1994) and formatted using BOXSHADE 2.7 (Kay Hoffman, unpublished). Identities are boxed in red and similarities are boxed in blue. The location of primers K1184 and K1187, which were used to search for additional copies of *DM_SNF1A* (Section 3.2.3.4), have been boxed in green and yellow, respectively. Sequence names include the descriptive name of the protein (when available, the accession number is used otherwise) and the organism (HUM: human, RAT: rat, DM: *D. melanogaster*, CEL: *C. elegans*, ATH: *A. thaliana*, TOB: tobacco, POT: potato, HOR: barley, RYE: rye, YSC: *S. cerevisiae*, CAN: *C. albicans*.) HUM_PKA-CA represents the sequence for the human cAMP-activated protein kinase C- α subunit and which is included to provide an outgroup for phylogenetic trees. Information regarding each sequence analyzed is shown in Appendix I.

```

HUM_AMPK1 234 Y T Q - - Y N P S V S S K H Q Q Q Q A - - - K K E E K Q D K R F P P E D S Y -
RAT_AMPK1 232 Y T Q - - Y N P S V S S K H Q Q Q Q A - - - K K E E K Q D K R S F P P E D S Y -
HUM_AMPK2 232 Y T E - - Y N R S V S T M H Q Q Q Q A - - - K K E E K Q D S A F P P E D S Y -
RAT_AMPK2 232 Y T E - - Y N R S I A T M H Q Q Q Q A - - - N K E E K Q D S A F P P E D S Y -
DM_SNF1A 244 P E E - - Y N K Q V V M C Q Q Q Q A - - - N K E E K Q D S A F P P S S I E Q -
CEL_U58726 241 P T D - - F E R P I V M E H Q Q Q Q A - - - A A - - - K K I A E S Q K D A M Y F P P I N E S -
CEL_PAR2.3 240 T T Y - - S E K Q A A S T Q Q Q Q A - - - D K R V M E S Q R I D Y F P P E C E N -
CAN_SNF1 269 T T N - - Y A G A H H T R I V I M M N I - - - H H M E D D K Q D D E L P P D L S -
YSC_SNF1 270 T T K - - F P G A G K R I I M M I - - - H H M Q D D K V D E R L P P D L K P H -
RYE_RKIN1 233 I I I - - Y D L V S S R R I I M M I - - - G H M K V S Q M R R C A V P P D Y -
HOR_BKIN12 236 I I S - - Y D S P K I M V I I - - - I I - - - R P V S P K M H C A V P P D -
HOR_BKIN2 214 T T S - - H P L P R P R V I V I - - - I I - - - R P P S S K A R R R A V P P D -
ATH_SKIN10 236 T T S - - H S E P R P R I V I V I - - - I I - - - R P Q Q Q Q O A H R R A V P P D -
ATH_AKIN10 235 T T S - - H P G P R P R I V I V I - - - V - - - P P P Q Q M P P O A H R R A V P P D -
TOB_NPK5 235 S T S - - H A G A P R P R I I I - - - M - - - P P P M L P P O A H R R A V P P D -
POT_SNF1 235 T L S - - H A G A P R P R I I I - - - M - - - P P P M L P P O A H R R A V P P D -
HUM_CAMKI 238 D S Y W D D S F R E E K E E F - - - C E Q A L Q P I A G D T A L - - - - - - -
RAT_CAMKI 238 D S Y W D D S F R R H E E K E E F - - - C E Q A L Q P I A G D T A L - - - - - - -
DM_CAMKI 264 D S Y W D E S K H F K N C T V E Y - - - C K Q A L G A I S G N E A S - - - - - - -
YSC_CAMKI 261 H R Y W D S N K K O F L K A N Q L T S P - - - A A L E D P I C T E L K - - - - - - -
HUM_PKA-CA 257 R F S - - H F S D L K S R N M Q L T S P G N L K N G M N M K A T T D W I A Y - - - - -

```

```

HUM_AMPK1 286 - - - S S T - - - - - M K K C E - - - S E E E L S
RAT_AMPK1 284 - - - S S T - - - - - M K K C E - - - S E E E L S
HUM_AMPK2 284 - - - D A N - - - - - V K K C E - - - T E S E M N
RAT_AMPK2 284 - - - D A N - - - - - V A K C E - - - C T E S E M N
DM_SNF1A 296 - - - D S N - - - - - I T Y A A R C T - - - G V K E E T H N
CEL_U58726 293 - - - E A S - - - - - I I A A R T E - - - H V A E E T T S
CEL_PAR2.3 291 - - - E S S - - - - - I I Q S A E T P P E K I I Y F F I F R H F L K V K E E T G
CAN_SNF1 320 - - - K N K N - - S K I D - - - - - I I R A S V T M G - - - D C R V N G
YSC_SNF1 323 P E E E N E N N D S K K D G S S P D N D E M N I S S T M G - - - D - E Y E
RYE_RKIN1 284 M M Q Q A K M - - - - - T R - - - V L G - - - D - H C E
HOR_BKIN12 287 K A P K A K M - - - - - T R - - - V N L G - - - D - H C E
HOR_BKIN2 265 T A Q Q V K K - - - - - T M Q - - - I M G - - - M - Q L E
ATH_SKIN10 287 T V E Q T K K - - - - - M Q Q - - - V M N G - - - M - Q L E
ATH_AKIN10 286 T V Q Q A K K - - - - - Q Q Q - - - I M N G - - - M - H I E
TOB_NPK5 286 T M Q Q A K K - - - - - Q Q Q - - - V R G - - - M - N S V A
POT_SNF1 286 T M Q Q A K K - - - - - Q Q Q - - - V M G - - - M - N T E
HUM_CAMKI 283 - - - D K N - - - - - H Q S S E Q - K M - - - A K S - - -
RAT_CAMKI 283 - - - D K N - - - - - H Q S V S E Q - K M - - - A K S - - -
DM_CAMKI 309 - - - S R N - - - - - H G T V S E Q - K M - - - A K S - - -
YSC_CAMKI 306 - - - T H N - - - - - L P G L K E G - D A Q K - - - R N S - - V E
HUM_PKA-CA 308 - - - Q R K - - - - - T A P F P K F K G P G D - - - T S N - - F D

```

```

HUM_AMPK1 312 C Y N - - - - - R N H D P A E R - - - R I - - - - M
RAT_AMPK1 310 C Y N - - - - - R N H D P A E R - - - R I - - - - M
HUM_AMPK2 310 Y S - - - - - G D P D Q A E R - - - R I - - - - M
RAT_AMPK2 310 Y S - - - - - G D P D Q A E R - - - R I - - - - M
DM_SNF1A 322 L S - - - - - G D P H D Q A E R - - - K F A D D - A A N - Q I
CEL_U58726 319 A L G - - - - - D D P H H S E M - - - K I A D E - T - A K
CEL_PAR2.3 334 A L A - - - - - E D H H F C E R E N H K N - - - A D E S S Q
CAN_SNF1 351 V E K A N K Q V A A G N S - - S S Q S K S S N E L D R L K N H A L V K D L K - K - - S K S
YSC_SNF1 366 E - - - - - S S E D T P A F N E R D M K K S L I K D M K A N K - - - S V S
RYE_RKIN1 312 C - - - - - N R L N E E T Y Y - - - F R - - - A T S
HOR_BKIN12 315 W - - - - - N R L N E E T Y Y - - - F R - - - S T S
HOR_BKIN2 293 Q - - - - - K R L N E A T Y Y - - - L R - - - T T S
ATH_SKIN10 315 R - - - - - N R T N D A T T Y Y - - - F R - - - V P S
ATH_AKIN10 314 R - - - - - N R T N D G T T Y Y - - - F R - - - A S S
TOB_NPK5 314 C - - - - - N R V N E G T Y Y - - - Q F R - - - A S S
POT_SNF1 314 R - - - - - N R V N E G T P Y Y - - - H R - - - V S T
HUM_CAMKI 302 - - - - - K W K Q A F N T A R R - - - - - H M
RAT_CAMKI 302 - - - - - K W K Q A F N T A R R - - - - - H M
DM_CAMKI 328 - - - - - R W K Q A Y Y A T R R - - - - - Q M
YSC_CAMKI 329 R R R - - - - - L M K Q K L R D Y - - - - - T E S D
HUM_PKA-CA 330 D Y E - - - - - E E E R S I N - - - - - -

```

```

HUM_AMPK1 337 N E K D F L A S P P D - - - - - S F L D D - - - - - H H L T - -
RAT_AMPK1 335 N E K D F L A S P P D - - - - - S F L D D - - - - - H H L T - -
HUM_AMPK2 335 N Q A S E F L A S P P S G - - - - - S F M D D S A - - - - - M H I P P G
RAT_AMPK2 335 N Q A S E F L A S P P T G - - - - - S F M D D S A - - - - - M H I P P G
DM_SNF1A 354 N E N N F V A G S P P P P P P V - - - P Q S S M D H Q A P L A T V T V G G G T S A S S G T A - - - T P V P P V
CEL_U58726 349 L S E E F Q V P N K G - - - - - - - - - - - - - - - P G P
CEL_PAR2.3 364 K A E D F E I G - - - - - - - - - - - - - - - K T N K M G S - - - - - T S L P V G
CAN_SNF1 398 E N E S F L S Q P P - - - - - - - - - - - - - - - P S P P F N R G S T - - - - - S S A P G V
YSC_SNF1 404 D E D T F L S Q P P T F Q Q - - - - - - - - - - - - - - - Q S K S H Q S Q V D H E T A K - - - - - Q H A R R M
RYE_RKIN1 338 G Y G A H Q Q P M - - - - - - - - - - - - - - - - - - - - - R S F N E F
HOR_BKIN12 341 G Y G A D H Q H L M D - - - - - - - - - - - - - - - - - - - - - S S F S Q I
HOR_BKIN2 319 G Y G A E Q E M D - - - - - - - - - - - - - - - - - - - - - S G S M P M
ATH_SKIN10 341 G Y E S E Q E T D - - - - - - - - - - - - - - - - - - - - - G T P R M
ATH_AKIN10 340 G Y G A E Q E M E - - - - - - - - - - - - - - - - - - - - - Y G Y H Q I
TOB_NPK5 340 G Y G A E Q E M E - - - - - - - - - - - - - - - - - - - - - Y G Y M R I
POT_SNF1 340 G Y G A E Q E M E - - - - - - - - - - - - - - - - - - - - -
HUM_CAMKI 317 R K Q L G T S Q E G - - - - - - - - - - - - - - - - - - - - -
RAT_CAMKI 317 R K L Q L G T S Q E G - - - - - - - - - - - - - - - - - - - - -
DM_CAMKI 343 Q R M A L N S N S N A N F - - - - - - - - - - - - - - - - - - - - -
YSC_CAMKI 351 S D F D E G S Q A N G - - - - - - - - - - - - - - - - - - - - -
HUM_PKA-CA 342 E K C G K E S E F - - - - - - - - - - - - - - - - - - - - -

```

```

HUM_AMPK1 360 -----RPHPRVPPFLAETPRRHTLDLDELN-----PQESKHQG
RAT_AMPK1 358 -----RPHPRVPPFLAETPRRHTLDLDELN-----PQESKHQG
HUM_AMPK2 363 L-----KPHPRMPPLADSPKRCPLDALN-----TTPKSLA
RAT_AMPK2 363 L-----KPHPRMPPLADSPKRCPLDALN-----TTPKSLA
DM_SNF1A 407 AGGTPSSSTIP-----IRPHPRIAAASRDRQLMSVQTSGGG--AFPEKTARGG
CEL_U58726 366 -----VHRHPRIAAASRDRQLMSVQTSGGG--AFPEKTARGG
CEL_PAR2.3 387 ATTK-----SEKSRNVAKVVGKPSNVGRRKILEG--LAKKEQ
CAN_SNF1 426 QQSLTYQ---TLAT-VPDLSTLPMSTIAIPTSLSIHRAYMAETK---QNGDPSQ
YSC_SNF1 442 ASAITQORTYHQSP--FMDQYKEESTVSIPTSLSIHRAYMAETK---QNGDPSQ
RYE_RKIN1 349 -----SASPSTRSYLPMSNDSQGGG--LSP
HOR_BKIN12 359 TLS-----SASPSTRNYLPMSNDSQGGG--LSP
HOR_BKIN2 337 SPE-----T-----PSSASEARQYGSPLFGLRQ--LSP
ATH_SKIN10 359 RTP-----ACASPGGHWIPSEVDHYGLG--LSP
ATH_AKIN10 357 HPA-----SVASPSHRLPLMEYQGGV--LSP
TOB_NPK5 358 NSS-----VLLPCWQHLPLIMDFQQVG--LSP
POT_SNF1 358 NSN-----PLLPWQRFPLIMDYQAG--LSP
HUM_CAMKI 328 -----QGQTASHGELLTPVAGGPAAG--LSP
RAT_CAMKI 328 -----QGQTASHGELLTPVAGGPAAG--LSP
DM_CAMKI 356 -----SSNSSMQDSTTPTAATGAMTS--LSP
YSC_CAMKI 362 -----SVPPKATDTSQLSKKLSE--LSP
HUM_PKA-CA -----LSP

```

```

HUM_AMPK1 393 -----VRKAHRRRQ-SRNYAAACQDYEVM-
RAT_AMPK1 391 -----VRKAHRRRQ-SRNYAAACQDYEVM-
HUM_AMPK2 397 -----VKKAHRRRQ-SRNYAAACQDYEVM-
RAT_AMPK2 397 -----VKKAHRRRQ-SRNYAAACQDYEVM-
DM_SNF1A 453 T-----PIKRAHRRRQ-SRNYAAACQDYEVM-
CEL_U58726 397 -----KRAHRRRQ-SRNYAAACQDYEVM-
CEL_PAR2.3 422 -----KKLTHRRRQ-SRNYAAACQDYEVM-
CAN_SNF1 475 -QH-----APPPTKKSKTHRRRQ-SRNYAAACQDYEVM-
YSC_SNF1 493 ISP-----LVTKKSKTHRRRQ-SRNYAAACQDYEVM-
RYE_RKIN1 373 -----YRVERAARRRQ-SRNYAAACQDYEVM-
HOR_BKIN12 386 -----YYPVQRAARRRQ-SRNYAAACQDYEVM-
HOR_BKIN2 360 -----HFAAERAARRRQ-SRNYAAACQDYEVM-
ATH_SKIN10 386 -----QVPVDRARRRQ-SRNYAAACQDYEVM-
ATH_AKIN10 384 -----QYPVERARRRQ-SRNYAAACQDYEVM-
TOB_NPK5 383 -----QFPVERARRRQ-SRNYAAACQDYEVM-
POT_SNF1 383 -----QFPVERARRRQ-SRNYAAACQDYEVM-
HUM_CAMKI 349 -----CCCRDCCGTE-SPTLP-HQDYEVM-
RAT_CAMKI 349 -----CCCRDCCGTE-SPTLP-HQDYEVM-
DM_CAMKI 378 -----NVLSSQSSHAQENKSGGSTNAASQDYEVM-
YSC_CAMKI 383 -----QSLKSEHT-KAFAPQVNV-TVAEKEFEN-
HUM_PKA-CA -----LSP

```

```

HUM_AMPK1 431 P Y V RK-NPVTSTYSKSLQYQVD-----SRTLLFRSDDITEAK-SGTATPQ
RAT_AMPK1 429 P Y V RK-NPVTSTFSKSLQYQVD-----SRTLLFRSDDITEAK-SGTATPQ
HUM_AMPK2 435 A H V RK-NPVTGNYVKSLQYLVD-----NRSLLFKSDDVVEQR-SGSSTPQ
RAT_AMPK2 435 A H V RK-NPVTGNYVKSLQYLVD-----NRSLLFKSDDVVEQR-SGSSTPQ
DM_SNF1A 493 P H V RQ-NVKTGKFSKSLQYQVD-----AKSLLFKSADVEVEQ--G--
CEL_U58726 434 P H IV RKPDAFAADPFKSLQYQVD-----QRSLLFKSADVEVEQ--G--
CEL_PAR2.3 459 M H IV SKPTPINPDVRSKSLQFALDKKEN-NKGLLFKGTEZEAVPPSRRCSRA
CAN_SNF1 524 L T V WKYDTS-AQFECGS-APMLM-----KMQIQLFQPNNY-
YSC_SNF1 541 L T L WKYDIG-NKTMTNEKPDLM-----KMVIQLFQPNNY-
RYE_RKIN1 415 C M C WCPGFP-RVSDMLDANHS-----VDCAIKNGDAN-
HOR_BKIN12 426 L M C WCPGFP-QVSAMLDSEHN-----VDSTIANGAD-
HOR_BKIN2 400 H M C WSPGFP-ENMHN-----GGVSAIANGL-
ATH_SKIN10 426 H M C WVPGLADGQNTMHNQH-----GRDSSIDCA-
ATH_AKIN10 424 H M C WVPMS--ADGMSHSDMN-----YGDSSIIMAA-
TOB_NPK5 423 P H C QWVPGVPGHHEGMSNNSH-IQ-----FQDSTVMGG-
POT_SNF1 423 Q M C WVPVSVPGHHEGMVNSHGNQ-----FQDSSIIMGG-
HUM_CAMKI -----LSP
RAT_CAMKI -----LSP
DM_CAMKI -----LSP
YSC_CAMKI 412 -INRVCSSDSDLP-GSDIKSLDEAKE-----KPGKDTKTZE-----LSP
HUM_PKA-CA -----LSP

```

```

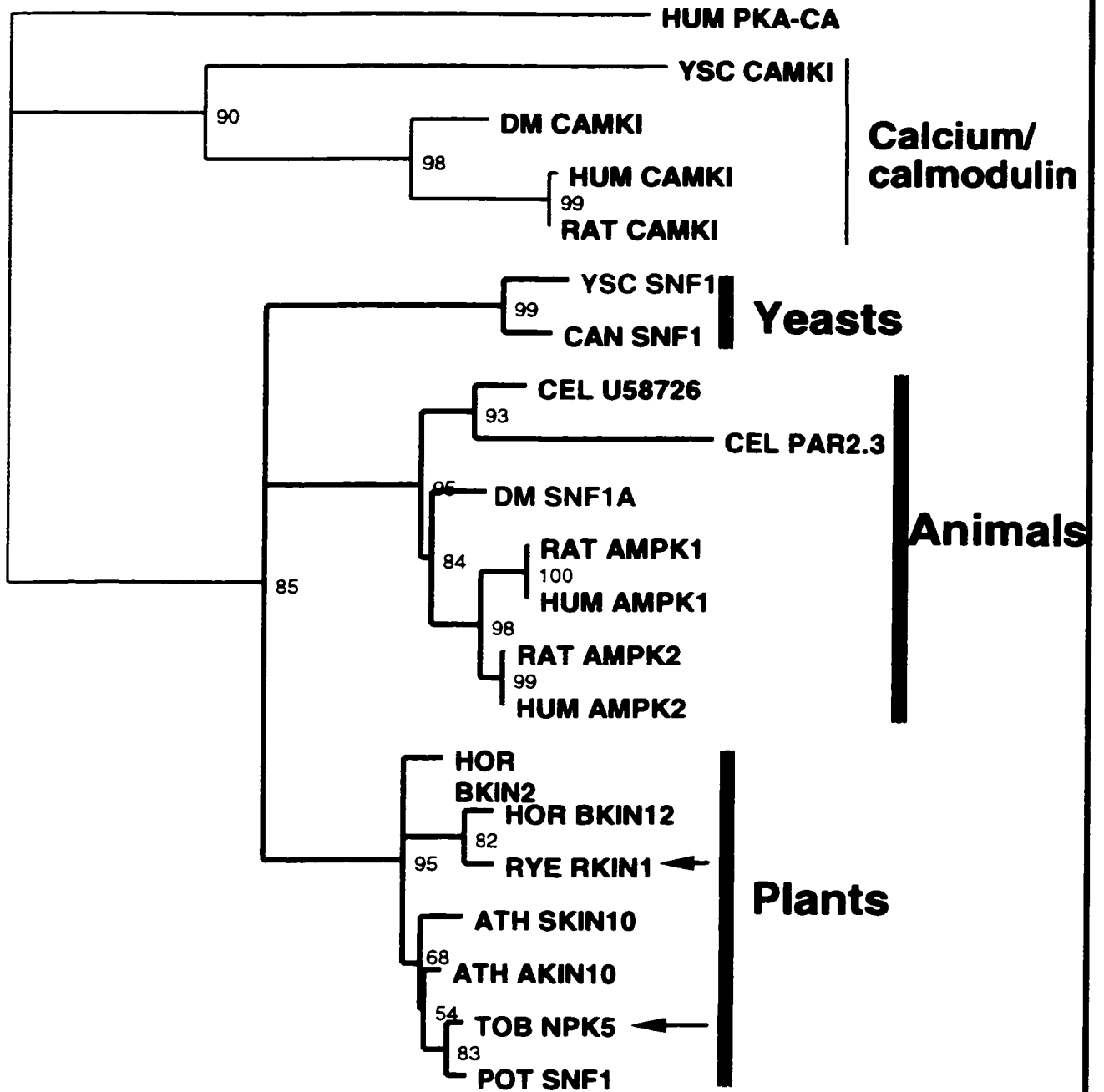
HUM_AMPK1 484 RSGSVSNYRS-----CQRSDSDAEAQGKS--S V S
RAT_AMPK1 482 RSGSISNYRS-----CQRSDSDAEAQGKP--S V S
HUM_AMPK2 488 RSCSAAGLHR-----FRSSFDSTAESHS--LSGS
RAT_AMPK2 488 RSCSAAGLHR-----PRSSVDSSTAENHS--LSGS
DM_SNF1A 539 -----DDVIME-----S
CEL_U58726 488 ASNSMPQKPA-----IRGTRTSSMPQAMS--M AS
CEL_PAR2.3 518 ASVSVTLAKSKSDLNGNSSK---VPMSPSPSPISPSVNIKPVRTDADASLKSBLNSS
CAN_SNF1 563 -----L-----VSFKFSG--W SA
YSC_SNF1 581 -----L-----VDFKFDG--W SS
RYE_RKIN1 454 -----S-----RLPAVVK--F IQ
HOR_BKIN12 465 -----G-----RSTHIVK--F IQ
HOR_BKIN2 436 -----D-----RSTHIVK--F IQ
ATH_SKIN10 464 -----M-----TSPTVIK--F IQ
ATH_AKIN10 464 -----V-----RSPVVK--F IQ
TOB_NPK5 463 -----V-----TIPNAVK--F IQ
POT_SNF1 465 -----T-----KLTNVVK--F IQ
HUM_CAMKI -----LSP
RAT_CAMKI -----LSP
DM_CAMKI -----LSP
YSC_CAMKI -----LSP
HUM_PKA-CA -----LSP

```


Figure 15. Maximum-likelihood distance tree derived from an alignment of SNF1/AMPK-like protein kinases.

The protein sequence alignment shown in Figure 14 was used to generate a maximum-likelihood tree with branch lengths using the program PUZZLE 4.0.2 (Strimmer and von Haeseler, 1996 and 1997). Branch support is based on 10000 quartet-puzzling steps. Names of the sequences include the name of the protein kinase (when available, otherwise the GenBank accession number is given) and organism (ATH: *A. thaliana*, CAN: *C. albicans*, CEL: *C. elegans*, DM: *D. melanogaster*, HOR: barley, HUM: human, MUS: mouse, POT: potato, RAT: rat, RYE: rye, TOB: tobacco, YSC: *S. cerevisiae*). The HUM_PKA-CA represents the sequence for the human cAMP-activated protein kinase C- α subunit and which was included as an outgroup. Information regarding sequences analyzed is shown in Appendix II. The well-supported branch of SNF1/AMPK kinases has been highlighted in green. The bars indicate three distinct sub-groups corresponding to the SNF1/AMPK kinases of yeasts, animals, and plants, as well as the group of Calcium/Calmodulin kinases. The arrows point to two plant kinases that have been shown to complement *snf1* mutants.

15.



3.3.3 Relationship between SNF1, AMPK and other SnRKs.

The human KP78 kinase has often been considered a sister group to the group composed of *SNF1* and its closest relatives because of its lower overall similarity to this group of kinases. Similarly, a number of plant “SNF1-related kinases” or SnRKs of unknown role have been cloned, submitted to GenBank and annotated as such despite having lower similarity values to *SNF1* than well-established plant homologues such as tobacco *NPK5* and rye *RKIN1*. In order to establish the relationship between some of the different protein kinase sequences that have been annotated as “SNF1-related kinases” in the various molecular sequence databases, these sequences were subjected to phylogenetic analysis along with *SNF1*, *AMPK* and their closest relatives.

3.3.3.1 Protein sequence alignments. The alignment of the thirty-six sequences included in this survey is shown in Figure 16. Thirty-one kinases either represent well known *SNF1/AMPK* homologues, sequences sharing high levels of similarity with human *AMPKs* and *SNF1*, or sequences sharing a lower level of similarity and which have been annotated in database entries as “similar to SNF1”, “SNF1-related”, or “similar to SNF1-related sequences”. One kinase, the human cAMP-activated PKA-C α , was used as an outgroup for all the other kinases. As in the analysis discussed in section 3.3.2, four other kinases (human, rat, *Drosophila* and *S. cerevisiae* Calcium/Calmodulin-dependent protein kinase I) were used to provide an example of a group of kinases for which conserved cellular role and regulation has been shown.

The divergence observed in the non-catalytic domain of some of the kinases included in the alignment precludes any meaningful alignment beyond the carboxy-termini of the respective catalytic domains. Nevertheless the catalytic-domain alignment shows great conservation and is easily alignable.

3.3.3.2 Phylogenetic analysis of the SNF1/AMPK and SNF1-related kinases. We used the alignment on Figure 16 to perform phylogenetic analysis on a group of kinases that includes the well-known SNF1/AMPK kinases and also includes sequences that are less similar to SNF1 but which have been termed “SNF1-related” in the various databases. Although the early branches are poorly resolved, it is possible to see that all of the kinases that were originally included in the previous analysis (Fig 14 and 15), the family of close SNF1/AMPK relatives, form a single monophyletic group (Fig 17, highlighted in green) with excellent support. The overall branching pattern and bootstrap values observed within the SNF1/AMPK group are consistent with those observed previously (Fig 15).

Notably, all of the other SNF1-related protein kinases included in this survey fall outside of the major SNF1/AMPK branch. Nonetheless, groupings seen in earlier analyses can still be seen in the tree shown in Figure 17. These include the SnRK2 group of plant kinases (Fig 17, highlighted in red) and the SnRK3 group of plant kinases (Fig 17, highlighted in yellow). Thus the tree based on the larger data set is consistent with the trees obtained with the smaller data sets. One notable difference is observed for the KP78-related kinases (Fig 17, highlighted in purple). In this maximum-likelihood tree, the group is poorly resolved and collapses into a multiforked node that includes rat *SNRK* (highlighted in black) and *P69EG3* (highlighted in blue) [however the support for this branching is only 51% so this node could itself be collapsed]. The topology of this region of the tree is different than what was seen in the maximum-likelihood tree of Figure 13. Similarly, protein distance and maximum-parsimony methods yield topologies that are consistent with the topology of the tree shown in Figure 13. In these trees there is good support for the KP78-related group of kinases as a separate group from rat *SNRK* and *Xenopus P69EG3*. These trees also show that Rat *SNRK* and *Xenopus P69EG3* fail to group with any other well-supported branches. Since

the number of taxa analyzed for Figure 17 is larger and involves a shorter sequence alignment,, it is likely that the ambiguity observed is the result of an insufficient number of informative sites.

The deep branches of the tree shown in Figure 17 lack statistical support, indicating that the relationship between any of the groups of kinase cannot be reliably inferred. The various kinase groups come together in a multiforked node, as was observed in the tree shown in Figure 13. It is very apparent that none of the groups of SNF1-related kinases included in this analysis actually shares a close relationship with the SNF1/AMPK group of kinases. Grouping the KP78-related kinases, the SnRK2 kinases, the SnRK3 kinases, rat *SNRK*, and *Xenopus P69EG3* with the SNF1/AMPK group of kinases appears to be artificial because the deep node is not only multiforked, but because the support for the branch as a whole is only 54%.

Figure 16. Protein sequence alignment of the catalytic domains of thirty-one SNF1/AMPK-related protein kinases.

All sequences were aligned to the amino terminus of the kinase catalytic domain (R258 on human Ampk- α 1) for subsequent phylogenetic analysis. Sequences were aligned with the program CLUSTALW 1.74 (Thompson *et al.*, 1994) and formatted using BOXSHADE 2.7 (Kay Hoffman, unpublished). Identities are boxed in red and similarities are boxed in blue. The sequences are named to include the descriptive name of the protein kinase (when available, the GenBank accession number is used otherwise) and the organism (ATH: *A. thaliana*, CAN: *C. albicans*, CEL: *C. elegans*, DM: *D. melanogaster*, HOR: barley, HUM: human, MUS: mouse, POT: potato, RAT: rat, RYE: rye, TOB: tobacco, WHT: wheat, XEN: *Xenopus laevis*, YSC: *S. cerevisiae*). HUM_PKA-CA represents the sequence for the human cAMP-activated protein kinase C- α subunit and which is included to provide an outgroup for phylogenetic trees. The CAMKI protein kinases represent Calcium/Calmodulin-dependent protein kinases. Information regarding sequences analyzed is shown in Appendix I.

HUM_AMPK1	199	S	A	H	T	---	CD	I	Y	Q	---	N	P	S																							
RAT_AMPK1	197	S	A	H	T	---	CD	I	Y	Q	---	N	P	S																							
HUM_AMPK2	197	A	A	H	T	---	RG	V	Y	E	---	Y	L	N																							
RAT_AMPK2	197	A	A	H	T	---	RG	V	Y	E	---	Y	L	N																							
DM_SNF1A	209	I	A	H	T	---	KS	I	F	P	I	E	---	Y	L	N	K	Q																			
CEL_U58726	206	A	A	H	T	---	KS	V	F	P	T	D	---	R	L	E	R	P																			
CEL_PAR2.3	205	A	A	H	T	---	KS	R	T	W	Y	---	S	H	E	K	Q																				
YSC_SNF1	235	V	V	R	R	---	SN	V	T	K	---	S	L	P	G																						
CAN_SNF1	234	A	V	R	R	---	SN	V	T	L	---	N	---	Y	L	A	G																				
TOB_NPK5	200	A	A	H	T	---	KG	N	I	S	---	H	L	A	G																						
POT_SNF1	200	A	A	H	T	---	KG	Y	I	T	---	H	L	A	G																						
ATH_AKIN10	200	A	A	H	T	---	KG	I	T	L	---	H	L	A	G																						
ATH_SKIN10	201	A	A	H	T	---	KG	I	T	L	---	H	L	A	G																						
HOR_BKIN2	179	A	A	H	T	---	KG	I	T	L	---	H	L	A	G																						
HOR_BKIN12	201	A	A	H	T	---	KG	T	I	L	---	Y	L	D	S																						
RYE_RKIN1	198	A	A	H	T	---	KG	T	I	L	---	I	---	Y	L	D	L																				
ATH_ASK1	183	T	V	V	A	Y	Q	E	D	K	N	R	T	I	Q	K	M	A	V	Q	K	I	D	Y	V	H	I	Q	D								
ATH_ASK2	183	T	V	V	A	Y	P	D	K	N	R	T	I	Q	R	M	A	V	Q	K	I	D	Y	V	H	I	Q	E									
WHT_PKABA1	173	T	V	V	A	Y	P	D	E	R	M	T	I	T	R	L	S	V	Q	S	D	Y	V	R	V	M	D										
ATH_S56718	202	T	V	V	A	Y	P	E	R	D	R	T	I	Q	R	L	S	V	T	S	E	D	L	H	L	P	E										
ATH_S71172	201	T	V	V	A	Y	P	E	R	D	R	T	I	Q	R	L	S	V	K	S	I	D	D	I	R	I	P	E									
WHT_WPK4	230	I	F	V	N	A	Y	H	Q	L	N	A	N	Y	---	V	Y	K	E	F	R	C	R	---	W	F	K	D									
ATH10237.1	206	A	F	V	N	A	Y	H	Q	L	N	A	N	Y	---	V	Y	K	E	F	R	C	R	---	W	F	K	D									
HUM_KP78	236	L	T	V	S	S	---	G	Q	L	K	E	R	E	R	---	V	L	R	K	R	I	F	---	Y	M	T	D									
MUS_EMK	233	L	T	V	S	S	---	G	Q	L	K	E	R	E	R	---	V	L	R	K	R	I	F	---	Y	M	T	D									
CEL_PAR1	350	L	T	V	S	S	---	G	Q	L	K	E	R	E	R	---	V	L	R	K	R	I	F	---	Y	M	T	D									
DM_SNF1B-1	256	A	L	T	V	S	---	G	G	T	L	K	E	R	E	R	---	V	L	R	K	R	V	Y	---	Y	I	M	D								
DM_SNF1B-2	243	S	L	T	V	S	---	G	T	L	K	E	R	D	R	---	V	L	R	K	R	V	Y	---	Y	I	E	I									
MUS_MSK	253	L	V	V	V	S	---	G	P	L	T	R	Q	R	---	V	L	E	R	F	R	I	F	---	F	M	Q	D									
RAT_SNRK	198	I	L	P	M	V	Q	P	O	E	A	D	S	E	T	L	T	M	---	M	D	C	K	T	V	P	---	R	V	A	G						
XEN_P69EG3	194	I	M	L	N	A	Y	Q	P	E	D	V	M	Y	K	---	M	R	K	E	I	K	---	W	L	P	G										
HUM_CAMKI	201	C	I	A	I	Y	P	Y	Y	D	A	K	E	Q	---	I	L	K	A	E	Y	E	D	S	Y	W	D	D	I	D	S						
RAT_CAMKI	201	C	I	A	I	Y	P	Y	Y	D	A	K	E	Q	---	I	L	K	A	E	Y	E	D	S	Y	W	D	D	I	D	S						
DM_CAMKI	227	I	I	A	I	Y	P	Y	Y	D	A	K	E	Q	---	I	L	K	A	E	Y	E	D	S	Y	W	D	D	I	D	S						
YSC_CAMKI	222	I	T	T	A	Y	S	A	R	A	R	Q	D	F	L	D	E	C	T	G	E	Y	G	P	V	K	H	R	S	---	Y	W	D	S	I	N	K
HUM_PKA-CA	222	W	A	L	T	E	H	A	Y	P	F	A	Q	P	I	Q	I	Y	---	V	S	K	V	R	F	S	---	H	F	S	D						

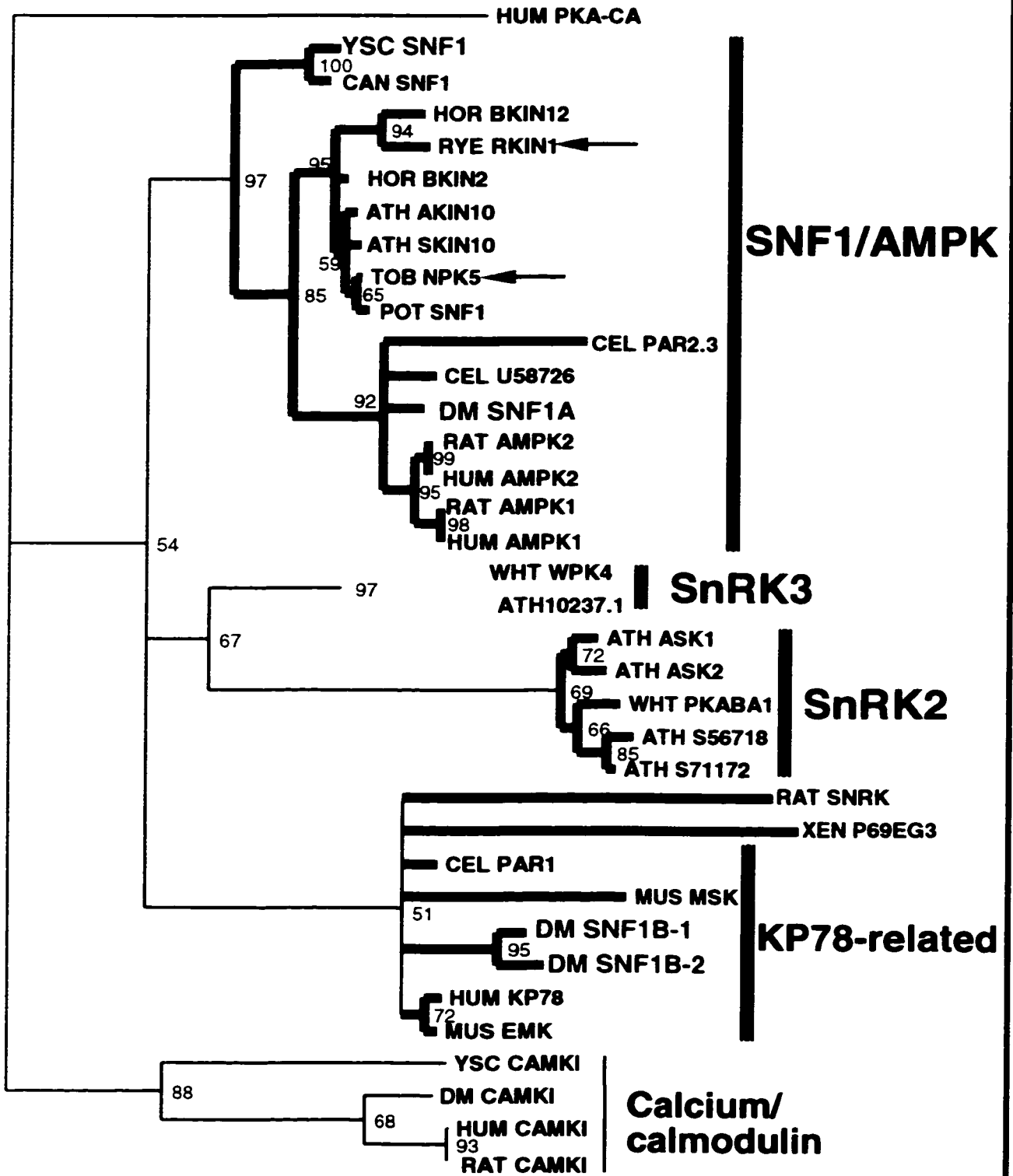


HUM_AMPK1	243	V	T	S	L	K	H	Q	Q	M	A	---	S	K	D	R	E	E						
RAT_AMPK1	241	V	I	S	L	K	H	Q	Q	M	A	---	S	K	D	R	E	E						
HUM_AMPK2	241	V	A	T	L	M	H	Q	Q	L	A	---	S	K	D	R	E	E						
RAT_AMPK2	241	I	A	T	L	M	H	Q	Q	L	A	---	S	K	D	R	E	E						
DM_SNF1A	253	V	V	N	V	C	Q	Q	Q	L	A	---	N	E	K	K	E	E						
CEL_U58726	250	I	V	N	L	H	H	C	C	M	A	---	S	K	D	V	I	A	E					
CEL_PAR2.3	249	A	A	D	S	T	Q	Q	Q	M	A	---	D	W	K	R	V	N	S					
YSC_SNF1	279	A	A	G	K	M	K	M	L	N	I	---	S	H	M	Q	D	D						
CAN_SNF1	278	A	A	H	L	T	R	V	M	L	N	I	---	S	H	M	E	D	D					
TOB_NPK5	244	A	R	D	P	R	I	I	M	M	---	P	R	M	P	P	P	P						
POT_SNF1	244	A	R	D	P	R	I	I	M	M	---	P	R	M	P	P	P	P						
ATH_AKIN10	244	A	R	D	P	R	I	I	M	V	---	P	R	Q	P	P	P	P						
ATH_SKIN10	245	A	R	D	P	R	I	I	V	I	---	P	R	Q	R	P	P	P						
HOR_BKIN2	223	A	R	D	P	R	I	I	M	I	---	P	R	R	E	S	P	P						
HOR_BKIN12	245	A	R	D	P	R	I	I	M	I	---	P	R	R	E	S	P	P						
RYE_RKIN1	242	V	R	D	S	R	I	I	M	I	---	P	R	R	E	S	P	P						
ATH_ASK1	233	C	K	N	L	S	R	I	F	V	A	N	S	L	I	---	A	K	K	S				
ATH_ASK2	233	C	K	H	L	S	R	I	F	V	T	N	S	A	I	---	L	K	K	N	P			
WHT_PKABA1	223	C	H	L	S	R	I	F	V	G	M	Q	Q	I	---	P	K	N	P					
ATH_S56718	252	C	H	L	S	R	I	F	V	A	A	T	I	---	P	T	S	D	K					
ATH_S71172	251	C	H	L	S	R	I	F	V	A	A	T	I	---	P	T	S	D	K					
WHT_WPK4	274	L	T	S	M	R	F	D	T	N	S	T	I	---	S	L	P	V	N	E	S	R		
ATH10237.1	250	A	R	S	I	Y	K	L	D	P	N	E	T	M	---	S	L	E	A	V	N	G	T	V
HUM_KP78	280	C	E	N	L	K	R	F	V	L	N	I	G	---	S	L	E	Q	N	K	D	R		
MUS_EMK	277	C	E	N	L	K	R	F	V	L	N	I	G	---	S	L	E	Q	N	K	D	R		
CEL_PAR1	394	C	E	N	L	K	R	F	V	L	N	I	G	---	S	L	E	Q	N	K	D	R		
DM_SNF1B-1	300	C	E	N	M	R	R	F	V	L	N	A	T	---	S	L	S	A	V	N	S	D	K	
DM_SNF1B-2	287	C	E	S	V	R	R	F	V	L	N	T	O	---	S	L	S	A	V	N	A	D	R	
MUS_MSK	297	C	E	T	R	R	V	A	I	---	S	L	S	A	V	N	S	D	K					
RAT_SNRK	242	C	H	L	S	R	I	F	V	T	N	S	A	I	---	P	K	N	P					
XEN_P69EG3	238	S	V	L	S	Q	Q	Q	K	I	---	S	L	S	A	V	N	S	D	K				
HUM_CAMKI	249	A	K	D	F	R	H	L	M	E	K	E	F	---	C	E	Q	A	L	Q	P			
RAT_CAMKI	249	A	K	D	F	R	H	L	M	E	K	E	F	---	C	E	Q	A	L	Q	P			
DM_CAMKI	275	A	K	H	F	K	N	L	M	C	T	V	E	Y	---	C	K	Q	A	L	Q	P		
YSC_CAMKI	272	A	K	H	F	L	K	A	N	L	S	F	P	---	A	A	L	L	E	D	P			
HUM_PKA-CA	266	L	K	D	E	R	N	L	Q	L	T	F	G	N	L	K	N	G	V	N	D	K	N	R

Figure 17. Maximum-likelihood distance tree derived from an alignment of thirty-one SNF1/AMPK-related protein kinases.

The protein sequence alignment shown in Figure 16 was used to generate a maximum-likelihood tree with branch lengths using the program PUZZLE 4.0.2 (Strimmer and von Haeseler, 1996 and 1997). Branch support is based on 10000 quartet-puzzling steps. Names of the sequences include the name of the protein kinase (when available, the GenBank accession number is given as name otherwise) and organism (ATH: *A. thaliana*, CAN: *C. albicans*, CEL: *C. elegans*, DM: *D. melanogaster*, HOR: barley, HUM: human, MUS: mouse, POT: potato, RAT: rat, RYE: rye, TOB: tobacco, YSC: *S. cerevisiae*). The HUM_PKA-CA represents the sequence for the human cAMP-activated protein kinase C- α subunit. HUM_CAMKI represents the human calcium/calmodulin-dependent protein kinase. These two sequences were used to provide an outgroup for the phylogenetic trees. Information regarding each sequence analyzed is shown in Appendix I. Well-supported branches (SNF1/AMPK, SnRK2, SnRK3, KP78-related groups) and single taxa not belonging to well-supported branches (rat *SNRK* and *Xenopus P69EG3*) have been highlighted in colour. The arrows point to two plant kinases that have been shown to complement *snf1* mutants.

17.



4. DISCUSSION AND CONCLUSIONS

4.1 Homology-based PCR cloning of members of a protein kinase subfamily.

We originally set out to identify and clone *Drosophila* genes encoding proteins related to the SNF1/AMPK family of protein kinases by using a homology-based PCR method. The high level of sequence conservation shared by all ePK catalytic domains can be both an advantage and a disadvantage due to the large number of potential amplifiable targets. With hundreds of potential target sequences in the *Drosophila* genome, we biased our PCR primers towards the subfamily of protein kinases of interest. Primers were designed based on regions encompassing both invariant catalytic domain residues as well as sites conserved between yeast SNF1 and rat AMPK.

Two sense and one antisense primers were initially designed (Fig 1.A and 1.B). The sense primer K929 (G-S/T-F-G-K-V-K) is based on the consensus motif GXGXXGXV found in subdomain I. In the crystal structure of PKA-C α , subdomain I folds into a β -strand/turn/ β -strand structure that acts as a flap that covers and anchors the non-transferrable phosphates of ATP and the four invariant residues on the consensus motif stabilize the β -strands (Knighton *et al.*, 1991a , b; Bossmeyer *et al.*, 1993). The sense primer K932 (V-H-R-D-L-K-P-E-N) was based on the consensus sequence HRDLKXXN of subdomain VIb. This consensus sequence is part of what has been termed the “catalytic loop” since the aspartate residue (corresponding to D166 in PKA-C α) acts as the catalytic base that accepts the proton from the hydroxyl group on the substrate (Hanks and Hunter, 1995). The antisense primer K930 (Y-A/V-G-P-E-V-D-I/L/V-W) is based on the sequence around the invariant aspartic acid residue in subdomain IX (D220 in PKA-C α) which is involved in stabilizing the catalytic loop (Hanks and Hunter, 1995). From our subsequent success

with these primers we showed that, by appropriately choosing PCR priming sites, it was possible to target a small subset of sequences from a very large pool of related kinase sequences.

The presence of two copies of the *DM_SNF1B* gene prompted us to ask whether other copies of the *DM_SNF1A* gene might exist in the *Drosophila* genome. This question is of practical importance in our studies because of the possible evolutionary link between *Drosophila* and yeast glucose repression pathways. We were interested in the possibility that there may yet be a *Drosophila* kinase bearing higher similarity to the yeast *SNF1* and which would be more likely to be involved in glucose repression. Especially considering the high similarity between *DM_SNF1A* and the mammalian AMPKs and the fact that the role in glucose repression does not appear to have been conserved in the mammalian AMPKs.

During the initial PCR cloning we had biased the primers by using sequence information from both yeast *SNF1* and rat *AMPK- α* because we were had no *a priori* knowledge of whether the *Drosophila* kinase would be share higher similarity to either of these two kinases. From an evolutionary point one might expect the two animal kinases to be more similar to each other. On the other hand, possible functional similarities between the yeast and *Drosophila* kinases might lead to a higher than expected level of similarity because of convergent evolution. The protein deduced from *DM_SNF1A* was subsequently shown to bear higher similarity to the catalytic subunit of mammalian AMPKs than to that of *SNF1*. In order to look for sequences that might be more similar to yeast *SNF1*, we used amino sequence information from *SNF1* and from two plant kinases that have been shown to complement the *snf1* mutation (*RKINI*: Alderson *et al.*, 1991; *NPK5*: Muranaka *et al.*, 1994). The new primers were fully degenerate so as to hybridize to any potential target sequence (section 2.1). Although the new primers were designed without any regard towards compatibility with the mammalian AMPK sequences, both primers were based on peptides that also

happen to be conserved in the *AMPKs*. The primers could, in theory, be used to amplify the majority of *SNF1/AMPK* kinases included in our phylogenetic survey of close *AMPK* relatives (Fig 14, green and yellow boxes). As a *bona fide* positive control for the new primers, amplification was attempted on human cDNA from intestine, skeletal muscle and liver. In liver we could amplify both $\alpha 1$ and $\alpha 2$ isoforms, whereas in intestine and muscle we could only find $\alpha 1$ and $\alpha 2$ respectively. When these same primers were used on *Drosophila* genomic DNA, the only sequences amplified corresponded to *DM_SNF1A* despite screening numerous clones obtained from several independent PCR amplifications. Ironically, *DM_SNF1A* is not fully conserved within one of the priming sites (having a Q residue instead of G/R residue) although the mismatch is not within the critical 3' end of the primer.

When yeast genomic DNA was used as a template, we found that the products obtained represented *SNF1* and the lesser related kinases *KIN1* and *KIN2* (Levin *et al.*, 1987). The amplification of these two kinases was not entirely expected since *Kin1* and *Kin2* are only approximately 56% similar to *Snf1* within the catalytic domain [in contrast, *Dm_Snf1a* and *Snf1* are 77% similar within the catalytic domain]. The lack of any close relatives to these kinases has previously led to their inclusion at the base of the branch that leads to *AMPK* and *SNF1* (Hanks and Hunter, 1995). An inspection of potential hybridization sites in those two sequences reveals that a combination of full primer sequence degeneracy and low annealing temperatures during PCR were contributing factors in amplifying the *KIN1* and *KIN2* fragments, despite the fact that neither were optimal targets for amplification. That *KIN1* and *KIN2* could be amplified despite sharing weak similarity to *SNF1* would suggest that failure to amplify *SNF1*-like sequences other than *DM_SNF1A* from *Drosophila* genomic DNA is the result of their absence from the genome. Consistent with this hypothesis, hybridizations with the *DM_SNF1A* probe and performed under

relaxed-stringency conditions failed to pick up any secondary hybridization signals (Figure 4.B). Nonetheless, it is impossible to rule out non-conservation at potential primer binding sites, or sub-optimal PCR conditions for the amplification of other potential kinases. One fact worth noting, however, is that the failure to amplify any novel SNF1-like sequences from yeast genomic DNA is also consistent with the absence of any other such sequences in the yeast genome. Results from the kinase survey on the completed yeast genome sequence (Hunter and Plowman, 1997) corroborate these PCR results.

4.2 Evolutionary conservation of *DM SNF1B-1* and *DM SNF1B-2*

DM_SNF1B-1 and *DM_SNF1B-2* were found as tandem gene copies within 3.8 Kb of continuous *Drosophila* genomic sequence. The genes are separated by a small 189 bp. spacer which separates the respective stop and start codons. At the DNA level *DM_SNF1B-1* and *DM_SNF1B-2* have ~ 63% identity. Their respective G+C% contents are also quite different (56.1% and 50.5% respectively). At the amino acid level *DM_SNF1B-1* and *DM_SNF1B-2* share 83% identity and 90% similarity within the catalytic domain, although they are more diverged at the carboxy-terminal ends (34% identity and 44% similarity).

Both genes share approximately 86% similarity with human KP78 within the catalytic domain and 40% in the non-catalytic domain (overall: \approx 43% identity, \approx 57% similarity). The sequences can be easily aligned up to approximately 55 residues past the C-terminus of the catalytic domain. This region encompasses the SNH domain determined by Becker *et al.* (1996), and which they have hypothesized as defining the family of “SNF1-related kinases”. Based on sequence similarity alone, however, it is difficult to speculate on whether *DM_SNF1B-1* and *DM_SNF1B-2* represent *bona fide* *Drosophila* KP78 homologues. Although they are found within a well supported branch that leads to KP78 and the mouse kinase EMK (Inglis *et al.*, 1993), members of this branch do not share the high levels of sequence similarity that are observed within other groups of mammalian kinases. Mouse EMK and human KP78 cluster together on our trees (Figures 13 and 17) but their sequence similarity levels appear to be lower than might be expected for human and mouse orthologous kinase genes. The same can be said about the mouse MSK kinase (Ruiz *et al.*, 1995) because although it branches with the KP78-related group with good support, sequence similarity levels observed between these kinases are certainly much lower than the levels that are observed between the rat and human AMPK- α isoforms. Thus the grouping of KP78-related kinases may or may not reflect functional similarities.

It is interesting to note that although the two protein sequences encoded by *DM_SNF1B-1* and *DM_SNF1B-2* are nearly co-linear, a region (corresponding to amino acids 558 to 637 in *DM_Snf1b-1*) appears to have been lost in *DM_Snf1b-2*. This extra loop found in the sequence of *DM_Snf1b-1* is easily aligned with a corresponding loop in human *Kp78*. The functional significance of these differences is not clear since very little is known about *KP78* and other kinases that group close to it. Recently, the original GenBank entry for *KP78* was amended and the provisional title of MAP/microtubule affinity-regulating kinase 3 was assigned. This was clearly the result of a paper in which Drewes *et al.* (1997) report on the cloning of two rat protein kinase genes that encode products that phosphorylate microtubule-associated proteins, causing microtubule dissociation. Although the two rat protein kinases do share high sequence similarity to human *KP78*, their non-catalytic domains are so different from that of *KP78* that they can not represent orthologous genes. Rat and human orthologous kinase genes might be expected to be over 95% identical at the amino acid level and may be expected to share a large degree of sequence similarity within the non-catalytic domain. The differences observed between the rat MAP kinases and *KP78* are of an equal order of magnitude to those observed when comparing *KP78* to the *AMPKs* and the non-catalytic domains are also unrelated. It would therefore appear that, once again, the role of a kinase has been inferred based on considerable, yet insufficient, similarity levels to another kinase. Recently, however, Peng *et al.* (1998) have isolated a human cDNA which is identical to *KP78*, and which they have dubbed *C-TAK1*. They have subsequently shown that C-Tak1 can phosphorylate Cdc25c, a dual specificity kinase that regulates entry into mitosis. They have therefore hypothesized that a role for C-Tak1 (*Kp78*) may be to regulate the Cdc25c, thus regulating entry into mitosis. These results are certainly consistent with the initial observations of Maheshwari *et al.* (1991, unpublished) that *Kp78* is a protein that is lost in some types of transplantable pancreatic carcinoma cells.

Using a PCR based approach, we were able to identify fragments from genes which appear to correspond to previously characterized mammalian kinase genes: *AMPK- α* and *KP78*. Ample evidence has been accumulated about the conserved role of the mammalian *AMPK- α* isoforms and yeast *SNF1* and their status as species homologues. In the case of *KP78*, the relationship to *SNF1*, if any, is much less clear. Thus far the relationship between the two has been suggested solely on the basis of sequence similarity. Ultimately, the role of *KP78* and other similar kinases will have to be assessed *in vivo*. In this respect, the cloning of similar genes in *Drosophila* will hopefully represent a first step towards future functional studies on these novel kinases. Powerful genetic approaches in *Drosophila* may enable elucidation of the function of *DM_SNF1B-1* and *DM_SNF1B-2* by providing a more tractable system in which to carry out such future studies.

4.3 *Drosophila* SNF1A: a putative homologue of the mammalian AMPKs.

The *Drosophila* *DM_SNF1A* gene encodes a protein kinase which has 60% overall similarity to the yeast Snf1 kinase that is involved in glucose derepression. *DM_Snf1a* has an even higher level of sequence similarity with the mammalian Ampk α -subunits (69% overall). Although not entirely unexpected, the much higher level of similarity between the mammalian *AMPKs* and *DM_SNF1A* is a little surprising considering the conservation between the yeast and *Drosophila* glucose repression pathways and especially considering the expected role of a *Drosophila* *SNF1* homologue in glucose repression.

A comparison of the protein encoded by *DM_SNF1A* to the two human Ampk α -subunits reveals that with the addition of a small number of insertions/deletions (indels) within the non-catalytic domain, the sequences can be aligned across their entire lengths (Fig 11.A). The catalytic domain of *DM_Snf1a* shares 92% similarity at the amino acid level with each human isoform whereas its non-catalytic domain is ~ 51% and ~53% similar to $\alpha 1$ and $\alpha 2$ respectively. One basic aspect conserved in all three kinases is the threonine residue of the “activation loop” (T184 on *DM_Snf1a*) that needs to be phosphorylated for kinase activity in both Snf1 and the Ampk- α subunits. The presence of the same residue in *DM_Snf1a* would indicate the presence of an upstream kinase kinase that would be responsible for its activation. Similarly, the extremely high levels of sequence identity that are observed in the catalytic domain predict similar catalytic domain conformation and it is very likely that *DM_Snf1a* will share similar substrate recognition motifs with both kinases (Dale *et al.*, 1995).

The more distinctive nature of kinase non-catalytic domains allows a number of interesting observations to be made by comparing the primary sequences of all three kinase non-catalytic domains. As mentioned previously *DM_Snf1a*, Ampk- $\alpha 1$ and Ampk- $\alpha 2$ are fully alignable within

their non-catalytic domains. Similarly the non-catalytic domain of the yeast Snf1 can also be fully aligned to all three kinases with the inclusion of indels (Figure 11.B). Consistent with the higher levels of sequence similarity shared between *Drosophila* DM_Snf1a and the mammalian Ampk- α isoforms, these three sequences also share six different indels with respect to the yeast Snf1 primary sequence (Figure 11.B, boxed in green). Interestingly however, *Drosophila* DM_Snf1a and yeast Snf1 share five indels with respect to the two mammalian Ampk- α isoforms (Figure 11.B, boxed in blue). Thus the *Drosophila* DM_Snf1a non-catalytic domain has a mosaic structure with some features held in common with both Ampk- α and Snf1, whereas others are shared only with the Ampk- α isoforms, and others still are shared only with the yeast Snf1.

Considerable efforts have been made to characterize the non-catalytic domains of both yeast Snf1 (Jiang and Carlson, 1996) and Ampk- α 1 (Dyck *et al.*, 1996; Crute *et al.*, 1998). Because of the high level of similarity shared between *Drosophila* DM_Snf1a and its yeast and mammalian counterparts, it is therefore possible to draw on the functional analysis of both homologues in order to determine features unique to the individual proteins as well as features shared by all. The region from L392 to V518 of Snf1 has been previously shown to correspond to the autoinhibitory domain of the Snf1 kinase (termed the Snf1-RD in section 1.4.1.4). The Snf1-RD binds to the kinase catalytic domain thus keeping it inactive (Jiang and Carlson, 1997). It is by interacting with the Snf1-RD that Snf4 is able to release the kinase catalytic domain from auto-inhibition. Except for a small insertion of five residues, DM_Snf1a has a similar region along its sequence that is co-linear with the Snf1 autoinhibitory region. Jiang and Carlson (1996; 1997) have also defined two functional regions of the Snf1 non-catalytic domain that are involved in interacting with Snf4 (the yeast Ampk- γ homologue) and Sip2 (the yeast Ampk- β homologue) respectively. The region from L392 to P495 of Snf1 is responsible for interacting with Snf4 and the corresponding region in both mammalian Ampk- α isoforms has three deletions with respect to Snf1. These deletions are likely to

have a large functional impact and may be at least partly responsible for the failure of experiments involving genetic complementation of Snf1-deficient yeast strains with heterologously expressed mammalian Ampk- α subunits. These deletions may abrogate possible interactions between yeast Snf4 and mammalian Ampk- α subunits. A similar situation has been described for the mammalian β -adrenergic receptor kinases (β ARK-1 and β ARK-2). Chuang *et al.* (1997) had previously characterized a critical domain within the C-terminal non-catalytic domain of β ARK involved in binding the $\beta\gamma$ subunits of the heterotrimeric guanine-nucleotide-binding regulatory G-protein. They observed that the core of this domain was poorly conserved between the two β ARK isoforms and then proceeded to show that differences within this domain lead to functional differences between the isoforms with respect to affinity for the regulatory G-protein (Chuang *et al.*, 1997).

Interestingly, despite having much higher similarity levels with the mammalian Ampk- α subunits than with yeast Snf1 an exceptional feature of the *Drosophila* DM_Snf1a protein is that it lacks these three deletions within the putative Snf4 interaction region and thus similar complementation assays may stand a better chance of success. Another interesting possibility regarding the sequence elements conserved between yeast and *Drosophila* Snf1 homologues may be that they are required for interaction with the glucose repression system, a function that appears to have been lost in the mammalian Ampk- α isoforms.

The region between P515 and N633 of yeast Snf1 has been shown to be necessary for Sip2 interactions (Jiang and Carlson, 1997). Within this region, both *Drosophila* DM_Snf1a and the mammalian Ampk- α subunits bear two deletions with respect to yeast Snf1. At the same time, DM_Snf1a and Snf1 also share a 24-residue deletion with respect to the mammalian Ampk- α subunits. At the functional level Dyck *et al.* (1996) have shown that the mammalian Ampk- α isoforms differ from Snf1 in that they require both β - and γ -subunits in order for full activation to take effect whereas Snf1 only requires Snf4 (the Ampk- γ homologue). It will be interesting to

explore whether the differences mentioned here contribute to the observed functional differences between the Ampk- α isoforms and Snf1. Similarly, it will be interesting to explore the nature of subunit interactions in *Drosophila* and whether only the Ampk- γ homologue is necessary for kinase activity or whether both γ - and β -subunits are required. It is worth noting that Erin Yoshida (1998) has cloned the putative Snf4 homologue from *Drosophila*, and thus with the cloning of two of the Snf1 kinase subunit homologues from *Drosophila*, it should be possible to start addressing these issues.

Crute *et al.* (1998) have recently performed a detailed analysis of non-catalytic domain function in Ampk- α subunits. Their findings have shown that despite some differences, the role of the non-catalytic domain of Ampk- α subunits is analogous to that of the Snf1 non-catalytic domain. Considering the many similarities between DM_Snf1a and its yeast and mammalian counterparts, it is therefore likely that the many aspects of its regulation will also be conserved. Crute *et al.* (1998) have defined the region from L313 to G392 in Ampk- α 1 as a region involved in auto-inhibition and AMP binding. The corresponding region in yeast Snf1 and *Drosophila* DM_Snf1a contains 4 insertions/deletions and it is tempting to speculate that these differences in primary structure are at least partly responsible for the observed lack of allosteric activation of Snf1 by AMP. Similarly, it will be interesting to test whether DM_Snf1a is itself activated by AMP.

As previously alluded to in section 3.2.3.3 (also shown in Table 1), DM_Snf1a shares a slightly higher degree of sequence identity with Ampk- α 2 than it does with Ampk- α 1. Although the overall percentage similarities are very similar, the differences are more readily apparent if: (a) one calculates pairwise protein distances to determine the divergence between sequences [since pairwise protein distance estimates incorporate an evolutionary model, they are more accurate estimates of sequence divergence than pairwise percentage difference calculations], and (b) one looks at the catalytic and non-catalytic domains separately [Comparing divergence rates of catalytic and non-

catalytic domains separately is conceptually similar to comparing the rates of evolution of synonymous vs. non-synonymous codon positions]. As can be seen in Table 2, the pairwise Dayhoff distance estimates for the catalytic domains of the human Ampk- α 1/DM_Snfla and human Ampk- α 2/DM_Snfla pairs are 0.23929 and 0.21825 respectively, which represents a slight difference in sequence divergences. If a similar kind of analysis is performed on the non-catalytic domains, the difference becomes more obvious because of the increased rate of change of the non-catalytic domain. For the non-catalytic domains, the human Ampk- α 1/DM_Snfla and human Ampk- α 2/DM_Snfla pairwise Dayhoff distances are 0.77354 and 0.64395 respectively, which represents a more definite difference in sequence divergence. These distance estimates highlight what could be interpreted as a difference in the evolutionary rates of the human α 1 and α 2 isoforms with respect to DM_Snfla. When one looks at the rat α 1 and α 2 isoforms, the same type of analysis yields similar results. Since both Ampk- α gene copies should be diverging from DM_Snfla at an equal rate these differences are unexpected. Such patterns of molecular evolution are often evidence of gene duplication followed by functional divergence of one of the gene copies (Nadeau and Sankoff, 1997; also reviewed by Ohta, 1998; Wagner, 1998; and others). In order to determine whether the observed differences constitute a definite difference in evolutionary rates it is possible to perform a relative-rate test (Li and Tanimura, 1987). I used the program RRTree of Robinson *et al.* (1998) in order to calculate the statistical support for the observed differences in protein divergence. Analysis using RRTree revealed that the sequence divergence differences between α 1 and α 2 isoforms of both rat and human are not statistically significant. Nevertheless, differences that may be statistically insignificant may yet lead to real differences at the functional level. This possibility is supported by preliminary evidence provided by Vavvas *et al.* (1997) and Salt *et al.* (1998) that suggests possible functional differences between the α 1 and α 2 isoforms. Clearly, the difference in divergence rates observed within the non-catalytic domain of the AMPK- α isoforms is much more

pronounced than that which is observed within the catalytic domain, the highly functionally constrained half of the protein. It has been well established that the non-catalytic domain is involved in regulating kinase activity and in interacting with other subunits and/or cellular components (Jiang and Carlson, 1997; Crute *et al.*, 1998). Differences within the non-catalytic domain might thus be expected if two protein kinases interact with different cellular components. An interesting possibility might then be that DM_Snf1a and AMPK- α 2 share conserved cellular roles while AMPK- α 1 has evolved a novel cellular role by virtue of changes within its non-catalytic domain. In this scenario after the *Drosophila*/human divergence the original AMPK- α gene underwent duplication, yielding α 1 and α 2 isoforms. The fact that one gene copy (α 2) is diverging from DM_Snf1a at a slower rate might then be the result of shared functional constraints with the DM_Snf1a resulting from a shared cellular role. Conversely, the α 1 isoform would be diverging from DM_Snf1a at a faster rate if it has acquired a novel cellular role, which would necessitate changes to the non-catalytic domain.

4.4 Molecular evolution of the SNF1/AMPK protein kinase family.

When Hanks and collaborators first took interest in performing sequence analysis of eukaryotic protein kinases (Hanks *et al.*, 1988; Hanks and Quinn, 1991; Hanks and Hunter, 1995) they were able to show how the alignment and comparison of kinase sequences could allow the development of hypotheses concerning the conserved three-dimensional structure of the catalytic domain and the important role played by the twelve invariant residues. Their subsequent phylogenetic analysis of a data set encompassing all of the then-known kinase sequences, 65 sequences, showed that there is a link between kinase domain sequence similarity and gross-level functional similarities. This finding then made it possible to make preliminary functional predictions for novel kinases based on molecular sequence data. For example, based on molecular sequence alone it is possible to predict whether a putative kinase is a serine/threonine, tyrosine, or dual-specificity kinase (Hanks and Quinn, 1991; Hanks and Hunter, 1995). Similarly, early kinase phylogenies showed that their small data set would breakdown into well-defined groups of highly related kinases (Hanks *et al.*, 1988; Hanks and Quinn, 1991). Among the clusters of like-kinases observed by Hanks *et al.* (1988) were: the protein tyrosine kinases, the cyclic nucleotide-dependent protein kinases, the calcium-phospholipid-dependent kinases, the calcium-calmodulin-activated kinases, the SNF1 kinases, the CDC2-like kinases, and the Raf-Mos proto-oncogene kinases, among others. The success of early kinase phylogenies in correlating function to phylogenetic grouping is due in large part to the fact that the data set was small and biased towards the inclusion of well-defined kinases, their multiple isoforms, and their homologues in various organisms (ie. unambiguously identifiable protein kinases that had been obtained through homology-based cloning

techniques). However, the pace of discovery of novel putative kinase sequences has continued unabated and the kinase gene-tree has become more and more densely populated. This is in large part due to the emergence of techniques that rely on less extensive levels of sequence similarity, such as PCR. As the kinase data set has become more expansive it has become apparent that there are many limitations in trying to correlate phylogenetic inference with the physiological role of an unknown protein kinase. Most of these novel kinases fall outside of the previously well-defined groups of functionally related kinases, making it exceedingly difficult, if not impossible, to make functional predictions based on molecular data alone. We have reached a point in kinase research where molecular cloning of novel sequences has far outstripped the functional studies required to make meaningful predictions on the physiological role.

Because of the large number of sequences present in the various databases with possible ties to the yeast *SNF1* kinase, we carried out a survey on some of these kinases and performed phylogenetic analyses in order to study the evolution of the group of SNF1/AMPK and SNF1-related kinases. During the early stages of the phylogenetic analyses shown here, it soon became apparent that although many sequences have been annotated in the various databases as “SNF1-related”, their overall similarity levels to *SNF1* are well below a hypothetical “threshold” necessary for inclusion into the family of *bona fide* *SNF1* and *AMPK* homologues. Various phylogenetic analyses (Becker *et al.*, 1996; Halford and Hardie, 1998), including the ones presented here, show that yeast *SNF1* and its closest relatives [(the mammalian *AMPKs*, *C. elegans* *PAR2.3*, and the SnRK1 group of plant kinases (which includes rye *RKIN1* and tobacco *NPK5*)] all cluster together as a single monophyletic group in phylogenetic trees, to the exclusion of all other putative SNF1-related kinases. Many of these other SNF1-related sequences cluster into well-supported groups of

their own and the phylogenetic trees shown here reveal that although most of the recent nodes are well supported, the deeper nodes have weak support and are poorly resolved. Since the deeper nodes of the trees are multiforked, it is difficult to claim that any of these kinases represent *SNF1* or *AMPK* relatives, or even that the level of sequence similarity they share with the *SNF1/AMPK* kinases has much functional significance. Any functional predictions for kinases which lie outside of well-supported nodes would require additional physical or biochemical evidence, something that is currently lacking since most of these kinases have been identified by molecular sequence data alone. Because a large number of the kinases studied in this thesis (including *DM_SNF1B-1* and *DM_SNF1B-2*) fall outside the main group of kinases that could be considered *SNF1/AMPK* homologues, the following discussion cannot adequately address them. I will therefore focus on *SNF1*, *AMPK*, and the “true” *SNF1/AMPK* homologues.

Although a lot of the “*SNF1*-related” protein kinases share relatively high levels of sequence similarity to *SNF1/AMPK* within the catalytic domain, the level of conservation within the non-catalytic domain is minimal. It is therefore impossible to align these “*SNF1*-related” sequences to *SNF1* beyond the SNH (*SNF1* homology) domain of Becker *et al.* (1996), a small cluster of conservation approximately thirty-to-forty amino acids from the C-terminus of the catalytic domain. The significance of the SNH domain is itself subject to debate, since the region of similarity is ill defined and standard sequence alignment algorithms are often unable to align the motif with any reliability for the majority of more distant *SNF1/AMPK* “relatives”. In contrast, despite the fact that the C-terminal domain of kinases is subject to most divergence, the “true” *SNF1/AMPK* homologues are distinguished by reasonable levels of sequence conservation observed along the full length of the protein sequence (Fig 14). As I have previously discussed, the C-terminal domain

exerts regulatory control on kinase activity and is involved in interacting with other cellular components, including other kinase subunits, and large differences between different types of kinases is expected. Although the different SNF1/AMPK homologues do accumulate many differences in their non-catalytic domains, it is possible to still observe clusters of conservation across the entire length of the non-catalytic domain. These clusters of conserved residues suggest that the non-catalytic domain is not free to change within groups of related kinases. The kinases included in Figure 14 include kinases of yeast, plant, fruitfly, nematode, human and rat origin and yet it is possible to align all the sequences reasonably well across their entire length. This is in marked contrast with what is observed when comparing the non-catalytic domains of unrelated kinases since these are essentially unalignable.

The conserved regions observed along the alignment of Figure 14 are very reminiscent of the twelve conserved subdomains of the kinase catalytic domain, all of which have structural significance in the context of the three-dimensional folding of the kinase catalytic domain (Hanks *et al.*, 1988). The “islands of conservation” in the non-catalytic domain may themselves have a role in structures involved in interacting with other conserved cellular components. For example, Snf4 is known to interact with the non-catalytic region of Snf1 and the fact that Npk5 (the tobacco Snf1 homologue) can interact with Snf4 *in vivo* (Jiang and Carlson, 1996) shows that at least some of the sequence conservation observed in the non-catalytic region of the kinase is reflected by conservation in protein-protein interactions. Paradoxically, a close examination of the proteins encoded by *NPK5*, *RKIN1*, and mammalian *AMPK- α* genes shows few differences that could account for the observed functional differences with regards to *snf1* complementation. The result from complementation of yeast *snf1* mutants with *DM_SNF1A* is likely to provide us with valuable

insight on sequence/functional conservation of these kinases. Similarly, experiments with chimaeric constructs in which the non-catalytic domains of various animal and plant homologues are fused to a given catalytic domain may also provide insight on the functional aspects of the non-catalytic domain.

As can be seen from the protein distance tree in Figure 15 (and to a lesser extent the protein distance tree seen in Figure 17), the sixteen AMPK/SNF1 kinases can be grouped into at least three major groups that cluster together with strong support and form a monophyletic group. The yeast *SNF1s* from *S. cerevisiae* and *Candida albicans*, the animal kinases (*D. melanogaster*, *C. elegans*, rat and human) and the plant kinases form well-supported branches. One interesting observation is that despite the fact that tobacco Npk5 and rye Rkin1 complement the *snf1* mutation in yeast cells (Alderson *et al.*, 1991; Muranaka *et al.*, 1995) these kinases neither cluster together nor with the yeast Snf1 (Fig 15). It will be interesting to see if any other of the plant kinases included in here show the ability to complement the *snf1* mutation. Conversely, since the human *AMPKs* have been unable to complement the *snf1* mutation, it will be interesting to see if the other animal kinases are also unable to complement *SNF1* function *in vivo*. In this respect, *DM_SNF1A* is unique because, although it clearly forms part of the group of animal *AMPKs*, the *Drosophila* and yeast glucose repression pathways appear to be uniquely conserved.

4.5 Molecular conservation of the glucose repression cascade in *Drosophila*.

The genesis of this project can be traced back to the fortuitous observation that the *Drosophila* α -amylase gene is glucose repressible when introduced into yeast cells (Hickey *et al.*, 1994). This phenomenon was found to be specific to the α -amylase gene since the *Adh* gene, which is not glucose repressible in *Drosophila*, fails to show any glucose regulation when heterologously expressed in yeast cells. At the purely mechanistic level, in order to elicit the appropriate glucose response when introduced into yeast cells, the *Drosophila* α -amylase promoter must interact with the specific regulatory elements responsible for yeast glucose repression. Specifically, if the α -amylase promoter is able to interact with the transcription factor Mig1 through the presence of Mig1-specific *cis* elements, then it should be able to recruit the remainder of the yeast repression machinery. With that in place, the *Drosophila* α -amylase gene can become a part of the yeast glucose regulatory circuit that would cause its repression in the presence of glucose and would cause its derepression when cells are shifted to low glucose conditions.

In the evolutionary context, one very important question that can be asked is if the *Drosophila* and yeast glucose repression pathways are functionally conserved. The larger question becomes if it is possible to imply the existence of a regulatory pathway from the conservation of DNA motifs which form the downstream end of a regulatory pathway? This question can only begin to be addressed through the systematic study and comparison of the components of each individual regulatory pathway. Nevertheless, there is precedent for the conservation of upstream and downstream components of a regulatory pathway. For example, transcriptional regulation of genes by cAMP depends on conserved *cis* elements (CRE) responsible for recruiting cAMP-responsive transcription factors (CREB) that ultimately couple gene expression to cAMP signalling (reviewed by Walton and Rehfuss, 1990; Sassone-Corsi, 1998).

With this in mind, the conservation of *cis* elements that interact with known *trans* acting factors involved in coupling glucose signalling to the regulation of gene expression is highly suggestive of an overall conservation of the regulatory pathway. We set out to examine the possible conservation of known components of the yeast glucose repression pathway in *Drosophila* and we first set our sight on elucidating the exact nature of *Drosophila* glucose repression at the transcriptional level. Efforts were first aimed at defining the minimal *cis* elements required for α -amylase glucose repression and towards the cloning of the putative transcription factor that, by analogy to the yeast glucose repression pathway, would play the role of the Mig1 protein.

I initially used PCR to try and amplify putative *Drosophila* homologues of the yeast *MIG1* gene. From the results shown by Hickey *et al.* (1994), the yeast Mig1 zinc-finger protein is able to interact with the *Drosophila* α -amylase promoter. A *Drosophila* Mig1 homologue likely has a similar recognition sequence and its zinc-finger motif would be expected to show a great deal of conservation. I therefore designed degenerate primers from within a 180 bp region encompassing the Mig1 zinc-finger region. This finger region is extremely well conserved between Mig1 and its putative *Aspergillus nidulans* homologue CreA (Dowzer and Kelly, 1992) but represents the only well-conserved feature between these two homologous proteins. After a number of unsuccessful experiments, it became apparent that the high content of six-fold degenerate amino acids within the zinc-finger domains did not allow for the design of highly specific forward and reverse PCR primers specific for a Mig1-like sequence. I attempted to circumvent this problem by designing a primer based on the conserved His-Thr-Gly-Glu-Lys-Pro motif [HTGEKP] at the junction of the two adjacent zinc-fingers (Nehlin and Ronne, 1990), a motif that is not exclusive to Mig1; although the zinc-finger family of transcription factors is very large, the HTGEKP motif is found in a smaller

subset family members. This primer was used along with vector primers on a larval λ gt-11 cDNA library as template. This approach successfully yielded a fragment from a known zinc-finger sequence [the fat body specific factor *AEF-1* (Falb and Maniatis, 1992)] as well as three other novel zinc-finger-like sequences. Since the *Drosophila* genome is likely to code for a large number of zinc-finger transcription factors and none of the sequences amplified appeared to match very closely with the Mig1 zinc-fingers, it thus appeared this approach would lack the specificity required to target the amplification of a Mig1-like zinc-finger transcription factor from *Drosophila*. A putative *Drosophila* Mig1 transcription factor thus appeared to be beyond the reach of homology-based cloning methods. Similarly, genetic approaches such as those that were used to originally clone *MIG1* in yeast had also proved unsuccessful (H. Ronne, pers. comm.) and thus we decided to set sights on other possible conserved components of the glucose repression pathway. It was at this time that initial reports of the cloning of the mammalian *AMPKs* first started to surface outlining the conservation of *SNF1* and *AMPK* at the molecular sequence level. Because of its role in the glucose repression pathway and the obvious evolutionary conservation with the mammalian *AMPKs*, we thus decided to pursue possible *SNF1* homologues in *Drosophila*.

The work presented here outlines the cloning of three novel protein kinases, one of which is thought to represent a *bona fide* homologue of the *SNF1/AMPK* family of protein kinases. If our hypothesis is correct and the yeast and *Drosophila* glucose repression pathways have been functionally conserved in evolution, the role of *DM_SNF1A* in *Drosophila* physiology should be analogous to that played by *SNF1* in yeast. Its role would thus be that of releasing genes from glucose repression during the shift from high to low glucose conditions. At the same time, the cellular roles of yeast Snf1 and the mammalian Ampk- α isoforms, while similar, are not completely

overlapping. In this respect the high degree of conservation between Snfla and its mammalian counterparts may also mean that the *Drosophila* Snfla shares functional similarities with the two Ampk- α isoforms that are not shared with Snfl.

Along with the cloning of the putative *DM_SNF1A* regulatory subunit (Yoshida, 1998), the results here provide evidence for another conserved component of the glucose repression pathway in *Drosophila* with a homologue in yeast, thus providing further support for the functional conservation of *Drosophila* and yeast glucose repression pathways.

4.6 Evolution of SNF1/AMPK-mediated cascades.

A great deal of biochemical and genetic characterization has gone towards understanding the role of SNF1 and its mammalian counterpart, the AMPK. Aside from the molecular data that initially hinted at a possible conservation of biochemical and physiological characteristics, many resources have been devoted to understanding the many aspects of form and function of these individual kinases and the pathways they are involved in. It is the comparison of these characteristics that has ultimately allowed for the development of hypotheses regarding the evolutionary conservation in the role of the *SNF1*, *AMPK* and their homologues.

Although, at first glance, glucose repression and regulation of fatty acid and cholesterol biosynthesis appear to be unrelated tasks, subsequent characterization of the mammalian and yeast SNF1/AMPK mediated pathways has revealed an unexpected degree of convergence. The similarities shared by SNF1 and AMPK are many, and range from similarities in primary sequence (Mitchelhill *et al.*, 1994; Carling *et al.*, 1994), to similarities in substrate specificity (Mitchelhill *et al.*, 1994; Dale *et al.*, 1995), similarities at the level of the different kinase subunits and their interactions (Stapleton *et al.*, 1994; Gao *et al.*, 1996; Woods *et al.*, 1996), their shared phosphorylation and inactivation of acetyl-CoA carboxylase (Mitchelhill *et al.*, 1994; Woods *et al.*, 1994), their shared *in vitro* inactivation by the phosphatase PP2A and reactivation by partially purified Ampkk (Woods *et al.*, 1994; Wilson *et al.*, 1996). Although the other plant and animal SNF1/AMPK homologues have not been as well characterized, there is already ample evidence [as reviewed in section 1.4.7] that the conservation extends to these homologues as well. On the strength of these findings it has been proposed that these protein kinases have a common origin and furthermore, that the various homologues may have other as of yet unreported functions (Hardie, 1994a, b).

Hardie and MacIntosh (1992) have argued that the SNF1/AMPK mediated cascade has a very ancient origin and may represent one of the earliest signal transduction pathways because nutrient sensing is likely to have been one of the first roles for signal transduction pathways early in evolution. Hardie *et al.* (1998) have recently proposed that the cascades mediated by the SNF1/AMPK kinases are involved in detecting changes in cellular energy status and enacting changes in metabolism, gene expression, and perhaps cell proliferation in response to stresses which compromise the energy status of the cell. In doing so, SNF1/AMPK may protect cells against environmental stress by switching off biosynthesis to conserve ATP. Thus the novel function of providing protection against environmental stress may have evolved from a more ancient response to nutritional stress, as it is still seen in yeast, and as it may well be observed in *Drosophila*.

The evolutionary study of the regulatory pathway mediated by SNF1/AMPK protein kinases has so far centered on establishing the similarities between the mammalian and yeast systems, and trying to elucidate molecular details that are shared between the two. Now that the evolutionary conservation has been well established, I believe that it is time to also start studying the differences between the various systems, as it will undoubtedly yield rich insight into the evolution of gene regulation and signal transduction systems. The study of different aspects of the pathway as it exists in present day organisms should provide us with evidence for the evolution and adaptation of a signalling pathway as organisms have changed over time.

Drosophila thus provides a unique opportunity to study the evolution of SNF1/AMPK kinases because of the proposed conservation between yeast and *Drosophila* glucose repression pathways. In many ways, glucose repression is a “relic” of a time in the evolution of eukaryotes when, because of unicellularity, cells did not have a buffer against drops in extracellular glucose

levels. *Drosophila* provides an “in-between” reference point through which the evolution of the pathway can be studied from both the perspective of a metazoan as well as from the perspective of an organism that still carries out glucose repression.

The differences observed between the mammalian Ampk- α and Snf1 are likely a reflection of differences in physiological role as the original signal transduction pathway evolved from a role in sensing nutrient starvation in unicellular organisms to sensing other cellular stresses which are associated with ATP depletion (Hardie and MacIntosh, 1992). The fact that the lone *Drosophila* *SNF1* homologue is much more similar to the mammalian *AMPKs* despite a putative role in glucose repression raises the question of whether the mammalian *AMPK* has role in glucose signalling. At present time, the list of possible physiological responses associated with activation of these kinases is much larger than may have been predicted based on the extensive yeast and mammalian data that had been accumulated up until the time this thesis work started.

In many respects, the greater contribution of the work presented here is that it raises questions about the evolution of *SNF1/AMPK* kinases, and about the evolution of signalling pathways they mediate. The cloning of a *Drosophila* *SNF1* homologue should represent a first step towards the functional characterization of *SNF1/AMPK* mediated signalling in *Drosophila*. This in turn should provide with the basis for comparative studies on the evolution of the pathway.

4.7 Conclusions

The main goal behind the work described in this thesis was to extend the functional similarities between the yeast and *Drosophila* glucose repression systems in an effort to shed light into the evolution of this regulatory pathway. Whereas previous work sought to determine similarities at the level of transcriptional regulation [ie. the *cis* elements and *trans*-acting factors involved in glucose repression], the cloning of the putative *Drosophila* homologue of the yeast SNF1 kinase, along with the cloning of the putative homologue of the SNF4 regulatory subunit (Yoshida, 1998), establishes that conservation has been maintained in the upstream components of the pathway.

I used a PCR-based approach to identify putative SNF1 kinases from *Drosophila melanogaster*. Three genes showing similarity to SNF1 were subsequently cloned and sequenced. Two of these genes, *DM_SNF1B-1* and *DM_SNF1B-2* share high similarity (approximately 83% within the catalytic domain) with the putative human protein kinase gene KP78 (Maheshwari *et al.*, unpublished; accession # 279966), a kinase whose loss of expression has been associated with the transformation of pancreatic cells into carcinoma cells. KP78 is also a member of an increasing number of protein kinases showing similarity to yeast SNF1. *DM_SNF1B-1* and *DM_SNF1B-2* are clustered in tandem within a 4 Kb region of the *Drosophila* genome. Although the gene-pair shows high conservation in the core catalytic domain, the two copies are quite divergent at their carboxy-terminal ends, indicating perhaps a difference in cellular roles.

The third gene that was cloned and sequenced, *DM_SNF1A*, not only shows high similarity to yeast SNF1, but it shows even greater similarity to the mammalian AMP-activated protein kinases (AMPKs) [~92% similarity in the catalytic domain]. The AMPKs are well-established

homologues of SNF1 and plant SNF1-related kinases (Hardie and Carlson, 1998) and this group of kinases is thought to play a role in enacting ATP-conservation measures and activating alternate pathways of ATP generation in response to cellular ATP depletion.

I performed phylogenetic analysis of SNF1, some of its known homologues, and a number of other kinases exhibiting various levels of sequence similarity to SNF1 in order to study the evolution of this family of proteins. The analyses revealed that SNF1 and well known homologues [including the mammalian AMPKs and plant kinases known to complement *snf1* mutants] all group together with very strong support, to the exclusion of any of the other lesser-related kinases. This group, the SNF1/AMPK kinases, includes the *Drosophila DM_SNF1A*, which therefore likely represents a *bona fide* SNF1 homologue in *Drosophila*, and should play a role in the *Drosophila* glucose repression pathway. The other protein kinases showing similarity to SNF1 break up into groups of their own, including: the plant SnRK2 and SnRK3 groups, and the KP78-related group. The latter includes *DM_SNF1B-1* and *DM_SNF1B-2*, the two other *Drosophila* protein kinase genes reported here. The relationship between these groups of kinases and the SNF1/AMPK kinases is not clear, although phylogenetic trees show very weak support at the base of the tree such that the node connecting the SNF1/AMPK kinases to the SnRK2, SnRK3, and KP78-related kinases is multiforked. There is little similarity between these groups within the non-catalytic domain, which is the ultimate determinant of the functional role of a kinase due to its function in interacting with regulatory subunits and other cellular components. It thus appears that the similarity between SnRK2, SnRK3, and KP78-related kinases and SNF1 has no functional significance.

The primary sequence encoded by the *DM_SNF1A* exhibits a mosaic structure within its non-catalytic regulatory domain. It is extremely well conserved with the two mammalian AMPK- α isoforms, much more so than with yeast SNF1, however it also includes several features held in

common with SNF1 and which are not found in the mammalian AMPKs. These features may be associated with functional differences that are observed between yeast SNF1 and the AMPKs: the lack of an allosteric AMP effect on yeast SNF1, the inability of AMPK- α isoforms to complement *snf1* mutants, the apparent absence of a role in glucose repression for AMPK- α . These issues should become clearer once it has been determined whether *DM_SNF1A* can complement *snf1* mutants, once biochemical characterization of the encoded protein takes place and once it has been determined whether, as expected, *DM_SNF1A* has retained a role in glucose repression. Considering the depth of the biochemical characterization of yeast and mammalian homologues, similar studies on the *Drosophila* kinase should provide valuable insight on the functional evolution of this family of kinases.

5. REFERENCES

- Aguan, K., Scott, J., See, C.G. and Sarkar, N.H. 1994. Characterization and chromosomal localization of the human homologue of a rat AMP-activated protein kinase-encoding gene: a major regulator of lipid metabolism in mammals. *Gene* **149**: 345-350.
- Alderson, A., Sabelli, P.A., Dickinson, J.R., Cole, D., Richardson, M., Kreis, M., Shewry, P.R., and Halford, N.G. 1991. Complementation of Snf1, a mutation affecting global regulation of carbon metabolism in yeast, by a plant protein kinase cDNA. *Proc Natl Acad Sci USA* **88**: 8602-8605.
- Anderberg, R.J. and Walker-Simmons, M.K. 1992. Isolation of a wheat cDNA clone for an abscisic acid-inducible transcript with homology to protein kinases. *Proc Natl Acad Sci USA* **89**: 10183-10187.
- Ashcroft, F.M., Proks, P., Smith, P.A., Ammala, C., Bokvist, K. and Rorsman, P. 1994. Stimulus-secretion coupling in pancreatic beta cells. *J Cell Biochem* **55**: 54-65.
- Ball, K.L., Dale, S., Weekes, J., and Hardie, D.G. 1994. Biochemical characterisation of two forms of 3-hydroxy-3-methylglutaryl-CoA reductase kinase from cauliflower (*Brassica oleracea*). *Eur J Biochem* **219**: 743-750.
- Barford, D., Amit, K.D. and Egloff, M.P. 1998. The structure and mechanism of protein phosphatases: insights into catalysis and regulation. *Annu Rev Biophys Biomol Struct* **27**: 133-164.
- Barker, J.H.A., Slocombe, S.P., Ball, K.L., Hardie, D.G., Shewry, P.R., and Halford, N.G. 1996. Evidence that barley 3-hydroxy-3-methylglutaryl-Coenzyme A reductase kinase is a member of the sucrose nonfermenting-1-related protein kinase family. *Plant Physiol* **112**: 1141-1149.
- Barrit, G.J. 1992. Communication within animal cells. Oxford University Press.
- Becker, W., Heukelbach, J., Kentrup, H. and Joost, H.G. 1996. Molecular cloning and characterization of a novel mammalian protein kinase harboring a homology domain that defines a subfamily of serine/threonine kinases. *Eur J Biochem* **235**: 736-743.
- Beg, Z.H., Allmann, D.W. and Gibson, D.M. 1973. Modulation of 3-hydroxy-3-methylglutaryl coenzyme A reductase activity with cAMP and with protein fractions of rat liver cytosol. *Biochem Biophys Res Commun* **54**: 1362-1369.
- Beg, Z.H., Stonik, J.A. and Brewer H.B. Jr. 1978. 3-Hydroxy-3-methylglutaryl coenzyme A reductase: regulation of enzymatic activity by phosphorylation and dephosphorylation. *Proc Natl Acad Sci U S A* **75**: 3678-3682.
- Beg, Z.H., Stonik, J.A. and Brewer, H.B. Jr. 1979. Characterization and regulation of reductase kinase, a protein kinase that modulates the enzymatic activity of 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *Proc Natl Acad Sci U S A* **76**: 4375-4379.

- Benkel, B.F., Abukashawa, S., Boer, P.H. and Hickey D.A. 1987. Molecular cloning of DNA complementary to *Drosophila melanogaster* α -amylase mRNA. *Genome* **29**: 510-515.
- Benkel, B.F., and Hickey, D.A. 1986a. The interaction of genetic and environmental factors in the control of amylase gene expression in *Drosophila melanogaster*. *Genetics* **114**: 943-954
- Benkel, B.F., and Hickey, D.A. 1986b. Glucose repression of amylase gene expression in *Drosophila melanogaster*. *Genetics* **114**: 137-144.
- Benkel, B.F., and Hickey, D.A. 1987. A *Drosophila* gene is subject to glucose repression. *Proc Natl Acad Sci USA* **84**: 1337-1339.
- Beri, R.K., Marley, A.E., See, C.C., Sopwith, W.F., Aguan, K., Carling, D., Scott, J., and Carey, F. 1994. Molecular cloning, expression and chromosomal localization of human AMP-activated protein kinase. *FEBS Letters* **356**: 117-121.
- Boer, P.H. and Hickey, D.A. 1986. The α -amylase gene in *Drosophila melanogaster*: nucleotide sequence, gene structure and expression motifs. *Nucleic Acids Res* **14**: 8399-8411.
- Boles, E., and Hollenberg, C.P. 1997. The molecular genetics of hexose transport in yeasts. *FEMS Microbiol Rev* **21**: 85-111.
- Boots, M.R., Yeh, Y.M. and Boots, S.G. 1980. Hypocholesterolemic agents VII: Inhibition of β -hydroxy- β -methylglutaryl-CoA reductase by monoesters of substituted glutaric acids. *Pharm Sci* **69**: 506-509.
- Bossemeyer, D., Engh, R.A., Kinzel, V., Ponstingl, H. and Huber, R. 1993. Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0A structure of the complex with Mn^{2+} adenylyl imidodiphosphate and inhibitor peptide PKI(5-24). *EMBO J* **12**: 849-859.
- Brownsey, R.W., Edgell, N.J., Hopkirk, T.J. and Denton R.M. 1984. Studies on insulin-stimulated phosphorylation of acetyl-CoA carboxylase, ATP citrate lyase and other proteins in rat epididymal adipose tissue. Evidence for activation of a cyclic AMP-independent protein kinase. *Biochem J* **218**: 733-743.
- Call, K.M., Glaser, T., Ito, C.Y., Buckler, A.J., Pelletier, J., Haber, D.A., Rose, E.A., Kral, A., Yeger, H., Lewis, W.H., *et al.* 1990. Isolation and characterization of a zinc-finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell* **60**: 509-520.
- Carling, D., Clarke, P.R., Zammit, V.A., and Hardie, D.G. 1989. Purification and characterization of AMP-activated protein kinase : Copurification of acetyl-CoA carboxylase kinase and 3-hydroxy-3-methylglutaryl-CoA reductase kinase activities. *Eur J Biochem* **186**: 129-136.

- Carling, D., Aguan, K., Woods, A., Verhoeven, A.J., Beri, R.K., Brennan, C.H., Sidebottom, C., Davison, M.D., and Scott, J. 1994. Mammalian AMP-activated protein kinase is homologous to yeast and plant protein kinases involved in the regulation of carbon metabolism. *J Biol Chem* **269**: 11442-11448.
- Carling, D., Zammit, V.A. and Hardie, D.G. 1987. A common bicyclic protein kinase cascade inactivates the regulatory enzymes of fatty acid and cholesterol biosynthesis. *FEBS Lett* **223**: 217-222.
- Carlson, M., Osmond, B.C. and Botstein, D. 1981. Mutants of yeast defective in sucrose utilization. *Genetics* **98**: 25-40.
- Carlson, M., Osmond, B.C., Neigeborn, L., and Botstein, D. 1984. A suppressor of *snf1* mutations causes constitutive high-level invertase synthesis in yeast. *Genetics* **107**: 19-32.
- Carlson, C.A. and Kim, K.H. 1973. Regulation of hepatic acetyl coenzyme A carboxylase by phosphorylation and dephosphorylation. *J Biol Chem* **248**: 378-80.
- Cassart, J.-P., Georis, J., Ostling, H., Ronne, H. and Vandenhoute, J. 1995. The MIG1 repressor from *Kluyveromyces lactis*: cloning, sequencing and functional analysis in *Saccharomyces cerevisiae*. *FEBS Lett* **371**: 191-194.
- Cavener, D.R. 1987. Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. *Nucleic Acids Res* **15**: 1353-1361.
- Celenza, J.L. and Carlson, M. 1984a. Structure and expression of the SNF1 gene of *Saccharomyces cerevisiae*. *Mol Cell Biol* **4**: 54-60.
- Celenza, J.L. and Carlson, M. 1984b. Cloning and genetic mapping of SNF1, a gene required for expression of glucose-repressible genes in *Saccharomyces cerevisiae*. *Mol Cell Biol* **4**: 49-53.
- Celenza, J.L., and Carlson, M. 1986. A yeast gene that is essential for release from glucose repression encodes a protein kinase. *Science* **233**: 1175-1180.
- Celenza, J.L., Carlson, M. 1989. Mutational analysis of the *Saccharomyces cerevisiae* SNF1 protein kinase and evidence for functional interaction with the SNF4 protein. *Mol Cell Biol* **9**: 5034-5044.
- Celenza, J.L., Eng, F.J. and Carlson, M. 1989. Molecular analysis of the SNF4 gene of *Saccharomyces cerevisiae*: evidence for physical association of the SNF4 protein with the SNF1 protein kinase. *Mol Cell Biol* **9**: 5045-5054.
- Chang, C. Kwok, S.F., Bleecker, A.B., and Meyerowitz, E.M. 1993. *Arabidopsis* ethylene-response gene *ETR-1*: similarity of product to two-component regulators. *Science* **262**: 539-544.
- Ching, Y.P., Davies, S.P., Hardie, D.G. 1996. Analysis of the specificity of the AMP-activated protein kinase by site-directed mutagenesis of bacterially expressed 3-hydroxy 3-methylglutaryl-CoA reductase, using a single primer variant of the unique-site-elimination method. *Eur J Biochem* **237**: 800-808.

- Chomczynski, P. and Sacchi, N. 1987. Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**: 156-159.
- Chuang, T.T., Pompili, E., Paolucci, L., Sallese, M., De Gioia, L., Salmona, M., De Blasi, A. 1997. Identification of a short sequence highly divergent between beta-adrenergic-receptor kinases 1 and 2 that determines the affinity of binding to betagamma subunits of heterotrimeric guanine-nucleotide-binding regulatory proteins. *Eur J Biochem* **245**: 533-540.
- Clarke, P.R. and Hardie, D.G. 1990. Calmodulin-dependent multiprotein kinase and protein kinase C phosphorylate the same site on HMG-CoA reductase as the AMP-activated protein kinase. *FEBS Lett* **269**: 213-217.
- Clarke, P.R. and Hardie, D.G. 1990. Regulation of HMG-CoA reductase: identification of the site phosphorylated by the AMP-activated protein kinase in vitro and in intact rat liver. *EMBO J* **9**: 2439-2446.
- Cohen, P. and Hardie, D.G. 1991. The actions of cyclic AMP on biosynthetic processes are mediated indirectly by cyclic AMP-dependent protein kinase. *Biochim Biophys Acta* **1094**: 292-299.
- Cooper, J.P., Roth, S.Y., and Simpson, R.T. 1994. The global transcriptional regulators, SSN6 and TUP1, play distinct roles in the establishment of a repressive chromatin structure. *Genes Dev* **8**: 1400-1410.
- Correll, C.C., Ng, L. and Edwards, P.A. 1994. Identification of farnesol as the non-sterol derivative of mevalonic acid required for the accelerated degradation of 3-hydroxy-3-methylglutaryl-coenzyme A reductase. *J Biol Chem* **269**: 17390-17393.
- Corton, J.M., Gillespie, J.G., and Hardie, D.G. 1994. Role of the AMP-activated protein kinase in the cellular stress response. *Current Biol* **4**: 315-324.
- Corton, J.M., Gillespie, J.G., Hawley, S.A. and Hardie, D.G. 1995. 5-aminoimidazole-4-carboxamide ribonucleoside. A specific method for activating AMP-activated protein kinase in intact cells? *Eur J Biochem* **229**: 558-65.
- Cowell, I.G. 1994. Repression versus activation in the control of gene transcription. *Trends Biochem Sci* **19**: 38-42.
- Crute, B.E., Seefeld, K., Gamble, J., Kemp, B.E. and Witters, L.A. 1998. Functional domains of the alpha1 catalytic subunit of the AMP-activated protein kinase. *J Biol Chem* **273**: 35347-35354.
- Dale, S., Wilson, W.A., Edelman, A.M., and Hardie, D.G. 1995. Similar substrate recognition motifs for mammalian AMP-activated protein kinase, higher plant HMG-CoA reductase kinase-A, yeast SNF1, and mammalian calmodulin dependent protein kinase-I. *FEBS Letters* **361**: 191-195.

- Davies, S.P., Carling, D. and Hardie, D.G. 1989. Tissue distribution of the AMP-activated protein kinase, and lack of activation by cyclic-AMP-dependent protein kinase, studied using a specific and sensitive peptide assay. *Eur J Biochem* **186**: 123-128.
- Davies, S.P., Hawley, S.A., Woods, A., Carling, D., Haystead, T.A. and Hardie, D.G. 1994. Purification of the AMP-activated protein kinase on ATP- γ -sepharose and analysis of its subunit structure. *Eur J Biochem* **223**: 351-357.
- Davies, S.P., Helps, N.R., Cohen, P.T. and Hardie, D.G. 1995. 5'-AMP inhibits dephosphorylation, as well as promoting phosphorylation, of the AMP-activated protein kinase. Studies using bacterially expressed human protein phosphatase-2C α and native bovine protein phosphatase-2AC. *FEBS Lett* **377**: 421-425.
- Davis, C.A., Riddell, D.C., Higgins, M.J., Holden, J.J.A. and White, B.N. 1985. A gene family in *Drosophila melanogaster* coding for trypsin-like enzymes. *Nucleic Acids Res* **13**: 6605-6619.
- Denton, R.M. and Brownsey, R.W. 1983. The role of phosphorylation in the regulation of fatty acid synthesis by insulin and other hormones. *Philos Trans R Soc Lond B Biol Sci* **302**: 33-45.
- DeVit, M.J., Waddle, J.A. and Johnston, M. 1997. Regulated nuclear translocation of the Mig1 glucose repressor. *Mol Biol* **8**: 1603-1618.
- Doane, W.W., Treat-Clemons, L.G., Gemmill, R.M., Levy, J.N., Hawley, S.A., Buchberg, A.M. and Paigen K. 1983. Genetic mechanism for tissue-specific control of α -amylase expression in *Drosophila melanogaster*. *Isozymes Curr Top Biol Med Res* **9**: 63-90.
- Douglas, P., Pigaglio, E., Ferrer, A., Halfords, N.G. and MacKintosh, C. 1997. Three spinach leaf nitrate reductase-3-hydroxy-3-methylglutaryl-CoA reductase kinases that are required by reversible phosphorylation and/or Ca^{2+} ions. *Biochem J* **325**: 101-109.
- Dowzer, C.E.A. and Kelly, J.M. 1989. Cloning of the *creA* gene from *Aspergillus nidulans*: a gene involved in carbon catabolite repression. *Curr Genet* **15**: 457-459.
- Dowzer, C.E.A., and Kelly, J.M. 1991. Analysis of *creA* gene, a regulator of carbon catabolite repression in *Aspergillus nidulans*. *Mol Cell Biol* **11**: 5701-5709.
- Drouin, G., Prat, F., Ell, M., Clarke, P.G.D. 1999. Detecting and characterizing gene conversions between multigene family members. *Mol Biol Evol* **16**: 1369-1390
- Dyck, J.R.B., Gao, G., Widmer, J., Stapleton, D., Fernandez, C.S., Kemp, B.E., Witters, L.A. 1996. Regulation of 5'-AMP-activated protein kinase activity by the non-catalytic β and γ subunits. *J Biol Chem* **271**: 17798-17803d.
- Edmondson, D.G., Smith M.M. and Roth, S.Y. 1996. Repression domain of the yeast global repressor Tup1 interacts directly with histones H3 and H4. *Genes Dev* **10**: 1247-1259.

- Entian, K.D. 1980. Genetic and biochemical evidence for hexokinase PII as a key enzyme involved in carbon catabolite repression in yeast. *Mol Gen Genet* **178**: 633-637.
- Entian, K.D. 1986. Glucose repression: a complex regulatory system in yeast. *Microbiol Sci* **3**: 366-371.
- Erickson, J.R. and Johnston M. 1993. Genetic and molecular characterization of *GAL83*: its interaction and similarities with other genes involved in glucose repression in *Saccharomyces cerevisiae*. *Genetics* **135**: 655-664.
- Estruch, F., Treitel, M.A., Yang, X. and Carlson M. 1992. N-terminal mutations modulate yeast SNF1 protein kinase function. *Genetics* **132**: 639-650.
- Falb, D. and Maniatis, T. 1992. A conserved regulatory unit implicated in tissue-specific gene expression in *Drosophila* and man. *Genes Dev* **6**: 454-465.
- Falk, J.J., Bass, R.B., Butler, S.L., Chervitz, S.A. and Danielson, M.A. 1997. The two-component signalling pathway of bacterial chemotaxis: a molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annu Rev Cell Dev Biol* **13**: 457-512.
- Felsenstein, J. 1993. Phylogeny inference package, Version 3.5. University of Washington.
- Feng, Z., Wilson, S.E., Peng, Z.Y., Schlender, K.K., Reiman, E.M., Trumbly, R.J. 1991. The yeast *GLC7* gene required for glycogen accumulation encodes a type 1 protein phosphatase. *J Biol Chem* **266**: 23796-23801.
- Ferrer. A. and Hegardt, F.G. 1984. Phosphorylation of 3-hydroxy-3-methylglutaryl coenzyme A reductase by microsomal 3-hydroxy-3-methylglutaryl coenzyme A reductase kinase. *Arch Biochem Biophys* **230**: 227-237.
- Ferrer. A. Caelles, C., Massot, N. and Hegardt, F.G. 1985. Activation of rat liver cytosolic 3-hydroxy-3-methylglutaryl coenzyme A reductase kinase by adenosine 5'-monophosphate. *Biochem Biophys Res Commun* **132**: 497-504.
- Fields, S. and Song, O. 1989. A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245-246.
- Foretz, M., Carling, D., Guichard, C., Ferre P. and Foufelle F. 1998. AMP-activated protein kinase inhibits the glucose-activated expression of fatty acid synthase gene in rat hepatocytes. *J Biol Chem* **273**: 14767-14771.
- Gao, G., Fernandez, D., Stapleton, A.S., Auster, A.S., Widmer, J., Dyck, J.R.B., Kemp, B.E. and Witters, L.E. 1996. Non-catalytic alpha- and beta-subunit isoforms of the 5'-AMP-activated protein kinase. *J Biol Chem* **271**: 8675-8681.

- Gao, G., Ollinger, K. and Brunk, U.T. 1994. Influence of intracellular glutathione concentration of lipofuscin accumulation in cultured neonatal rat cardiac myocytes. *Free Radic Biol Med* **16**: 187-194.
- Gao, G., Widmer, J., Stapleton, D., Teh, T., Cox, T., Kemp, B.E., and Witters, L.A. 1995. Catalytic subunits of the porcine and rat 5'-AMP-activated protein kinase are members of the SNF1 protein kinase family. *Bioch Biophys Acta* **1266**: 73-82.
- Gancedo, J. M. 1998. Yeast carbon catabolite repression. *Microbiol Molec Biol Rev* **62**: 334-361.
- Garton, A.J., Campbell, D.G., Carling, D., Hardie, D.G., Colbran, R.J., Yeaman, S.J. 1989. Phosphorylation of bovine hormone-sensitive lipase by the AMP-activated protein kinase. A possible antilipolytic mechanism. *Eur J Biochem* **179**: 249-254.
- Garton, A.J., Campbell, D.G., Cohen, P. and Yeaman, S.J. 1988. Primary structure of the site on bovine hormone-sensitive lipase phosphorylated by cyclic AMP-dependent protein kinase. *FEBS Lett* **229**: 68-72.
- Gemmill, R.M., Schwartz, P.E. and Doane, W.W. 1986. Structural organization of the Amy locus in seven strains of *Drosophila melanogaster*. *Nucleic Acids Res* **14**: 5337-5352.
- Giguere, V., Ong, E.S., Segui, P. and Evans, R.M. 1987. Identification of a receptor for the morphogen retinoic acid. *Nature* **330**: 624-629.
- Gilman, A.G. 1987. G-proteins: transducers of receptor-generated signals. *Biochemistry* **56**: 615-649.
- Goldsmith, E.J. and Cobb, M.H. 1994. Protein kinases. *Curr Op Struct Biol* **4**: 833-840.
- Goldstein, J.L. and Brown, M.S. 1990. Regulation of the mevalonate pathway. *Nature* **343**: 425-430.
- Graves, J.D. and Krebs, E.G. 1999. Protein phosphorylation and signal transduction. *Pharmacol Ther* **82**: 111-121.
- Gronemeyer, H. and Pongs, O. 1980. Localization of ecdysterone on polytene chromosomes of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* **77**: 2108-2112.
- Guo, S. and Kemphues, K.J. 1995. Par-1, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81**: 611-620.
- Haj-Ahmad, Y., Hickey, D.A. 1982. A molecular explanation of frequency-dependent selection in *Drosophila*. *Nature* **299**: 350-352.

- Halford, N.G. and Hardie D.G. 1998. SNF1-related protein kinases: global regulators of carbon metabolism in plants? *Plant Mol Biol* **37**: 735-748.
- Halford, N.G., Vicente-Carbajosa, J., Sabelli, P.A., Shewry, P.R., Hannappel, U. and Kreis, M. 1992. Molecular analyses of a barley multigene family homologous to the yeast protein kinase gene SNF1. *Plant J* **2**: 791-797.
- Halford N.G., Man, A.L., Barker, J.H.A., Monger, W., Shewry, P.R., Smith, A.M., Purcell, P.C. 1994. Investigating the role of plant SNF1-related protein kinases. *Biochem Soc Trans* **22**: 953-957.
- Hanks, S. K., Quinn, A.M. and Hunter, T. 1988. The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* **241**: 42-52.
- Hanks, S.K., and Quinn, A.M. 1991. Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol* **200**: 38-62.
- Hanks, S. K. and Hunter, T. 1995. The Eukaryotic protein kinase superfamily. In "The protein kinase facts book, protein-serine kinases" Hardie, G and Hanks, S. (eds.) Academic Press. p.7-47.
- Hannappel, U., Vicente-Carbajosa, J., Barker, J.H., Shewry, P.R. and Halford, N.G. 1995. Differential expression of two barley SNF1-related protein kinase genes. *Plant Mol Biol* **27**: 1235-1240.
- Hardie, D.G. 1989. Regulation of fatty acid synthesis via phosphorylation of acetyl-CoA carboxylase. *Prog Lipid Res* **28**: 117-146.
- Hardie, D.G., Carling, D., Halford, N. 1994. Roles of the Snf1/Rkin1/AMP-activated protein kinase family in the response to environmental and nutritional stress. *Semin Cell Biol* **5**: 409-416.
- Hardie, D.G. and Carling, D. 1997. The AMP-activated protein kinase fuel gauge of the mammalian cell? *Eur J Biochem* **246**: 259-273.
- Hardie, D.G. and MacKintosh, R.W. 1992. AMP-activated Protein Kinase - An Archetypal Protein Kinase Cascade ? *BioEssays* **14**: 699-704.
- Hardie, D.G. 1992. Regulation of fatty acid and cholesterol metabolism by the AMP-activated protein kinase. *Biochem Biophys Acta* **1123**: 231-238.
- Hardie, D.G. 1994a. Ways of coping with stress. *Nature* **370**: 599-600.
- Hardie, D.G. 1994b. An emerging role for protein kinases: the response to nutritional and environmental stress. *Cell Signal* **6**: 813-821.

- Hardie, D.G. and Guy, P.S. 1980. Reversible phosphorylation and inactivation of acetyl-CoA carboxylase from lactating rat mammary gland by cyclic AMP-dependent protein kinase. *Eur J Biochem* **110**: 167-177.
- Hardie, D.G. 1992. Regulation of fatty acid and cholesterol metabolism by the AMP-activated protein kinase. *Biochim Biophys Acta* **1123**: 231-238.
- Hardie, D.G., Carling, D., Carlson, M. 1998. The AMP-activated/SNF1 protein kinase subfamily: metabolic sensors of the eukaryotic cell? *Annu Rev Biochem* **67**: 821-855.
- Hardie, D.G., Salt, I.P., Hawley, S.A. and Davies, S.P. 1999. AMP-activated protein kinase: an ultrasensitive system for monitoring cellular energy charge. *Biochem J* **338**: 717-722.
- Hardy, T.A., Huang, D., and Roach, P.J. 1994. Interactions between cAMP-dependent and SNF1 protein kinases in the control of glycogen accumulation in *Saccharomyces cerevisiae*. *J Biol Chem* **269**: 27907-27913.
- Haribabu, B., Hook, S.S., Selbert, M.A., Goldstein, E.G., Tomhave, E.D., Edelman, A.M., Snyderman, R. and Means, A.R. 1995. Human calcium-calmodulin dependent protein kinase I: cDNA cloning, domain structure and activation by phosphorylation at threonine-177 by calcium-calmodulin dependent protein kinase I kinase. *EMBO J* **14**: 3679-3686.
- Hawley, S.A., Selbert, M.A., Goldstein, E.G., Edelman, A.M., Carling, D. and Hardie, D.G. 1995. 5'-AMP activates the AMP-activated protein kinase cascade, and Ca²⁺/calmodulin activates the calmodulin-dependent protein kinase I cascade, via three independent mechanisms. *J Biol Chem* **270**: 27186-27191.
- Hawley, S.A., Davison, M., Woods, A., Davies, S.P., Beri, R.K., Carling, D., Hardie, D.G. 1996. Characterization of the AMP-activated protein kinase kinase from rat liver and identification of threonine 172 as the major site at which it phosphorylates AMP-activated protein kinase. *J Biol Chem* **271**: 27879-27887.
- Haystead, T.A., Moore, F., Cohen, P. and Hardie, D.G. 1990. Roles of the AMP-activated and cyclic-AMP-dependent protein kinases in the adrenaline-induced inactivation of acetyl-CoA carboxylase in rat adipocytes. *Eur J Biochem* **187**: 199-205.
- Henin, N., Vincent, M.F., Gruber, H.E., Van den Berghe, G. 1995. Inhibition of fatty acid and cholesterol synthesis by stimulation of AMP-activated protein kinase. *FASEB J* **9**: 541-546.
- Henin N., Vincent M.F., and Van den Berghe, G. 1994. Stimulation of rat liver AMP-activated protein kinase by AMP analogues. *Biochim Biophys Acta* **1290**: 197-203.
- Hickey, D.A. and Benkel, B. 1982. Regulation of amylase activity in *Drosophila melanogaster*: effects of dietary carbohydrate. *Biochem Genet* **20**: 1117-1129.

- Hickey, D.A., Benkel, K.I., Fong, Y., and Benkel, B.F. 1994. A *Drosophila* gene promoter is subject to glucose repression in yeast cells. *Proc Natl Acad Sci USA* **91**: 11109-11112.
- Hickey, D.A., Benkel, B.F. and Magoulas, C. 1989. Molecular biology of enzyme adaptations in higher eukaryotes. *Genome* **31**: 272-283.
- Holland R., Witters L.A., and Hardie, D.G. 1984. Glucagon inhibits fatty acid synthesis in isolated hepatocytes via phosphorylation of acetyl-CoA carboxylase by cyclic-AMP-dependent protein kinase. *Eur J Biochem* **140**: 325-333.
- Hubbard, E.J., Jiang, R. and Carlson M. 1994. Dosage-dependent modulation of glucose repression by MSN3 (STD1) in *Saccharomyces cerevisiae*. *Mol Cell Biol* **14**: 1972-1978.
- Hughes, D.A. 1994. Signal transduction. Histidine kinases hog the limelight. *Nature* **369**: 187-188.
- Hunter, T. and Cooper, J.A. 1985. Protein-tyrosine kinases. *Annu Rev Biochem* **54**: 897-930.
- Hunter, T. 1987. A thousand and one protein kinases. *Cell* **50**: 8293-8295.
- Hunter, T., Angel, P., Boyle, W.J., Chiu, R., Freed, E., Gould, K.L., Isacke, C.M., Karin, M., Lindberg, R.A. and van der Geer, P. 1988. Targets for signal-transducing protein kinases. *Cold Spring Harb Symp Quant Biol* **53** (Pt 1): 131-142.
- Hunter, T. 1991. Protein kinase classification. *Methods Enzymol* **200**: 3-37.
- Hunter, T. 1994. 1001 protein kinases redux : towards 2000. *Semin Cell Biol* **5**: 367-376.
- Hunter, T. 1995. Protein kinases and phosphatases: the yin and yang of protein phosphorylation and signalling. *Cell* **80**: 225-236.
- Hunter, T. and Plowman, G.D. 1997. The protein kinases of budding yeast: six score and more. *Trends Biochem Sci* **22**: 18-22.
- Hutber, C.A., Hardie, D.G. and Winder, W.W. 1997. Electrical stimulation inactivates muscle acetyl-CoA carboxylase and increases AMP-activated protein kinase. *Am J Physiol* **272**: E262-E266.
- Ingebritsen, T.S., Lee, H.S., Parker, R.A. and Gibson, D.M. 1978. Reversible modulation of the activities of both liver microsomal hydroxymethylglutaryl coenzyme A reductase and its inactivating enzyme. Evidence for regulation by phosphorylation-dephosphorylation. *Biochem Biophys Res Commun* **81**: 1268-1277.
- Ingebritsen, T.S., Parker, R.A. and Gibson, D.M. 1981. Regulation of liver hydroxymethylglutaryl-CoA reductase by a bicyclic phosphorylation system. *J Biol Chem* **256**: 1138-1144.

- Ingebritsen, T.S., Stewart, A.A. and Cohen P. 1983. The protein phosphatases involved in cellular regulation. 6. Measurement of type-1 and type-2 protein phosphatases in extracts of mammalian tissues; an assessment of their physiological roles. *Eur J Biochem* **132**: 297-307.
- Ingebritsen, T.S. and Cohen, P. 1983. The protein phosphatases involved in cellular regulation. 1. Classification and substrate specificities. *Eur J Biochem* **132**: 255-261.
- Inglis, J., Lee, M. and Hill, R.E. 1993. Emk, a protein kinase with homologues in yeast maps to mouse chromosome 19. *Mamm Genome* **4**: 401-403.
- Jain, J., McCaffrey, P.G., Liner, Z., Kerppola, T.K., Lambert, J.N., Verdine, G.L., Curran, T. and Rao A 1993. Calcineurin phosphatase activity in T lymphocytes is inhibited by FK506 and cyclosporin A. *Nature* **365**: 352-355
- Jiang, R. and Carlson, M. 1996. Glucose regulates protein interactions within the yeast SNF1 protein kinase complex. *Genes Dev* **10**: 3105-3115.
- Jiang, R. and Carlson, M. 1997. The Snf1 protein kinase and its activating subunit, Snf4, interact with distinct domains of the Sip1/Sip2/Gal83 component in the kinase complex. *Mol Cell Biol* **17**: 2099-2106.
- Johnston, M., and Carlson, M. 1992. Regulation of carbon and phosphate utilization. In "The molecular and cellular biology of the yeast *Saccharomyces* : gene expression" Jones, E.W., Pringle, J.R., and Broach, J.R. (eds.) Cold Spring Harbor Laboratory Press. pp.193-281.
- Joseph, L.J., Le Beau, M.M., Jamieson, G.A. Jr., Acharya, S., Shows, T.B., Rowley, J.D. and Sukhatme, V.P. 1988. Molecular cloning, sequencing, and mapping of EGR2, a human early growth response gene encoding a protein with "zinc-binding finger" structure. *Proc Natl Acad Sci USA* **85** (19): 7164-7168. Published erratum appears in *Proc Natl Acad Sci USA* **86**: 515.
- Kalderon, D. and Rubin, G.M. 1988. Isolation and characterization of *Drosophila* cAMP-dependent protein kinase genes. *Genes Dev* **2**: 1539-1556.
- Karin, M. 1994. Signal transduction from the cell surface to the nucleus through the phosphorylation of transcription factors. *Curr Opin Cell Biol* **6**: 415-424.
- Katzenellenbogen, J.A. and Katzenellenbogen, B.S. 1996. Nuclear hormone receptors: ligand-activated regulators of transcription and diverse cell responses. *Chem Biol* **3**: 529-536.
- Keith, M.L., Rodwell, V.W., Rogers, D.H. and Rudney, H. 1979. In vitro phosphorylation of 3-hydroxy-3-methylglutaryl coenzyme A reductase: analysis of ³²P-labelled, inactivated enzyme. *Biochem Biophys Res Commun* **90**: 969-975.
- Keleher, C.A., Redd, M.J., Schultz, J., Carlson, M., and Johnson, A.D. 1992. Ssn6-Tup1 is a general repressor of transcription in yeast. *Cell* **68**: 709-719.

- Kemp, B.E., Mitchelhill, K.I., Stapleton, D., Michell, B.J., Chen, Z.P., and Witters, L.A. 1999. Dealing with energy demand: the AMP-activated protein kinase. *Trends Biochem Sci* **24**: 22-25.
- Kennelly, P.J. and Potts M. 1996. Fancy meeting you here ! A fresh look at “prokaryotic” protein phosphorylation. *J Bact* **178**: 4759-4764.
- Knighton, D.R., Zheng J., Ten Eyck L.F., Ashford V.A., Xuong N.-H., Taylor, S.S. and Sowadski, J.M. 1991. Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**: 407-414.
- Kozak, M. 1997. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res* **15**: 8125-8148.
- Krebs, E.G. 1994. The growth of research on protein phosphorylation. *Trends Biochem Sci* **19**: 439.
- Kudo, N., Barr A.J., Snehal, D. and Lopaschuk, G.D. 1995. High rates of fatty acid oxidation during reperfusion of ischemic hearts are associated with a decrease in malonyl-CoA levels due to an increase in 5'-AMP-activated protein kinase inhibition of acetyl-CoA carboxylase. *J Biol Chem* **270**: 17513-17520.
- Kudo, N., Gillespie, J.G., Kung, L., Witters, L.A., Schulz, R., Clanachan, A.S. and Lopaschuk, G.D. 1996. Characterization of 5'AMP-activated protein kinase activity in the heart and its role in inhibiting acetyl-CoA carboxylase during reperfusion following ischemia. *Biochim Biophys Acta* **1301**: 67-75.
- Lakatos, L. and Banfalvi, Z. 1997. Nucleotide sequence of a cDNA clone encoding an SNF1 protein kinase homologue from *Solanum tuberosum*. *Plant Physiology* **113**: 1004.
- Lefkowitz, R.J. 1993. G protein-coupled receptor kinases. *Cell* **74**: 409-412.
- Le Guen, L., Thomas, M., Bianchi, M., Halford, N.G. and Kreis, M. 1992. Structure and expression of a gene from *Arabidopsis thaliana* encoding a protein related to SNF1 protein kinase. *Gene* **120**: 249-254.
- Lent, B.A. and Kim, K.H. 1983. Phosphorylation and activation of acetyl-coenzyme A Carboxylase kinase by the catalytic subunit of cyclic AMP-dependent protein kinase. *Arch Biochem Biophys* **225**: 972-978.
- Lent, B. and Kim, K.H. 1982. Purification and properties of a kinase which phosphorylates and inactivates acetyl-CoA carboxylase. *J Biol Chem* **257**: 1897-1901.
- Leonard, C.J., Aravind, L., and Koonin, E.V. 1998. Novel families of putative protein kinases in bacteria and archaea: evolution of the “eukaryotic” protein kinase superfamily. *Genome Res.* **8**: 1038-1047.
- Li, W.H. and Tanimura, M. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* **326**: 93-96.

- Louis, N.A. and Witters, L.A. 1992. Glucose regulation of acetyl-CoA carboxylase in hepatoma and islet cells. *J Biol Chem* **267**: 2287-2293.
- Ludin K., Jiang R., and Carlson M. 1998. Glucose-regulated interaction of a regulatory subunit of protein phosphatase 1 with the Snf1 protein kinase in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **95**: 6245-6250.
- Macintosh, R.W., Davies S.P., Clarke P.R., Weekes, J., Gillespie J.G., Gibb B.J., and Hardie D.G. 1992. Evidence for a protein kinase cascade in higher plants: 3-hydroxy-3-methylglutaryl-CoA reductase kinase. *Eur J Biochem* **209**: 923-931.
- Maeda, T., Wurgler-Murphy, S.M., and Saito, H. 1994. A two-component system that regulates an osmosensing MAP kinase cascade in yeast. *Nature* **369**: 242-245.
- Magoulas, C., Bally-Cuif, L., Loverre-Chyurlia, A., Benkel, B. and Hickey, D. 1993. A short 5'-flanking region mediates glucose repression of amylase gene expression in *Drosophila melanogaster*. *Genetics* **134**: 507-515.
- Magoulas, C., Loverre-Chyurlia, A., and Hickey, D.A. 1992. Amylase *cis*-acting sequences mediate alleviation of glucose repression by cAMP in *Drosophila*. *Biochem Cell Biol* **70**: 751-757.
- Makinde, A.O., Gamble, J. and Lopaschuk G.D. 1997. Upregulation of 5'-AMP-activated protein kinase is responsible for the increase in myocardial fatty acid oxidation rates following birth in the newborn rabbit. *Circ Res* **80**: 482-489.
- Maldonado, F. and Hanks, S.K. 1988. A cDNA clone encoding human cAMP-dependent protein kinase catalytic subunit C α . *Nucleic Acids Res* **16**: 8189-8190.
- Man, A.L., Purcell, P.C., Hannappel, U. and Halford, N.G. 1997. Potato SNF1-related protein kinase: molecular cloning, expression analysis and peptide kinase activity measurements. *Plant Mol Biol* **34**: 31-43.
- Mayer-Jaekel, R.E. Okhura, H. Gomes, R. Sunkel, C.E., Baumgartner, S., Hemmings, B.A. and Glover D.M. 1993. The 55 kD regulatory subunit of *Drosophila* protein phosphatase 2A is required for anaphase. *Cell* **72**: 621-633.
- McGarry, J.D., Leatherman, G.F. and Foster D.W. 1978. Carnitine palmitoyltransferase I. The site of inhibition of hepatic fatty acid oxidation by malonyl-CoA. *J Biol Chem* **253**: 4128-4136.
- Michell, B.J., Stapleton, D., Mitchelhill, K.I., House, C.M., Katsis, F., Witters, L.A. and Kemp, B.E. 1996. Isoform-specific purification and substrate specificity of the 5'-AMP-activated protein kinase. *J Biol Chem* **271**: 28445-28450.
- Mills, S.E., Foster, D.W. and McGarry, J.D. 1983. Interaction of malonyl-CoA and related compounds with mitochondria from different rat tissues. Relationship between ligand binding and inhibition of carnitine palmitoyltransferase I. *Biochem J* **214**: 83-91.
- Mitchell, P.J. and Tjian, R. 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA-binding proteins. *Science* **245**: 371-378.

- Mitchelhill, K.I., Stapleton, D., Gao, G., House, H., Michell, B., Katsis, F., Witters, L.A., and Kemp, B.E. 1994. Mammalian AMP-activated protein kinase shares structural and functional homology with the catalytic domain of yeast Snf1 protein kinase. *J Biol Chem* **269**: 2361-2364.
- Monger, W.A., Thomas, T.H., Purcell, P.C. and Halford, N.G. 1997. Identification of a sucrose nonfermenting-1-related protein kinase in sugar beet (*Beta vulgaris* L.). *Plant growth Regul* **22**: 181-188.
- Moore, F., Weekes, J. and Hardie, D.G. 1991. Evidence that AMP triggers phosphorylation as well as direct allosteric activation of rat liver AMP-activated protein kinase. A sensitive mechanism to protect the cell against ATP depletion. *Eur J Biochem* **199**: 691-697.
- Munday, M.R., Carling, D. and Hardie, D.G. 1988. Negative interactions between phosphorylation of acetyl-CoA carboxylase by the cyclic AMP-dependent and AMP-activated protein kinases. *FEBS Lett* **235**: 144-148.
- Munoz-Dorado, J., Inouye, S., and Inouye, M. 1993. Eukaryotic-like protein serine/threonine kinases in *Myxococcus xanthus* a developmental bacterium exhibiting social behaviour. *J Cell Biochem* **51**: 29-33.
- Muoio, D.B., Seefeld, K., Witters L.A. and Coleman, R.A. 1999. AMP-activated kinase reciprocally regulates triacylglycerol synthesis and fatty acid oxidation in liver and muscle: evidence that sn-glycerol-3-phosphate acyltransferase is a novel target. *Biochem J* **338**: 783-791.
- Muranaka, T., Banno, H. and Machida, Y. 1994. Characterization of tobacco protein kinase NPK5, a homologue of *Saccharomyces cerevisiae* SNF1 that constitutively activates expression of the glucose-repressible SUC2 gene for a secreted invertase of *S. cerevisiae*. *Mol Cell Biol* **14**: 2958-2965.
- Nehlin, J.O., and Ronne, H. 1990. Yeast MIG1 repressor is related to the mammalian early growth response and Wilms' tumour finger proteins. *EMBO* **9**: 2891-2898.
- Nehlin, J.O., Carlber, M., and Ronne, H. 1991. Control of yeast GAL genes by MIG1 repressor : a transcriptional cascade in the glucose response. *EMBO* **10**: 3373-3377.
- Neideracher, D. and Entian K.-D. 1991. Characterization of Hex2 protein, a negative regulatory element necessary for glucose repression in yeast. *Eur J Biochem* **200**: 311-319.
- Neigeborn L. and Carlson, M. 1984. Genes affecting the regulation of SUC2 gene expression by glucose repression in *Saccharomyces cerevisiae*. *Genetics* **108**: 845-858.
- Ness, G.C., Spindler, C.D. and Benton, G.A. 1980. Characteristics of magnesium-adenosine triphosphate-dependent inactivators of 3-hydroxy-3-methylglutaryl coenzyme A reductase. *J Biol Chem* **255**: 9013-9016.
- Nesterova M.V., Ulmasov, K.A., Shlyapnikov, S.V. and Severin, E.S. 1981. The autophosphorylation reaction in the mechanism of activation of pig brain cyclic AMP-dependent protein kinase. *Biochim Biophys Acta* **660**: 110-116

- Ohkura, H., Kinoshita N., Minatani S., Toda, S. and Yanagida, M. 1989. The fission yeast *dis2⁻* gene required for chromosome disjoining encodes one of two putative type I protein phosphatases. *Cell* **57**: 997-1007.
- Ohta, T. 1988. Time for acquiring a new gene by duplication. *Proc Natl Acad Sci USA* **85**: 3509-3512.
- Ohya, Y., Kawasaki, H., Suzuki, K., Londesborough, J. and Anraku, Y. 1991. Two yeast genes encoding calmodulin-dependent protein kinases. Isolation, sequencing and bacterial expressions of CMK1 and CMK2. *J Biol Chem* **266**: 12784-12794.
- Ostling, J., Carlberg, M. and Ronne H. 1996. Functional domains in the Mig1 repressor. *Mol Cell Biol* **16**: 753-761.
- Ozcan, C., Dover, J., Rosenwald, A.G., Wolf, S., and Johnston, M. 1996. Two glucose transporters in *Saccharomyces cerevisiae* are glucose sensors that generate a signal for induction of gene expression. *Proc Natl Acad Sci USA* **93** : 12428-12432.
- Ozcan, S., Dover, J., and Johnston, M. 1998. Glucose sensing and signalling by two glucose receptors in the yeast *Saccharomyces cerevisiae*. *EMBO J* **17**: 2566-2573.
- Park, Y.S., Hong, S.W., Oh, S.A., Kwak, J.M., Lee, H.-H. and Nam, H.G. 1993. Two putative protein kinases from *Arabidopsis thaliana* contain highly acidic domains. *Plant Mol Biol* **22**: 615-624.
- Peng, C.Y., Graves, P.R., Ogg, S., Thoma, R.S., Byrnes, M.J. III, Wu, Z., Stephenson, M.T. and Piwnicka-Worms, H. 1998. C-TAK1 protein kinase phosphorylates human Cdc25C on serine 216 and promotes 14-3-3 protein binding. *Cell Growth Differ* **9**: 197-208.
- Picciotto, M.R., Czernik, A.J. and Nairn, A.C. 1993. Calcium/calmodulin-dependent protein kinase I. cDNA cloning and identification of autophosphorylation site. *J Biol Chem* **268**: 26512-26521.
- Ponticos, M., Long Lu, Q., Morgan, J.E., Hardie, D.G., Partridge, T.A., and Carling, D. 1998. Dual regulation of the AMP-activated protein kinase provides a novel mechanism for the control of creatine kinase in skeletal muscle. *EMBO J* **17**: 1688-99.
- Popov, K.M., Zhao, Y., Shimomura, Y., Kuntz, M.J., and Harris, R.A. 1992. Branched-chain α -ketoacid dehydrogenase kinase. Molecular cloning, expression, and sequence similarity with histidine protein kinases. *J Biol Chem* **267**: 13127-13130.
- Popov, K.M., Kedishvili, N.Y., Zhao, Y., Shimomura, Y., Crabb, D.W., and Harris, R.A. 1993. Primary structure of pyruvate dehydrogenase kinase establishes a new family of eukaryotic protein kinases. *J Biol Chem* **268**: 26602-26606.

- Redd, M.J., Arnaud MB, and Johnson AD. 1997. A complex composed of Tup1 and Ssn6 represses transcription *in vitro*. *J Biol Chem* **272**: 11193-11197.
- Ribeiro, R.C., Kushner, P.J. and Baxter, J.D. 1995. The nuclear hormone receptor gene superfamily. *Annu Rev Med* **46**: 443-453.
- Robinson, M., Gouy, M., Gautier, C. and Mouchiroud, D. 1998. Sensitivity of the relative-rate test to taxonomic sampling. *Mol Biol Evol* **15**: 1091-1098.
- Ronne, H. 1995. Glucose repression in fungi. *Trends Genet* **11**: 12-17.
- Rose, M., Albig, W. and Entian K.-D. 1991. Glucose repression in *Saccharomyces cerevisiae* is directly associated with hexokinase phosphorylation by hexokinases PI and PII. *Eur J Biochem* **199**: 511-518.
- Rudolph, J., and Oesterhelt, D. 1995. Chemotaxis and phototaxis require *CheA* histidine kinase in the archaeon *Halobacterium salinarium*. *EMBO J* **14**: 667-673.
- Ruiz, J.C., Conlon, F.L. and Robertson, E.J. 1994. Identification of novel protein kinases expressed in the myocardium of the developing mouse heart. *Mech Dev* **48**: 153-164.
- Saier, M.H. Jr. 1993. Introduction: protein phosphorylation and signal transduction in bacteria. *J Cell Biochem* **51**: 1-6.
- Salt, I.P., Johnson, G., Ashcroft, S.J.H., and Hardie, D.G. 1998. AMP-activated protein kinase is activated by low glucose in cell lines derived from pancreatic beta cells, and may regulate insulin release. *Biochem J* **335**: 533-539.
- Sano, H. and Youssefian, S. 1994. Light and nutritional regulation of transcripts encoding a wheat protein kinase homologue is mediated by cytokinins. *Proc Natl Acad Sci USA* **91**: 2582-2586.
- Sassone-Corsi, P. 1998. Coupling gene expression to cAMP signalling: role of CREB and CREM. *Int J Biochem Cell Biol* **30**: 27-38.
- Sato, R., Goldstein, J.L. and Brown, M.S. 1993. Replacement of serine-871 of hamster 3-hydroxy-3-methylglutaryl-CoA reductase prevents phosphorylation by AMP-activated kinase and blocks inhibition of sterol synthesis induced by ATP depletion. *Proc Natl Acad Sci USA* **90**: 9261-9265.
- Schuster, S.S., Noegel, A.A., Oehme, F., Gerisch, G., Simon, M.I. 1996. The hybrid histidine kinase DokA is part of the osmotic response system of Dictyostelium. *EMBO J* **15**: 3880-3889.
- Shapiro, M.B. and Senapathy, P. 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* **15**: 7155-7174.

Sim, A.T. and Hardie, D.G. 1988. The low activity of acetyl-CoA carboxylase in basal and glucagon-stimulated hepatocytes is due to phosphorylation by the AMP-activated protein kinase and not cyclic AMP-dependent protein kinase. *FEBS Lett* **233**: 294-298

Smith, S.W., Overbeek, R., Woese, C.R., Gilbert, W. and Gillevet, P.M. 1994. The genetic data environment an expandable GUI for multiple sequence analysis. *Comput Appl Biosci* **10**: 671-675.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**: 503-517.

Stapleton, D., Gao, G., Michell, B.J., Widmer, J., Mitchelhill, K., Teh, T., House, C.M., Witters, L.A., and Kemp, B.E. 1994. Mammalian 5'-AMP-activated protein kinase non-catalytic subunits are homologues of proteins that interact with a yeast protein kinase. *J Biol Chem* **269**: 29343-29346.

Stapleton, D., Mitchelhill, K.I., Gao, G., Widmer, J., Michell, B.J., Teh, T., House, C.M., Fernandez, C.S., Cox, T., Witters, L.A. and Kemp, B.E. 1996. Mammalian AMP-activated protein kinase subfamily. *J Biol Chem* **271**: 611-614.

Stapleton, D., Woollatt, E., Mitchelhill, K.I., Nicholl, J.K., Fernandez, C.S., Michell, B.J., Witters, L.A., Power, D.A., Sutherland, G.R. and Kemp, B.E. 1997. AMP-activated protein kinase isoenzyme family: subunit structure and chromosomal location. *FEBS Lett* **409**: 452-456.

Streb, H., Irvine, R.F., Berridge, M.J. and Schulz. I. 1983. Release of Ca^{2+} from a non-mitochondrial intracellular store in pancreatic acinar cells by inositol-1,4,5-triphosphate. *Nature* **306**: 67-69.

Sullivan, J.E., Carey, F., Carling, D. and Beri, R.K. 1994. Characterisation of 5'-AMP-activated protein kinase in human liver using specific peptide substrates and the effects of 5'-AMP analogues on enzyme activity. *Biochem Biophys Res Commun* **200**: 1551-1556.

Sullivan, J.E., Brocklehurst, K.J., Marley, A.E., Carey, F., Carling, D. and Beri, R.K. 1994. Inhibition of lipolysis and lipogenesis in isolated rat adipocytes with AICAR, a cell-permeable activator of AMP-activated protein kinase. *FEBS Lett* **353**: 33-36

Thevelein, J.M. 1994. Signal transduction in yeast. *Yeast* **10**: 1753-1790.

Thompson, J.D., Higgins, D.G. and Gibson, T.J. 1994. CLUSTAL: improving the sensitivity of progressive mutiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nuc Acids Res* **22**: 4673-4680.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nuc Acids Res*, **24**: 4876-4882.

Thompson, D.B., Treat-Clemons, L.G. and Doane, W.W. 1992. Tissue-specific and dietary control of α -amylase gene expression in the adult midgut of *Drosophila melanogaster*. *Exp Zool* **262**: 122-134.

- Thompson-Jaeger, S., Francois, J., Gaughran, J.P., and Tatchell, K. 1991. Deletions of SNF1 affects the nutrient response of yeast and resembles mutations which activate the adenylate cyclase pathway. *Genetics* **129**: 697-706.
- Thornton, C., Snowden, M.A. and Carling, D. 1998. Identification of a novel AMP-activated protein kinase β subunit isoform that is highly expressed in skeletal muscle. *J Biol Chem* **273**: 12443-12450.
- Tillman T.S., Ganster R.W., Jiang, R., Carlson, M., and Schmidt, M.C. 1995. STD1 (MSN3) interacts directly with the TATA-binding protein and modulates transcription of the SUC2 gene of *Saccharomyces cerevisiae*. *Nucleic Acids Res* **23**: 3174-3180.
- Toda, T., Cameron, S., Sass, P., Zoller, M., and Wigler, M. 1987. Three different genes in *S. cerevisiae* encode the catalytic subunits of the cAMP-dependent protein kinase. *Cell* **50**: 277-287.
- Treitel, M.A. and Carlson, M. 1995. Repression by SSN6-TUP1 is directed by MIG1, a repressor/activator protein. *Proc Natl Acad Sci USA* **92**: 3132-3136.
- Treitel, M.A., Kuchin, S. and Carlson, M. 1998. Snf1 protein kinase regulates phosphorylation of the Mig1 repressor in *Saccharomyces cerevisiae*. *Mol Cell Biol* **18**: 6273-6280.
- Tzamarias, D. and Struhl, K. 1995. Distinct TPR motifs of Cyc8 are involved in recruiting the Cyc8-Tup1 co-repressor complex to differentially regulated promoters. *Genes Dev* **9**: 829-831.
- Tu, J. and Carlson, M. 1994. The GLC7 type 1 protein phosphatase is required for glucose repression in *Saccharomyces cerevisiae*. *Mol Cell Biol* **14**: 6789-6796.
- Tu, J. and Carlson, M. 1995. REG1 binds to protein phosphatase type 1 and regulates glucose repression in *Saccharomyces cerevisiae*. *EMBO J* **14**: 5939-5946.
- Tzamarias, D., and Struhl, K. 1994. Functional dissection of the yeast Cyc8-Tup1 transcriptional co-repressor complex. *Nature* **369**: 758-761.
- Vallier, L.G. and Carlson, M. 1994. Synergistic release from glucose repression by *mig1* and *ssn* mutations in *Saccharomyces cerevisiae*. *Genetics* **137**: 49-50.
- Varanasi, U.A., Klis, M., Mikesell, P.B. and Trumbly, R.J. 1996. The Cyc8 (Ssn6)-Tup1 co-repressor complex is composed of one Cyc8 and four Tup1 subunits. *Mol Cell Biol* **16**: 6707-6714.
- Vavvas, D., Apazidis, A., Saha, A.K., Gamble, J., Patel, A., Kemp, B.E., Witters, L.A. and Ruderman, N.B. 1997. Contraction-induced changes in acetyl-CoA carboxylase and 5'-AMP-activated kinase in skeletal muscle. *J Biol Chem* **272**: 13255-13261.
- Velasco, G., Geelen, M.J. and Guzman, M. 1997. Control of hepatic fatty acid oxidation by 5'-AMP-activated protein kinase involves a malonyl-CoA-dependent and a malonyl-CoA-independent mechanism. *Arch Biochem Biophys* **337**: 169-175.

Verhoeven, A.J.M., Woods, A., Brennan, C.H., Hawley, S.A., Hardie, D.G., Scott, J., Beri, R.K., and Carling, D. 1995. The AMP-activated protein kinase gene is highly expressed in rat skeletal muscle. Alternative splicing and tissue distribution of the mRNA. *Eur J Biochem* **228**: 236-243.

Wagner, A. 1998. The fate of duplicated genes: loss or new function ? *BioEssays* **20**: 785-788.

Wallimann, T. 1994. Bioenergetics. Dissecting the role of creatine kinase. *Curr Biol* **4**: 42-46.

Walsh, D.A., Perkins J.P., and Krebs, E.G. 1968. An adenosine 3', 5'-monophosphate-dependent protein kinase from rabbit skeletal muscle. *J Biol Chem* **243**: 3763-3765.

Walton K.M. and Rehfuss R.P. 1990. Molecular mechanisms of cAMP-regulated gene expression. *Mol Neurobiol* **4**: 197-210.

Wang, S. 1995. Molecular Evolution of Trypsin Genes in *Drosophila*. Ph.D. thesis. University of Ottawa, Ottawa, Canada.

Weekes, J., Hawley, S.A., Corton, J., Shugar, D. and Hardie, D.G. Activation of rat liver AMP-activated protein kinase by kinase kinase in a purified, reconstituted system. Effects of AMP and AMP analogues. *Eur J Biochem* **219**: 751-757.

Wera, S. and Hemmings, B.A. 1995. Serine/threonine protein phosphatases. *Biochem J* **311**: 17-29.

Wiemann,S., Voss,H., Kinzel,V. and Pyerin,W. 1991. Rat C- α catalytic subunit of the cAMP-dependent protein kinase: cDNA sequence and evidence that it is the only isoform expressed in myoblasts. *Biochim Biophys Acta* **1089**: 254-256.

Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S., Du, Z., Durbin, R., Favello,A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laister, N.,Latreille, P., Lightning, J., Lloyd, C., McMurray, A., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A.,Saunders, D., Shownkeen, R., Smaldon, N., Smith, A., Sonnhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M.,Vaughan, K., Waterston, R., Watson, A., Weinstock, L.,Wilkinson-Sproat, J. and Wohldman, P. 1994. 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*. *Nature* **368**: 32-38.

Wilson, W.A., Hawley, S.A., and Hardie, D.G. 1996. Glucose repression/derepression in budding yeast : SNF1 protein kinase is activated by phosphorylation under derepressing conditions and this correlates with a high AMP/ATP ratio. *Curr Biol* **6**: 1426-1434.

Winder, W.W. and Hardie, D.G. 1996. Inactivation of acetyl-CoA carboxylase and activation of AMP-activated protein kinase in muscle during exercise. *Am J Physiol* **270**: E299-E304.

- Winder, W.W., Wilson, H.A., Hardie, D.G., Rasmussen, B.B., Hutber, C.A., Call, G.B., Clayton, R.D., Conley, L.M., Yoon, S. and Zhou, B. 1997. Phosphorylation of rat muscle acetyl-CoA carboxylase by AMP-activated protein kinase and protein kinase A. *J Appl Physiol* **82**: 219-225.
- Witters, L.A. and Kemp, B.E. 1992. Insulin activation of acetyl-CoA carboxylase accompanied by inhibition of the 5'-AMP-activated protein kinase. *J Biol Chem* **267**: 2864-2867.
- Witters, L.A., Nordlund, A.C. and Marshall, L. 1991. Regulation of intracellular acetyl-CoA carboxylase by ATP depletors mimics the action of the 5'-AMP-activated protein kinase. *Biochem Biophys Res Commun* **181**: 1486-1492.
- Witters, L.A., Gao, G., Kemp, B.E. and Quistorff B. 1994. Hepatic 5'-AMP-activated protein kinase: zonal distribution and relationship to acetyl-CoA carboxylase activity in varying nutritional states. *Arch Biochem Biophys* **308**: 413-419.
- Woods, A., Munday, M.R., Scott, J., Yang, X., Carlson, M., and Carling, D. 1994. Yeast SNF1 is functionally related to mammalian AMP-activated protein kinase and regulates acetyl-CoA carboxylase in vivo. *J Biol Chem* **269**: 19509-19515.
- Woods, A., Salt, I., Scott, J., Hardie, D.G. and Carling, D. 1996. The $\alpha 1$ and $\alpha 2$ isoforms of the AMP-activated protein kinase have similar activities in rat liver but exhibit differences in substrate specificity *in vitro*. *FEBS Lett* **397**: 347-351.
- Wright, R.M., and Poyton, R.O. 1990. Release of two *Saccharomyces cerevisiae* cytochrome genes, *COX6* and *CYC1*, from glucose repression requires the *SNF1* and *SSN6* gene products. *Mol Cell Biol* **10**: 1297-1300.
- Xu, X.Z., Wes, P.D., Chen, H., Li, H.S., Yu, M., Morgan, S., Liu, Y. and Montell, C. 1998. Retinal targets for calmodulin include proteins implicated in synaptic transmission. *J Biol Chem* **273**: 31297-31307.
- Yang X., Hubbard E.J., Carlson M. 1992. A protein kinase substrate identified by the two-hybrid system. *Science* **257**: 680-682.
- Yang, X., Jiang, R., and Carlson, M. 1994. A family of proteins containing a conserved domain that mediates interaction with the yeast SNF1 protein kinase complex. *EMBO* **13**: 5878-5886.
- Yeaman, S.J. 1990. Hormone-sensitive lipase a multipurpose enzyme in lipid metabolism. *Biochim Biophys Acta* **1052**: 128-132.
- Yoshida E.N. 1998. Molecular Evolutionary Conservation of the Drosophila Alpha-Amylase Gene and its Regulatory Systems. Ph.D. thesis. University of Ottawa, Ottawa, Canada.
- Zagulski, M., Herbert, C.J. and Rytka, J. 1998. Sequencing and functional analysis of the yeast genome. *Acta Biochim Pol* **45**: 627-643.

Zheng, J., Knighton, D.R., Eyck, L.F., Karlsson, R., Xuong, N., Taylor, S.S. and Sowadski, J.M. 1993. Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry* **32**: 2154-2161.

Zimmerman, F.K., and Scheel, I. 1977. Mutants of *Saccharomyces cerevisiae* resistant to carbon catabolite repression. *Mol Gen Genet* **154**: 75-82.

Appendix I

Table 3. SNF1/AMPK and SNF1-related kinases used in the different phylogenetic analyses.

Name	Species	# of residues	Database ^A	Accession #	Reference
RAT_AMPK1	Rat	548	SP	P54645	Stapleton <i>et al.</i> (1996)
RAT_AMPK2	Rat	552	GB	AAA85033.1	Gao <i>et al.</i> (1995)
HUM_AMPK1	Human	550	DBJ	BAA36547.1	Yano, K. (unpublished)
HUM_AMPK2	Human	552	SP	P54646	Aguan <i>et al.</i> (1994)
CEL_U58726	<i>C. elegans</i>	562	GB	AAB00579.1	Wilson <i>et al.</i> (1994)
CEL_PAR 2.3	<i>C. elegans</i>	622	SP	P45894	Wilson, R. (unpublished)
DM_SNF1A	<i>Drosophila</i>	582	GB	AAB71397.1	Taboada and Hickey (this work)
YSC_SNF1	<i>S. cerevisiae</i>	633	SP	P06782	Celenza and Carlson (1986)
RAT_SNRK	Rat	746	EMB	CAA61563.1	Becker <i>et al.</i> (1996)
MUS_MSK	Mouse	481	GB	AAA67926.1	Ruiz <i>et al.</i> (1994)
HUM_KP78	Human	713	PIR	S27966	Maheshwari <i>et al.</i> (unpublished)
MUS_EMK	Mouse	775	PIR	S31333	Inglis <i>et al.</i> (unpublished)
DM_SNF1B-1	<i>Drosophila</i>	684	GB	AAB81836	Taboada and Hickey (this work)
DM_SNF1B-2	<i>Drosophila</i>	604	GB	AAB81836	Taboada and Hickey (this work)
CEL_PAR1	<i>C. elegans</i>	1192	GB	AAA97437.1	Guo and Kempfues (1995)
XEN_P69EG3	<i>X. laevis</i>	651	PIR	S52244	Roghi <i>et al.</i> (unpublished)
ATH_AKIN10	<i>A. thaliana</i>	512	GB	AAA32736.1	Le Guen <i>et al.</i> (1992)
ATH_ATSKIN1	<i>A. thaliana</i>	512	EMB	CAA64382.1	Lessard <i>et al.</i> (unpublished)
HOR_BKIN2	Barley	484	EMB	CAA57898.1	Hannappel <i>et al.</i> (1995)
TOB_NPK5	Tobacco	511	DBJ	BAA05649.1	Muranaka <i>et al.</i> (1994)
POT_SNF1	Potato	512	EMB	CAA65243.1	Man <i>et al.</i> (1997)
BKIN12	Barley	513	EMB	CAA46556.1	Halford <i>et al.</i> (1992)
RKIN1	Rye	502	GB	AAA33921.1	Alderson <i>et al.</i> (1991)
PKABA1	Wheat	332	GB	AAA96325.1	Anderberg and Walker-Simmons (1992)
ATH_S56718	<i>A. thaliana</i>	362	PIR	S56718	Gallois, P. (unpublished)
ATH_S71172	<i>A. thaliana</i>	361	PIR	S71172	Gallois, P. (unpublished)
ATH_ASK1	<i>A. thaliana</i>	363	GB	AAA02840.1	Park <i>et al.</i> (1993)
ATH_ASK2	<i>A. thaliana</i>	353	EMB	CAA78106.1	Park <i>et al.</i> (1993)
WHT_WPK4	Wheat	526	DBJ	BAA34675.1	Sano and Youssefian (1994)
ATH10237.1	<i>A. thaliana</i>	426	EMB	CAB10237.1	Bevan <i>et al.</i> (unpublished)
HUM_CAMKI	Human	370	SP	Q14012	Haribabo <i>et al.</i> (1995)
RAT_CAMKI	Rat	374	SP	Q63450	Picchiotto <i>et al.</i> (1992)
DM_CAMKI	<i>Drosophila</i>	405	EMB	CAA76937	Xu <i>et al.</i> (1998)
YSC_CAMKI	<i>S. cerevisiae</i>	446	SP	P27466	Ohya <i>et al.</i> (1991)
HUM_PKA-C α	Human	351	EMB	CAA30597.1	Maldonado and Hanks (1988)

A - Database symbols: SP- SWISS-PROT; GB - GenBank; EMB - European Molecular Biology Laboratory; PIR - Protein Identification Resource; DBJ - DNA data bank of Japan