A POTENTIAL HOMOLOG TO THE VERTEBRATE PRO-OPIOMELANOCORTIN GENE IN YEAST

by

Campbell Vance Kyle

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Physiology
University of Utah
December 1992

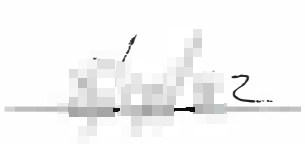
THE UNIVERSITY OF UTAH GRADUATE SCHOOL

SUPERVISORY COMMITTEE APPROVAL

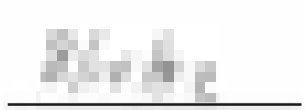
of a dissertation submitted by

Campbell Vance Kyle

This dissertation has been read by each member of the following supervisory committee and by majority vote has been found to be satisfactory.



Chair:



L. White, Ph.D.

Harold M. Brown Jr., Ph.D.

Ralph E. Whatley,

THE UNIVERSITY OF UTAH GRADUATE SCHOOL

FINAL READING APPROVAL

To the Graduate Council of the University of Utah:

I have read the dissertation of Campbell Vance Kyle in its final form and have found that (1) its format, citations and bibliographic style are consistent and acceptable; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the supervisory committee and is ready for submission to The Graduate School.

Date

William D. Odell, M.D., Ph.D.
Chair, Supervisory Committee

Approved for the Major Department

Approved for the Graduate Council

Dean of The Graduate School

ABSTRACT

The vertebrate pro-opiomelanocortin (POMC) gene codes for the large protein precursor to a number of small peptides with highly conserved sequence, including adrenocorticotropin (ACTH), melanocyte stimulating hormone (MSH), and β-endorphin. Previous evidence suggests that structurally and possibly functionally related sequences (homologs) are present in unicellular organisms, and that these may be encoded within a similarly organized genes. This work attempted to identify a POMC homolog in the protozoan, *Tetrahymena pyriformis*, and in fission yeast, *Saccharomyces cerevisiae*.

Using two different equilibrium -type radioimmunoassays with polyclonal antibodies, I initially identified ACTH-like immunoactivity in extracts of both species. However, the same extracts showed no reaction in an immunoradiometric assay requiring greater structural similarity to vertebrate ACTH. The low concentration meant purification of adequate amounts of material to sequence was impractical.

Southern analysis of Tetrahymena genomic DNA was unfruitful in identifying POMC-related sequences. However, yeast genomic DNA showed hybridizing bands when probed at low stringency with the rat POMC gene. Screening a yeast genomic DNA library under low stringency resulted in identification of one of the DNA sequences responsible. Analysis of hybridizing DNA showed a long open reading frame coding for a putative protein of 610 amino acids, showing 18% identity and 45% chemical similarity to the rat POMC precursor, with 30% identity to the vertebrate ACTH sequence. There was no

significant homology to other vertebrate POMC-derived peptides, and the protein showed no features suggesting cleavage into smaller peptides. The presence of a leucine zipper motif and multiple phosphorylation sites indicated an ability to dimerize and possibly bind DNA. A second open reading frame was also found on the opposite DNA strand 500 bases upstream, showing strong homology to DNA-unwinding proteins. A poly(dA-dT) stretch between both genes may function as a common regulatory sequence.

Northern analysis indicated the ORF1gene was expressed. However, deletion and overexpression analysis showed it was not likely to be responsible for the measured immunoactivity, that it was not essential, and that it had no effect on growth rate under normal aerobic conditions. Significance of the sequence similarity to POMC and possible functions are discussed.

TABLE OF CONTENTS

ABSTRACT iv

LIST OF TABLES viii

LIST OF FIGURES ix

ACKNOWLEDGEMENTS xi

1. INTRODUCTION 1

1.1 Pro-opiomelanocortin (POMC) in Vertebrates 1
1.2 Extrapolation to Primitive Organisms 7
1.3 Strategy 17

2. EXTRACTION AND IMMUNOASSAY OF UNICELLULAR ORGANISMS 18

2.1 Extraction Procedures 18
2.2 Immunoassay Procedures 21
2.3 Results of Immunoassays 24
2.4 Experiments to Exclude Artifacts 32
2.5 Conclusions from Immunoassay Experiments 38

3. SOUTHERN ANALYSIS OF GENOMIC DNA 40

3.1 Preparation of High Molecular Weight Genomic DNA 40
3.2 Transfer and Immobilization 43
3.3 Construction and Labelling of DNA Probes 45
3.4 Hybridization Strategy 50
3.5 Results of Southern Blotting 52

4. SCREENING OF YEAST GENOMIC LIBRARIES 57

4.1 Library Screening Methods 57
4.2 Results of Library Screening 60

5. ANALYSIS OF HYBRIDIZING YEAST FRAGMENTS 70

5.1 Methods 70
5.2 Sequence of First Open Reading Frame (ORF1) 74

5.3 Sequence of Second Open Reading Frame (ORF2) 92
5.4 Poly(dA-dT) Enhancer Sequence 98
5.5 Summary 100

6. STUDIES TO DETERMINE SIGNIFICANCE OF ORF1 101

6.1 Expression: Northern Blot Analysis 101
6.2 Genetic Manipulation Experiments 102

7. DISCUSSION AND SIGNIFICANCE 112

REFERENCES 116

LIST OF TABLES

Table		Page
1.	Immunoactivities resembling vertebrate peptides in unicellular organisms	8
2.	Genotype of yeast DY150 strain	19
3.	ACTH-like material in extracts of Tetrahymena	28
4.	ACTH-like material in extracts of yeast	29
5.	Design of oligonucleotides for probing ACTH-like sequences in unicellulars	46
6.	Design of oligonucleotides for probing β-endorphin-like sequences in unicellulars	47
7.	Components of hybridization buffer for Southern blotting	51
8.	Restriction fragment lengths (kb) of rat and yeast DNA (Rat POMC Exon 3 PCR probe)	55
9.	Proteins with closest homology to 610A.A. protein	88
10.	Codon preferences in yeast open reading frame	99
11.	Genotype of M226 yeast strain	106

LIST OF FIGURES

Figure		Page
1.	Primary sequence of vertebrate POMC translation products	2
2.	Structural and functional division of vertebrate POMC molecules	4
3.	Homology between Aplysia FMRF gene and vertebrate POMC and CRH sequences	12
4.	Polyclonal assay of Tetrahymena extract/medium -no inhibitors	26
5.	Polyclonal assay of Tetrahymena extract/medium -with protease inhibitors	27
6.	Polyclonal assay of yeast extract/medium -no protease inhibitors	30
7.	Polyclonal assay of yeast extract/medium -with inhibitors	31
8	Monoclonal immunoassay of extracts and medium	33
9.	Log-logit plots of Tetrahymena dose-response curves	35
10.	Log-logit plots of yeast dose-response curves	36
11.	Experiments to exclude protease artifacts	37
12.	Labelling of rat POMC gene (Exon 3) by PCR	49
13.	Southern blots using oligonucleotide probes	53
14.	Southern blots using rat POMC gene probe	54
15.	Yeast sequences hybridizing to oligonucleotide probes	61
16.	Restriction map and sequencing protocol of 3 kb Sac 1 fragment	66

Figure		Page
17.	Southern blots using rat POMC gene and yeast 1.1 kb Bam H1-Sal 1 fragment	68
18.	Sequence of putative 610 amino acid protein (ORF1)	75
19.	Strongest regional nucleotide homologies between ORF1 and rat POMC gene	78
20.	Alignment of amino acid sequence from ORF1 with rat POMC	80
21.	Derived amino acid sequence and functional motifs of ORF1	82
22.	Helical wheel of potential leucine zipper in 610A.A. protein compared with known leucine zippers	84
23.	Hydrophobicity profile and secondary structure analysis of the putative protein from ORF1	86
24.	Homology of 610 amino acid protein (ORF1) with other proteins	89
25.	Sequence of putative protein from ORF2	93
26.	Sequence and functional motifs of putative ORF2 protein	96
27.	Homology of putative ORF2 protein with other proteins	97
28.	Northern blot of yeast RNA probed with 3kb Sac1 fragment	103
29.	Southern blots to test knockout of ORF1	108
30.	Comparison of growth rate of different yeast strains in defined medium	111

ACKNOWLEDGMENTS

I gratefully thank my supervisors, Dr W.D. Odell and Dr R.F. Gesteland for their supervision and sponsorship during this work. I also wish to thank the other members of my Supervisory Committee for their time and help, divided among their many other responsibilities. Among them was Dr Dixon Woodbury, whose untimely demise unfortunately prevented him from seeing the completion of this work.

CHAPTER 1

INTRODUCTION

1.1 Pro-opiomelanocortin (POMC) in Vertebrates

1.1.1 Amino acid sequences and peptide products

The name "pro-opiomelanocortin," first coined in 1979 by Chretien et al. (1), refers to the large protein precursor to a number of small structurally and functionally diverse bioactive peptides. This 31 kD primary translation product of the pro-opiomelanocortin gene is cut by a series of enzymatic cleavages into progressively smaller peptide fragments (2, 3), with the final profile of peptide products being both species- and tissue-specific (4-6).

The primary amino acid sequence of POMC is shown in Figure 1 for a number of vertebrate species, as determined by either recombinant DNA technology (7-9), or by direct amino acid sequencing of the product peptides extracted from pituitary gland tissue (10-14). As Figure 1 shows, there is significant homology, both in overall structure and primary sequence of the protein, even between species evolutionarily separated by as much as 600 million years (such as between fish and higher terrestrial vertebrates, including rats and humans; 15). In all cases the initial translation product contains a signal peptide largely composed of hydrophobic amino acids, indicating that the newly synthesized product is directed into the lumen of the endoplasmic reticulum for subsequent processing and transport outside the cell and consistent with the "messenger" function of the product peptides (5).

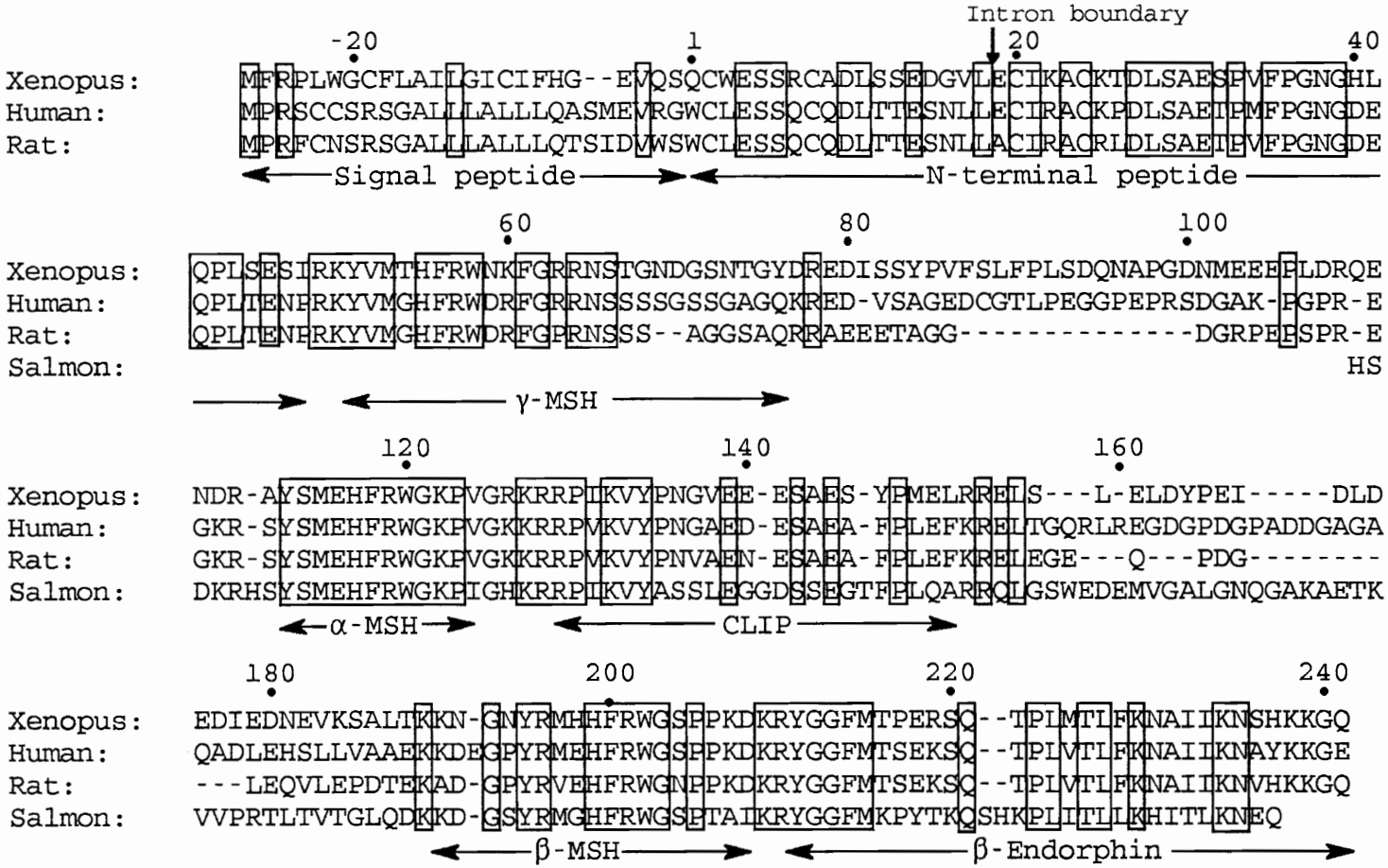


Fig. 1. Primary sequence of vertebrate POMC translation products. Boxed regions show identity in all species. Numbers are from point of signal peptide cleavage. The intron splice junction between second and third exons is shown by an arrow.

As the name suggests, the degree of conservation of the rest of the protein sequence corresponds fairly well to its division into several functionally different peptide groups, in three categories, as described below, (Figure 2).

1.1.1.1 "Opio" products

This part of the name refers to β-endorphin (β-EP) and its subsequent cleavage products, including α-endorphin (β-EP 1-17), γ-endorphin (β-EP 1-16), and Met-enkephalin (β-EP 1-6)(5, 6). These peptides bind with high affinity to receptors found mainly in neurological tissues and initially characterized by their binding of "opiate" compounds such as morphine; such peptides appear to function primarily as neurotransmitters, although other paracrine functions are also evident (10). In accordance with their transmitter function, the endorphins are also highly conserved between different vertebrate species, particularly the most N-terminal sequence (Tyr-Gly-Gly-Phe-Met) of Met-enkephalin.

1.1.1.2 "Melano" products

These are so-called because of their ability to disperse melanin pigment granules and thereby alter skin pigmentation, particularly in frogs and fish (14-16). There are three homologous sequences (termed alpha, beta and gamma) in virtually all vertebrate species, although the most N-terminal (gamma) sequence appears to be absent in fish (17, 18; see below). Such peptides have a conserved core sequence (Met-Glu-His-Phe-Arg-Trp), with variation in surrounding sequence and in their total length. The three forms differ significantly in their melanization potency, with β-MSH possessing only one third of the potency of the alpha peptide, and that of γ-MSH being lower still (8, 9). They also have other less well-defined actions, such as effects on thyroid function, electrolyte balance and behavior, and appear to function as neurotransmitters (10, 19, 20).

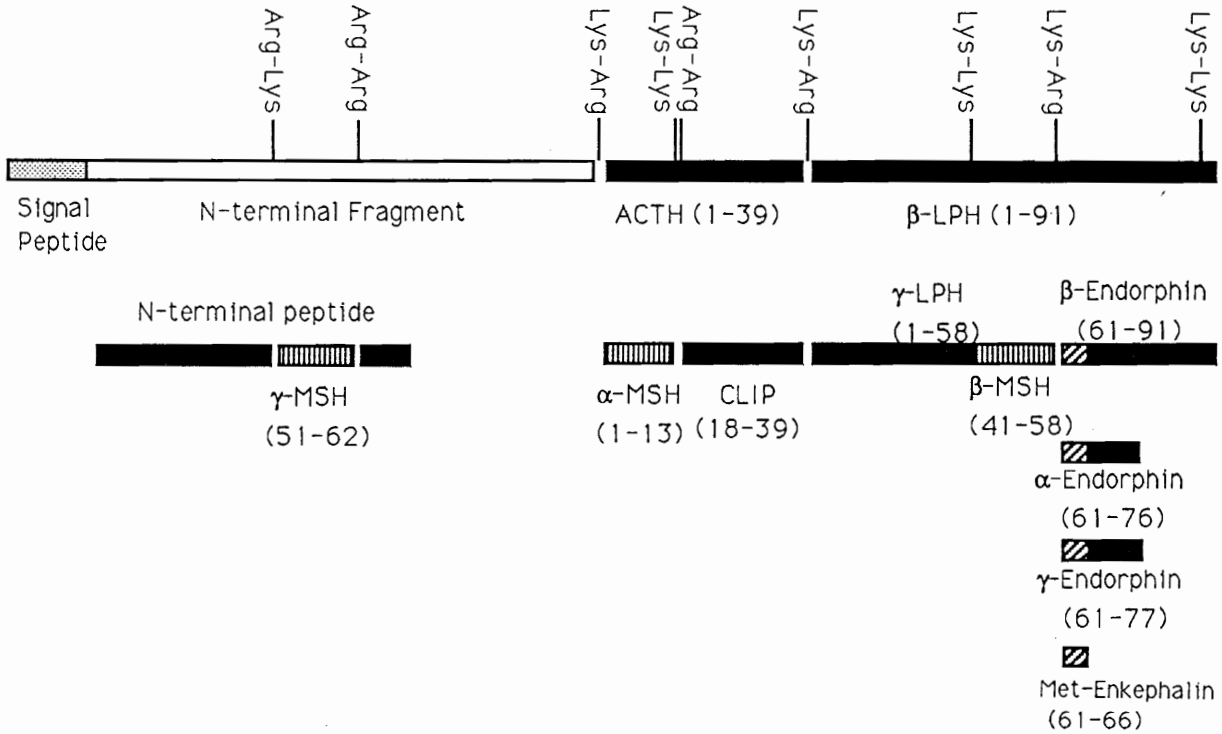


Fig. 2. Structural and functional division of vertebrate POMC molecules. Cleavage sites with resulting product peptides are as shown. Abbreviations: MSH = melanocyte-stimulating hormone; LPH = lipotropin; CLIP = corticotropin-like intermediate lobe peptide; ACTH = adrenocorticotropic hormone.

1.1.1.3 "Cortin" products

This term refers to peptides that stimulate the vertebrate adrenal cortex, principally the 39 amino acid peptide, adrenocorticotropin (ACTH). ACTH shows greater than 80% conservation between all vertebrate species (Figure 2), particularly in its N-terminal part, with close to 100% conservation in residues 1-10 corresponding to the core sequence of α-MSH. The most N-terminal region of the POMC molecule (N-terminal peptide or NPP; 21, 22) may also stimulate adrenocortical function and would therefore belong to this category. N-terminal peptides have a mitogenic effect (22) and also potentiate the steroidogenic effects of ACTH on the vertebrate adrenal gland (22, 23), but the reason for their strong conservation is still not fully explained. Finally, the short 19 amino acid sequence just upstream of ACTH has been postulated as specifically stimulating adrenal androgen production, although this sequence does not show strong conservation between vertebrates and its putative androgen-stimulating role is still disputed (24).

Between these three most conserved regions, there are two regions of low conservation in terms of both length and primary sequence, lying immediately upstream (J2 - "joining peptide") and downstream (J3) of the ACTH sequence respectively. Although the biological functions, if any, of these regions are not known, their lack of conservation suggests they do not have a critical role in vertebrates, and that they are therefore unlikely to be conserved in invertebrates.

1.1.2 Organization of the POMC gene

Vertebrate POMC genes all have basically the same organization. They comprise three exons, with the first exon encoding purely 5'-untranslated RNA upstream sequence. The second encodes the signal peptide and first 18 amino acids of the conserved N-terminal sequence. The third exon encodes all the rest

of the translated protein sequence and 3'-untranslated sequences, including virtually all of the known biologically active peptide sequences in POMC (5, 32).

Most species, including pig (27, 28), cow (29), rat (30, but see 31, 32) and man (33-35) have only one copy of the POMC gene. However, two genes appear to be present in toad (36), mouse (37-39) and salmon (40, 41). One of the mouse genes is a non-functional pseudogene (38), but both toad genes are functional and their very close homology may be due to the postulated evolutionary gene duplication in this species (42). In salmon, two distinct sets of α-MSH, β-MSH and β-EP products can be isolated, although only one complete fish gene has been isolated to date (40, 41).

Conservation of nucleotide sequences parallels that of the protein sequence, although codon degeneracy results in lower nucleotide sequence homology. There is also conservation of 5' and 3' nontranslated regions of POMC mRNA's from different vertebrate species, particularly the presence of a purine-rich stretch (80% A and G) of about 50 nucleotides upstream from the translation start site. This allows for the potential formation of stable hairpin loop structures with regulatory functions (25-28).

Differences in transcriptional start site and posttranscriptional processing result in at least two different sizes of mRNA in species with only one POMC gene, a predominant one of about 1100 bases in tissues such as the anterior pituitary gland, and a second transcript about 200 bases shorter in other tissues, such as the hypothalamus, pituitary intermediate lobe, and gonads. The level of conservation in the 3' nontranslated end of the gene is much lower (26). In most species it is around 160 nucleotides, but it may be considerably longer in fish, at 900 bases (41).

1.2 Extrapolation to Primitive Organisms

It is important to note in the above discussion that all the above structures are from vertebrate species; no POMC-related peptide has been sequenced in invertebrates, the one exception being an incidental finding of sequences distantly related to MSH and CLIP within a neuropeptide gene of Aplysia, termed the FMRF gene because of the repeated sequence Phe-Met-Arg-Phe within its protein product (43; see section 1.2.2). This leads to the obvious question of extension of this vertebrate data to invertebrates, and particularly unicellular organisms, as discussed below.

1.2.1. Antibody evidence

1.2.1.1 Reasons for optimism

There is considerable immunological evidence to suggest that a wide variety of vertebrate "hormones," both peptide and steroid, may have invertebrate homologs of considerable antiquity (44). Thus, substances resembling vertebrate hormones by conventional immunological and/or biological criteria are present in a wide variety of "simple" unicellular organisms, both prokaryotic and eukaryotic (44, 45). A list of some of these activities together with the organisms studied is found in Table 1.

Like most of the other peptides listed in Table 1, antibody evidence suggests that POMC-derived peptides also have a long evolutionary history. Studies have shown ACTH- and β-endorphin-like immunoactivity in numerous invertebrates, including insects, worms and molluscs, either by assay of tissue extracts or immunohistochemistry (71-77). Most studies suffer from problems of antibody specificity. However, Schoofs et al. (73) identified substances resembling α-MSH in both whole organism extracts and in specific neurons of the locust. Unlike most authors, their antisera were well characterized, with epitope

Table 1. *Immunoactivities resembling vertebrate peptides in unicellular organisms*

Immunoactivity.	Organism.
Peptide hormones:	
Insulin	*E. coli* (46), Tetrahymena (47,48)
Chorionic gonadotropin	*P. Maltophilia* (49,50)
-binding sites	*C. albicans* (51), *P. maltophilia*(52)
Somatostatin	Tetrahymena(53), *E. coli*(54)
Relaxin	Tetrahymena (55)
Calcitonin	Candida (56), Tetrahymena (57)
TSH	*C. perfringens* (58).
-binding sites	*E. coli*, *Y. enterocolitica* (59)
Neurotensin	*E. coli*, other bacteria(60)
ACTH, β-endorphin	Tetrahymena (61,62).
-binding activity	Tetrahymena(63), Amoeba (64)
Arginine vasotocin	Tetrahymena(65).
Placental lactogen (hPL)	Tetrahymena (66)
Steroids/Prostaglandins.	
17β-estradiol	*S. cerevisciae* (67)
-receptor	*S. cerevisciae* (68), *P. brasiliensis* (69)
Corticosterone binding	*C. albicans* (70)
Prostaglandin F2α	Tetrahymena (66)

mapping showing their monoclonal antibody to be directed against residues Met^4 Gly^{10} of vertebrate α-MSH, and the polyclonal to be mainly directed against residues Gly^{10}-Lys^{11}, both of which are contained within the biologically active part of the α-MSH molecule. The activity showed parallelism to synthetic vertebrate α-MSH in the polyclonal assay, and HPLC analysis showed a profile similar to α-MSH; further, the extracts showed α-MSH-like activity in a frog skin bioassay. Duvaux-Miret et al. (77) also identified high levels of immunoactivity in extracts of the helminth *Schistosoma mansoni* using several polyclonal antibodies raised against vertebrate ACTH, α-MSH and β-endorphin ("0.6 pmol/pair of worms" for β-endorphin). They identified a major peak of β-endorphin-like activity eluting in identical fashion to vertebrate β-endorphin on HPLC, but they did not show any sequence data to identify the peak. Most other reports involve immunohisto-chemical staining techniques purporting to show POMC-derived peptides in various invertebrate tissues, particularly neural tissue, but specificity of their antisera is uncertain and they give no structural or sequence data.

For unicellular organisms, molecules resembling ACTH and β-endorphin are included in the above list largely on the basis of work done by LeRoith et al. (61, 62), who reported ACTH- and β-endorphin-like immunoactivity in acid extracts of Tetrahymena. These activities appeared to have a range of molecular weights, with higher molecular weight fractions (possibly corresponding to larger precursor forms) having more than one type of immunoactivity. Bioactivity in a dispersed rat adrenal cell bioassay was also demonstrated, which could be removed by prior immunodepletion of the extracts.

1.2.1.2 Antibody evidence: reasons for caution

A logical, but perhaps simplistic conclusion from the above findings is that molecules closely resembling vertebrate hormones in structure and/or function are

present in such early life forms, and may bear an evolutionary relationship to their vertebrate counterparts. However, the significance of both immunoactivity and bioactivity in invertebrates using vertebrate assay systems should not be over-stretched, particularly with polyclonal antisera, and antibodies used at low titers. Quite apart from the need to exclude artifacts (78), antibodies recognize surface shapes rather than sequences, which usually are generated from several adjacent (but linearly separated) folds of a peptide chain, meaning such assay data often say little about either the primary sequence or overall structure of the protein responsible for the activity (79). In some cases, antibodies have recognized "activities" which are claimed to be a particular peptide, but when the molecules responsible were finally identified, have turned out to show little or no primary sequence resemblance to the vertebrate peptide against which the antibody was initially generated (80). For example, a locust hormone reacting strongly with a polyclonal antibody against vertebrate ACTH eventually was found to have a primary sequence more resembling vertebrate arginine vasopressin (ADH), with *no* recognizable sequence homology to ACTH (81, 82)! In other cases the recognition sequence has also turned out to be relatively small (3-4 amino acids at most) and apparently coincidental (83). Similarly, caution is also necessary in the interpretion of bioassay results, which are likewise susceptible to artifact; for example, both cAMP and testosterone can produce increased skin pigmentation in frog skin bioassays for MSH bioactivity (80).

1.2.2 Sequence evidence

1.2.2.1 Reasons for optimism

Sequence data on genes containing POMC-related sequences are scanty in invertebrates, with very few exceptions. Taussig and Scheller (43) made an incidental finding of sequences distantly related to MSH and CLIP within the

FMRF precursor gene of the mollusc Aplysia, a gene primarily coding for 28 copies of the tetrapeptide FMRFamide and one copy of FLRFamide (84). The α-MSH- and CLIP-related sequences showed approximately 40% and 25% respective homology to their vertebrate counterparts, but the two sequences were not contiguous and indeed were present in *reverse* order compared to vertebrate POMC (i.e., the CLIP-like sequence was more N-terminal). Also concerning was the presence of a long amino acid sequence showing 20-30% homology with vertebrate corticotropin releasing factor (CRF). Although the only homology to Met-enkephalin was the repeated C-terminal dipeptide Phe-Met, the gene was overall much closer in organization to the vertebrate pre-proenkephalin gene, which contains six copies of Met-enkephalin (YGGFM, using the one letter amino acid code; 84) and one copy of Leu-enkephalin (Leu for Met)(85, 86). A related FMRF precursor gene is also present in Drosophila, showing weaker homology still to MSH, CLIP and CRF (87); however, this gene is apparently *not* responsible for POMC-like immunoactivity in insects. The homology of the Aplysia sequences with relevant vertebrate sequences is shown in Figure 3.

Although only short, enkephalin sequences are also well conserved in some snails, but no data exist on the genes coding for these peptides (80, 88). Overall, although the vertebrate POMC and pre-proenkephalin genes are evolutionarily related on both sequence and structural grounds, the above data weigh against either the Aplysia FMRF gene or its Drosophila relative being a distinct ancestral POMC homolog and suggest the origin of the vertebrate POMC gene may be more obscure than previously thought. It is thus quite possible that one or both of the above genes is not the true progenitor to vertebrate POMC, but rather an evolutionary branch.

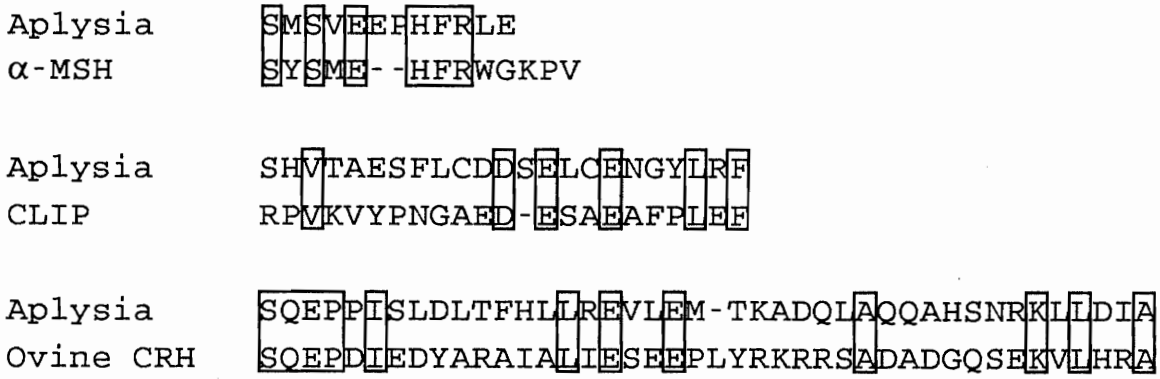


Fig. 3. Homology between Aplysia FMRF gene, and vertebrate POMC and CRH sequences. Vertebrate POMC-derived peptides include melanocyte-stimulating hormone (α-MSH) and corticotropin-like intermediate lobe peptide (CLIP). Homology with ovine CRH is also shown.

The only other genetic evidence is Southern blotting data by Duvaux-Miret et al. (77) suggesting sequences homologous to human POMC-derived peptides (MSH, NPP, Met-enkephalin) are present in annelids. However, the sequences responsible were not isolated and there were several major technical questions overlying the experimental methodology used by these authors.

1.2.2.2 Sequence evidence: reasons for caution

For vertebrates the rate of evolutionary change for ACTH is relatively low at 4.2% residue mutations per 100 million years and 13% per 100 million years for β-lipotropin (which includes the conserved β-endorphin part of the molecule) (89). However, the α-MSH- and CLIP-like sequences in Aplysia (the only known invertebrate sequences with clear similarity to their vertebrate counter-parts) indicate a divergence rate of about 6-7% and 10-12% per 100 million years for ACTH and β-lipotropin respectively. Given the very major assumptions that these are indeed evolutionary intermediates and that this rate of change holds for even more primitive invertebrates (90, 91), such data suggest that the "lookback time" would not be much further back than Aplysia and a protein with discernible

(20-30%) homology may indeed not be recognizable as far back as unicellular eukaryotes (1.5 billion years; 15, 92). Alternatively, if Aplysia is diverged from the main evolutionary pathway between yeast and vertebrates, the yeast homolog may be as much or even *more* closely related to vertebrate POMC than is its molluscan counterpart.

Overall, it is fair to say that the apparent plethora of immunological (and biological) evidence is overshadowed by a virtual absence of sequence data, not only in primitive unicellular life forms but also in more advanced invertebrate forms such as molluscs, arthropods and coelenterates, where it should be easier to identify and characterize a relative of the vertebrate hormone because of the shorter evolutionary distance. Although this lack of data is certainly not encouraging, nor is it sufficiently conclusive as to stop one from looking. It is therefore tempting to also look for other evidence to support or discredit a direct POMC lineage from unicellulars to vertebrates.

1.2.3 Organizational evidence

1.2.3.1 Reasons for optimism

The production of multiple small bioactive peptides as proteolytic cleavage products from large precursor protein molecules is a common and ancient one in biology (93). Even in primitive organisms, most invertebrate genes encode at least two (and often multiple) separate peptides, which may be identical, related, or completely different from each other (80). For example, the yeast alpha factor precursor contains four identical copies of its short pheromone product (94). Vertebrate-like peptides from numerous invertebrate species have also been shown to have multiple molecular weight forms (54, 61) and in some cases (e.g., insect glucagon-like activity) the shift to smaller molecular size has been duplicated by exposure to trypsin (95). This suggests many bioactive peptides are

synthesized in invertebrates as high molecular weight precursors, which are then processed to smaller more bioactive peptide products. Because the vertebrate POMC gene also conforms well to this prohormone-multiple product rule, it is reasonable to expect potential yeast MSH-, CLIP- and opioid-like sequences to be similarly arranged within a large prohormone in invertebrates.

The Tetrahymena data of LeRoith et al. (61, 62) indirectly suggest progressive cleavage of a large POMC-like prohormone molecule. When cell extracts were applied to a size fractionation column the immunoactivity was of multiple sizes, ranging from high molecular weight (>30 kDa) to significantly less than that of the vertebrate monomeric species (4.5 kDa for ACTH1-39; 3.6 kDa for β-EP); the larger species reacted with both antibodies, suggesting these larger forms contained both ACTH- and β-endorphin-like epitopes and were either the primary translation product or processing intermediates of a POMC-like gene. A similar heterogeneous molecular weight profile has also been suggested for insulin-like and somatostatin-like immunoactivity in *Escherichia coli* (47, 54).

1.2.3.2 "Organizational" evidence: reasons for caution

Despite the above evidence, antibody data in invertebrates (particularly insects) are *not* universal in supporting the concept of multiple peptides derived from a common POMC gene, even when one allows for difficulties in antibody specificity. In some cases insect neural cells staining for α-MSH activity show no reaction to antisera directed against other POMC-derived peptides, such as ACTH and β-endorphin (73, 74). When present, these and other immunoactivities (e.g., against Met-enkephalin) appear to be in other distinct neurons; further, γ-MSH activity is apparently uniformly absent (73). Such an overall discrepancy in immunostaining pattern is also found in other insect studies, and prompted De

Loof and Schoofs (80) to conclude elsewhere that they had significant doubts about whether a vertebrate-like POMC precursor gene exists in invertebrates.

1.2.4 Functional evidence

Because the "hormones" shown in Table 1 appear to have an ancient phylogenetic origin, they have undoubtedly withstood enormous evolutionary pressures, and it is logical to conclude that they have retained some fundamental function(s) more "basic" than their complex roles in multicellular organisms, perhaps even directly affecting cell viability, growth and differentiation, cell communication or defense (96, 97). Can evidence for a persistence of such fundamental roles at a local tissue level also be found in higher life forms? In the case of POMC-derived peptides, such evidence definitely exits. ACTH and opiate peptides function as specific neurotransmitters (73-75, 98-100) and immunomodulators (101-103) in both vertebrate and invertebrate. Additionally, POMC-derived peptides are locally expressed during organogenesis (104, 105), affect growth and differentiation in at least some tissues (106-108), and are present ubiquitously in vertebrate tissues, where in some cases such as the immune system, gut, and gonads, they are known to have important paracrine functions (109-110).

1.2.5 Other "indirect" evidence

1.2.5.1 Reasons for optimism

The metabolic similarity between yeast and human cells has led at least one distinguished scientist to consider yeast to be almost a "simplified 'prototype' version of the vertebrate cell" (111). As well as similarity in messenger molecules, there is strong conservation of many aspects of the associated messenger processing, regulatory and receptor machinery between vertebrates and

unicellular organisms. For example, the yeast KEX2 (112) and KEX1(113) enzymes are, respectively, an endopeptidase cleaving at pairs of basic amino acids, and a carboxypeptidase trimming back at the C-terminal ends of the resulting fragments. Together, these two *yeast* enzymes can correctly process the *vertebrate* POMC precursor in mammalian cells (114), leading to the obvious question that if such enzymes are present in yeast, what is their natural substrate?

Several of the acidic proteins normally found associated with peptide hormones (such as ACTH and β-EP) in vertebrate secretory granules are also conserved in unicellulars (115), and there is significant conservation of response pathways. Intercellular communication systems are well documented between unicellular organisms (116). Peptide receptors (117, 118), and second messenger pathways (G proteins, cyclic AMP, inositol triphosphate and protein kinases; 119, 120) are all conserved, indicating that the basic cellular response pathways for peptide stimuli is very similar. This conservation of response pathways is also present for steroid hormones (121).

1.2.5.2 Reasons for caution

Despite the pronounced similarities between mammalian and unicellular hormone/receptor systems, they are by no means identical. Yeast cells often process vertebrate peptide hormones in different or additional ways (e.g., extra proteolysis steps, 122; glycosylation, 123) to mammalian cells, and yeast steroid receptors differ from their vertebrate counterparts in both specificity and cellular context requirements (124). In almost every instance, research shows that invertebrates have an endocrine system which is both diverse and complex, with many invertebrate hormones of obscure evolutionary origins showing little homology to vertebrate peptides (80, 83). Particularly in insects and molluscs, not only one but many relatives to a vertebrate hormone can often be found,

giving the impression that the evolution of many hormones looks "more like a rather tangled bush rather than a tree with distinct branch points"(83) and that significant divergence into multiple forms with probably different functions has sometimes occurred even within each species (70).

1.3 Strategy

Several approaches are possible towards identifying a potential POMC-homolog in unicellular organisms (125, 126). All assume there is enough conservation of sequence at protein and genetic levels to enable antibody or DNA probes to distinguish the relevant sequence from other unrelated ones, although the evolutionary ancestry of these organisms suggests that the level of homology is likely to push the limits of existing techniques. Important steps therefore include: a) Assay for the presence or absence of substance(s) in extracts of unicellulars crossreacting with antibodies against vertebrate POMC peptides. If present, these can be characterized, purified and ultimately, if they can be purified to homogeneity in sufficient quantities, sequenced. The peptide homology with vertebrate POMC peptides can then be determined and this sequence used to design DNA probes for the gene. b) More direct searching for sequences homologous to the vertebrate gene by i) probing of Southern blots of unicellular genomic DNA with vertebrate sequences, and ii) amplification of homologous sequences from unicellular genomic DNA by the polymerase chain reaction (PCR), using "best guess" primers based on the vertebrate peptide sequence (127). Genomic Southern blots can indicate the number of "genes" and their degree of homology, while sequencing of PCR products will indicate whether the conserved primers span an open reading frame with homology to the vertebrate gene. Identification of the homologous gene(s) can then be achieved by screening a DNA library using the appropriate specific DNA probe (125).

CHAPTER 2

EXTRACTION AND IMMUNOASSAY OF UNICELLULAR ORGANISMS

2.1 Extraction Procedures

2.1.1 Tetrahymena

Wild type *Tetrahymena pyriformis* was obtained as a kind gift of Dr G. Herrick, Dept of Cellular, Viral and Molecular Biology, University of Utah. Cells were grown in one liter aliquots of either enriched medium consisting of either 2% Proteose Peptone (Gibco), 1% Yeast extract (Gibco), 0.1% Sequestrene (NEFS medium), or a completely synthetic medium (128). They were incubated under normal (atmospheric) aeration at 30°C, at 150 rpm. At the end of the logarithmic phase of growth (about 72 h) cells were harvested by centrifugation, and extracted according to the following protocol:

Cells were centrifuged in a Sorvall RC3C swinging bucket rotor at 1200 *g* for 10 min at 4°C. The medium (supernatant) was carefully removed, taking care to leave all the cells in the precipitate. They were then washed twice in ice cold distilled water, centrifuged as before, and the final cell pellet weighed prior to extraction. To extract, cells were homogenized in 5 volumes of glacial acetic acid (109, 110), using a Brinkmann (Westbury, NY) Polytron homogenizer (speed setting 8 for 3 min). The homogenate was then agitated for 15 min at 4°C, and centrifuged for 15 min at 10,000 *g*. The supernatant was shell frozen in a clean siliconized glass container and lyophilized until dry. The lyophilizate was

resuspended in 5 volumes of 0.1 M HCl and agitated for 15 min on ice (with or without the following protease inhibitors, all purchased from Sigma Chemical Co., St Louis, MO: Aprotonin, 2% v/v; Leupeptin, 1 μg/ml; EDTA, 1 mM; Phenyl-methylsulfonyl fluoride, 1 mM). The solution was then neutralized by addition of the minimum amount of 5 M NaOH needed to bring the pH to 7.4, then immediately centrifuged at 10,000 *g* for 15 min in a Sorvall RC3C centrifuge. The resulting supernatant was then decanted for assay, after checking that extract osmolality was within an acceptable range (less than 500 mOsm).

2.1.2 Yeast

To investigate the possibility of one or more ACTH-like molecules in yeast, a strain of budding yeast (*Saccharomyces cerevisiae*) was obtained as a kind gift from Dr D. Stillman (Dept. of Cellular, Viral and Molecular Biology, University of Utah). This yeast strain was wild type apart from several selectable auxotrophies, as shown in Table 2. Strain DY150 cells were grown at 30°C in a chemically defined medium (129) to the end of the logarithmic phase of growth. Growth in log phase in defined medium was not significantly different from that in enriched

Table 2. *Genotype of yeast DY150 strain*

MAT	a haploid
Auxotrophies	ura 3-52
	leu 2-3,112
	trp1-1
	his 3-11,15
	ade2-1
	can1-100 (oc)
Outside Source	KAN Y699 [W303]

medium, even though the final cell density in stationary phase was slightly lower (data not shown). The extraction protocol for yeast was identical to that for Tetrahymena except that prior removal of the cell wall was required before the yeast cells (spheroplasts) were extracted in acid. This was achieved by preparation according to one of the protocols detailed below.

2.1.2.1 Enzymatically

After initial harvesting and washing of yeast as described above, the cells were resuspended in a solution of 2 M Sorbitol and incubated at 37°C with intermittent gentle shaking. Digestion (using Yeast Lytic Enzyme; ICN Biologicals; 129) was monitored by observing the percentage of cells lysed by exposure to hypoosmolar conditions, as judged by a reduction in absorbance at 800 nm, compared to control (130). After greater than 75% of the cells had been lysed, they were precipitated by centrifugation, washed twice with distilled water and extracted with glacial acetic acid as above.

2.1.2.2 Grinding under liquid nitrogen

For large volumes of culture, yeast cells were lysed by grinding in a Waring blender under liquid nitrogen (127). After weighing, the harvested cells were resuspended in a minimal volume of distilled water and immediately snap frozen by dropping into liquid nitrogen, at which point they could then be stored at -70°C indefinitely. A small metal grinding chamber was used for homogenization with repeated bursts of 30 sec on speed setting "high", until the popcorn-like frozen yeast lumps became a fine yellow powder. This took five separate bursts per aliquot (50 g). The resulting fine powder was then suspended in glacial acetic acid for extraction by homogenization using the Polytron homogenizer, as described above.

2.1.3 Medium

The extraction procedure for medium was based on the affinity of small peptides for microfine silica, and the ability to reversibly elute them by washing in acidified acetone (131). After centrifugation of cells, medium was harvested by carefully siphoning off the supernatant. To further eliminate any cell debris, the medium was also filtered through Whatmann Number 1 filter paper. It was then mixed overnight with 1% w/v acid-prewashed Silica (QUSO) at 4°C. The silica was then precipitated by brief centrifugation, washed twice with distilled water, and bound peptides eluted using 40% acetone in 0.1 M HCl in two successive extractions. The acetone was evaporated overnight under nitrogen and the aqueous phase then lyophilized. Subsequent steps for the lyophilizate were as described above for the cell extracts.

Varying amounts of reconstituted extracts were assayed for ACTH-like activity in two different polyclonal assay systems (see below), both in the presence and absence of maximal concentrations of protease inhibitors, as recommended by the manufacturers (EDTA, 1 mM; Phenylmethylsulfonyl fluoride, 1 mM; Leupeptin, 1μg/ml; Trasylol, 2%v/v; Pepstatin, 1μg/ml - These inhibitors were all obtained from Sigma Chemical Co., St Louis, MO, except for Pepstatin and Leupeptin, which were obtained from Boehringer-Mannheim)

2.2 Immunoassay Procedures

Two different types of immunoassays were used to assay for ACTH activity in extracts. These are considered in the following discussions.

2.2.1 Polyclonal immunoassays: components

The polyclonal assays were "in-house" competitive immunoassays (78, 109). The components of these polyclonal assay systems were as follows:

2.2.1.1 ACTH1-39 standard

Synthetic human ACTH1-39 standard (Calbiochem, Irvine, CA) was stored at -70°C in 2 µg aliquots at a concentration of 100 µg/ml in 0.05 M HCl (snap frozen in siliconized plastic vials). Stock aliquots were thawed once only for use as either standard or hormone "label."

2.2.1.2 Iodinated (^{125}I) ACTH1-39 "label"

^{125}I (1 mCi/µl) was obtained from New England Nuclear, Boston, MA, or Amersham Corp., Arlington Hts, IL. Iodogen and sodium metabisulfite were obtained from Sigma Chemical Co, St. Louis, MO. Iodination was performed using the iodogen technique (132), because the gentle nature of this oxidizing agent minimized damage to the peptide, especially avoiding oxidation of Met^4 to the sulfoxide (133, 134). Iodogen (1 µg) was dissolved in heptane and coated onto the bottom of siliconized glass tubes by evaporation of the solvent. The tubes were then stored dessicated at 4°C in the dark. Iodination was performed by adding 1 µl of ^{125}I to 2 µg of synthetic ACTH1-39 in 20 µl of 0.01 M phosphate buffered saline (i.e.150 mM NaCl/ 0.01 M sodium phosphate solution pH 7.4; stored at -20°C) and 70 µl of 0.5 M phosphate buffered saline to provide extra buffering capacity. These were mixed gently in the iodogen tube by gentle rotation for 90 sec, after which 10 µl of 1 mg/ml sodium metabisulfite was added to quench the reaction and the products were then separated by gel filtration in 0.1 M HCl over a 10 cm Sephadex G10 column (Sepadex from Pharmacia Co., Uppsala, Sweden) which had been presiliconized and saturated with bovine serum albumin (RIA grade; Sigma). Complete separation between the organic and inorganic peaks could be achieved using this size of column. Specific activity of 250 µCi/µg was routinely obtained. Fractions corresponding to

labelled ACTH (first organic fraction) were pooled, mixed with an equal volume of methanol and stored at -20°C. This preparation remained stable for over 2 weeks.

2.2.1.3 Antibodies

Two different polyclonal antibodies were used: i) West antibody (raised in rabbit), which recognizes principally the mid (amino acids 11-17) region of vertebrate ACTH1-39 (70), and ii) Orth antibody (raised in rabbit) with specificity principally for the N-terminal half of vertebrate ACTH1-39 (α-MSH). Appropriate titer was determined by serial dilution of antibodies to determine specific binding at different dilutions.

2.2.2. Polyclonal immunoassays: assay conditions

The extracts were assayed in triplicate in siliconized 100 x 7 mm polypropylene tubes (Intermountain Scientific, Salt Lake City, UT). Constituents (in final assay volume of 1 ml) were:

a) ^{125}I-labelled ACTH1-39, 1500 cpm/tube.

b) Antibody at final dilution either 1:80,000 (West), or 1:500,000 (Orth).

c) Serial dilutions of extract, or unlabelled ("cold") ACTH1-39 standard (from 1 μg/ml to 1 pg/ml).

d) Assay buffer (0.01 M Phosphate buffered saline; 0.2% β-mercaptoethanol; Sigma) to final volume of 1 ml.

Tubes were incubated for 48 h at 4°C, after which 100 μl of 1:5 dilution of goat anti-rabbit antiserum (made by Dr R. Henninger, Brigham Young University, Provo, Utah) and 250 μl of 20% polyethylene glycol (average molecular weight 8000; Sigma) were added. After a further 4 h incubation at 4°C, 100 μl of outdated filtered human plasma (obtained with permission from the University of

Utah Medical Center Bloodbank) was added, and bound versus free counts were immediately separated by centrifugation at 2800 rpm for 20 min at 4°C in a tabletop centrifuge. Tubes were counted for 5 min in a gamma spectrometer. This assay system had a sensitivity of approximately 10 pg/ml, with intra-assay coefficient of variation below 5% and interassay coefficient of variation of 9%.

2.2.3 Monoclonal immunoassay

This was a commercial assay kit developed by Nichols Institute (San Juan Capistrano, CA.) and has been described previously (135). Briefly, the antigen is "sandwiched" between a polyclonal antibody bound by a tight avidin-biotin linkage to a polystyrene bead and an ^{125}I-labelled monoclonal antibody recognizing the C-terminal region of ACTH. In contrast to the previous competitive assay, therefore, increasing antigen is reflected by an increasing number of counts attached to the bead. This assay shows considerably greater specficity than the polyclonal assays described above, because there are much more stringent requirements for hormone recognition. To react in the assay, an antigen must possess epitopes recognized by two different antibodies as well as having a conformation enabling it to be "sandwiched" between them (136).

2.3 Results of Immunoassays

2.3.1 Polyclonal immunoassays

2.3.1.1 Tetrahymena

Initial immunoassay data for extracts of *Tetrahymena pyriformis* at the end of the logarithmic phase of growth in enriched medium (NEFS medium) showed much higher concentrations of ACTH-like immunoactivity in the medium, as compared with relatively small amounts in the cells, even in the presence of inhibitors (data not shown). This somewhat surprising result prompted me to

assay the individual medium components. Both Proteose Peptone and Yeast Extract (Gibco) showed significant immunoreactivity, equivalent to 11.7 and 2.7 μg/g dry weight respectively. Proteose peptone is derived from vertebrate tissues and is therefore almost certainly contaminated by vertebrate ACTH. However, while the presence of large amounts of activity in the Yeast Extract also suggested an immunoreactive species, the manufacturer (Gibco) claimed proprietary protection in not divulging whether this preparation was also contaminated by vertebrate tissue sources.

To avoid the possibility of medium contamination by vertebrate ACTH, Tetrahymena were regrown in chemically defined media as previously described under methods. The cells from 8 L of broth were harvested and extracted at the end of log phase. Substances with ACTH-like activity were extracted from the medium (4L) using microfine silica, with cells and cell debris being excluded by both careful aspiration of the supernatant after centrifugation and also passing the medium through a 0.45 μm filter (Millipore). The results of this analysis are seen in Figures 4 and 5. Glacial acetic extracts of the cells showed a dose-response pattern in the immunoassay, which was present with both polyclonal antibodies, directed principally at different parts of the ACTH molecule. Back-calculation gave a concentration of aproximately 250 pg/g wet weight of organisms using the antibody against the mid-portion of ACTH, assuming the immunoreactive substance(s) have similar affinity for the antibodies as does the ACTH standard. This concentration approximated that found by LeRoith et al. (70) of 0.16 pmol/g wet weight - if one assumes that the immunoreactive species had similar molecular weight(s) to vertebrate ACTH (4.5kD) their stated figure is equivalent to 700 pg/g.

Extracts from the medium also showed a small amount of immunoreactivity, although this was extremely small when expressed on a concentration basis.

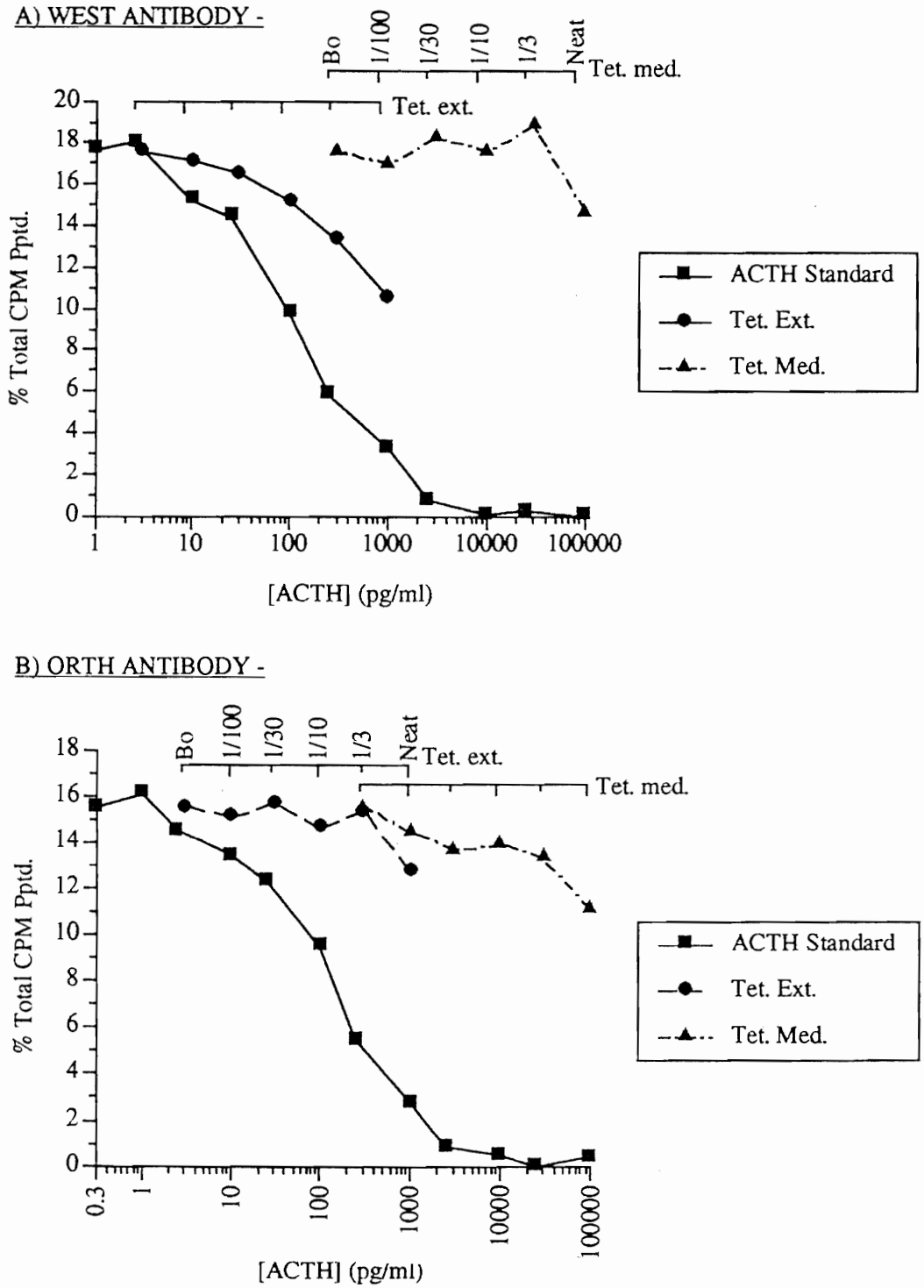


Fig. 4. Polyclonal assay of Tetrahymena extract/medium -no inhibitors. Dose-response curves for standard, extract and medium are as shown, with percent total counts precipitated (y axis) plotted against final concentration of ACTH (x axis), or at progressive dilutions of extract or medium.

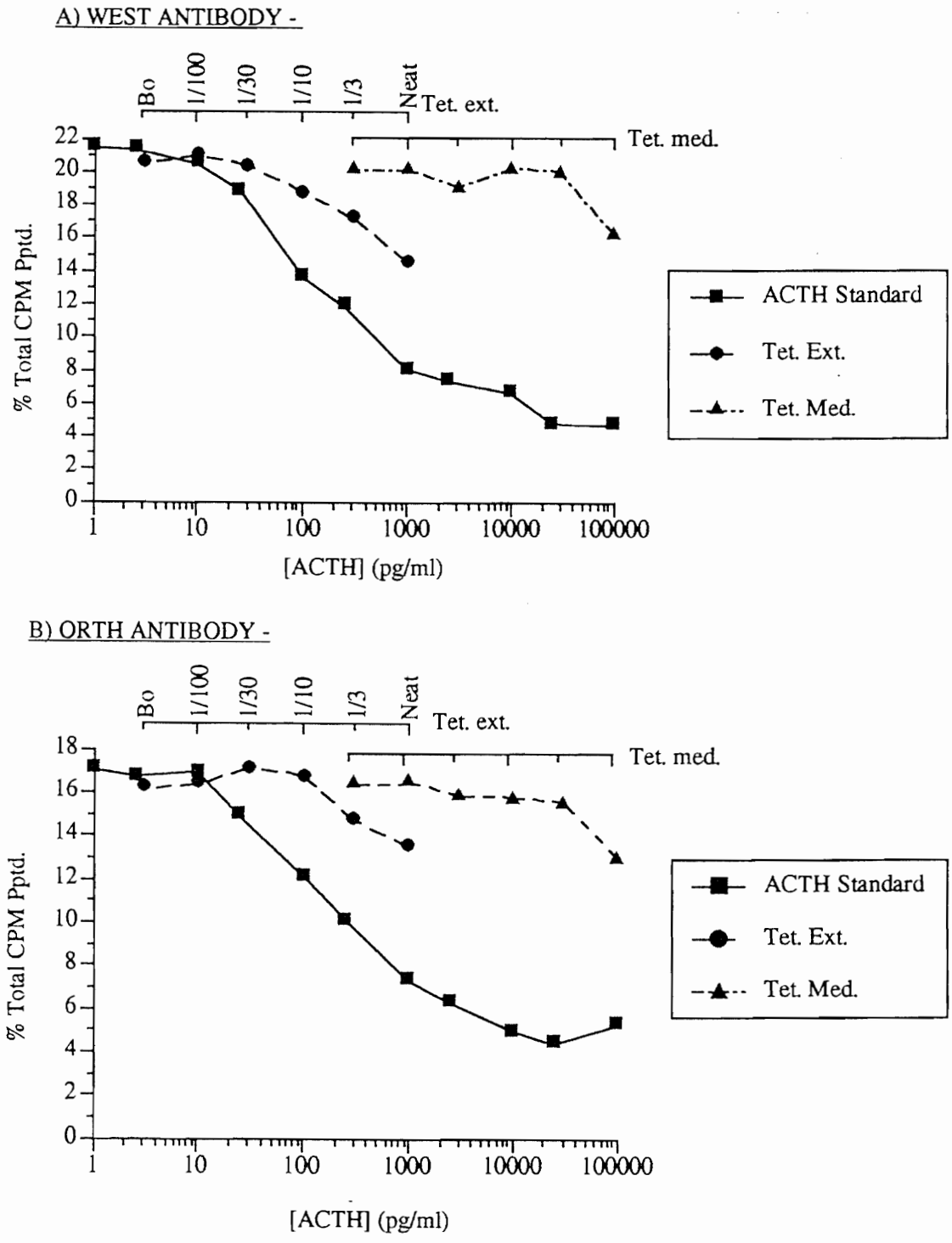


Fig. 5. Polyclonal assay of Tetrahymena extract/medium -with protease inhibitors. (1mM EDTA; 1mM PMSF; 2% v/v aprotonin; 1µg/ml leupeptin; 1µg/ml pepstatin). Dose-response curves with antibodies directed mainly against the mid-portion (A) and amino-terminal portion (B) of vertebrate ACTH 1-39 are shown. The assay of serial dilutions of extracts and medium is also shown.

However, medium alone (in which no cells had been grown) was also extracted after incubation under identical conditions and showed no immunoreactivity whatever (data not shown), thereby suggesting this was not the result of contamination from any medium component or from the glassware. Overall, there was no significant increase in immunoactivity seen when the extractions and assays were performed in the presence of maximal concentrations of multiple protease inhibitors, suggesting that proteolytic degradation was neither reducing the amount of immunoreactive species present in the extracts, nor was it resopnsible for giving spurious results from degradation of label. Later specific experiments to examine this issue confirmed these conclusions (section 2.4).

Table 3 shows derived concentrations for ACTH-like immunoreactivity from one representative experiment using 8 L of cultured Tetrahymena, assayed in the presence and absence of protease inhibitors.

2.3.1.2 Yeast

The fact that the proprietary preparation of Gibco Yeast Extract contained significant immunoreactivity prompted the investigation of whether immuno-reactive ACTH-like substances were present in yeast cells themselves. Yeast

Table 3. *ACTH-like material in extracts of Tetrahymena*

		Extract (pg/g)	Medium (pg/g cells)	Medium (pg/ml)
Midportion antibody.	No inhs.	250	60	0.3
(West)	Inhs.	230	150	0.9
N-terminal antibody	No inhs.	80	130	0.8
(Orth)	Inhs.	170	230	1.0

cells were therefore grown in defined medium to saturation, extracted in glacial acetic acid and assayed as previously described under methods. The results are shown in Figures 6 and 7. Increasing amounts of extract again resulted in a depression of counts precipitated, in a dose-response relationship. Calculation of the amount of ACTH-like material in the yeast extracts showed an initial concentration of approximately 210 and 123 pg/g wet weight of yeast cells (prior to spheroplast formation) using mid-portion and N-terminal portion antibodies to ACTH respectively, assuming that the antigen(s) were of similar size to 4.5 kD vertebrate ACTH and that they showed similar affinity for the antibodies. Less activity reacting with the N-terminal antibody was present in both extracts, with one showing less than 10% displacement below the B_0 value. Medium also showed no measurable ACTH-like activity using either antibody. Again, the addition of high concentrations of protease inhibitors did not significantly affect the overall results, indicating that significant degradation of ACTH-like immuno-activity was unlikely to have occurred during the extraction process and that proteolytic degradation of ^{125}I-labelled ACTH was not spuriously causing the depression of counts by the extract (see below). Details of concentrations of material in extracts and medium are shown in Table 4.

Table 4. *ACTH-like material in extracts of yeast*

		Extract 1 (pg/g)	Extract 2 (pg/g)	Medium. (pg/ml)
Midportion antibody. (West)	No inhs.	210	120	ND
	Inhs.	200	80	ND
N-terminal antibody (Orth)	No inhs.	120	ND	ND
	Inhs.	130	40	ND

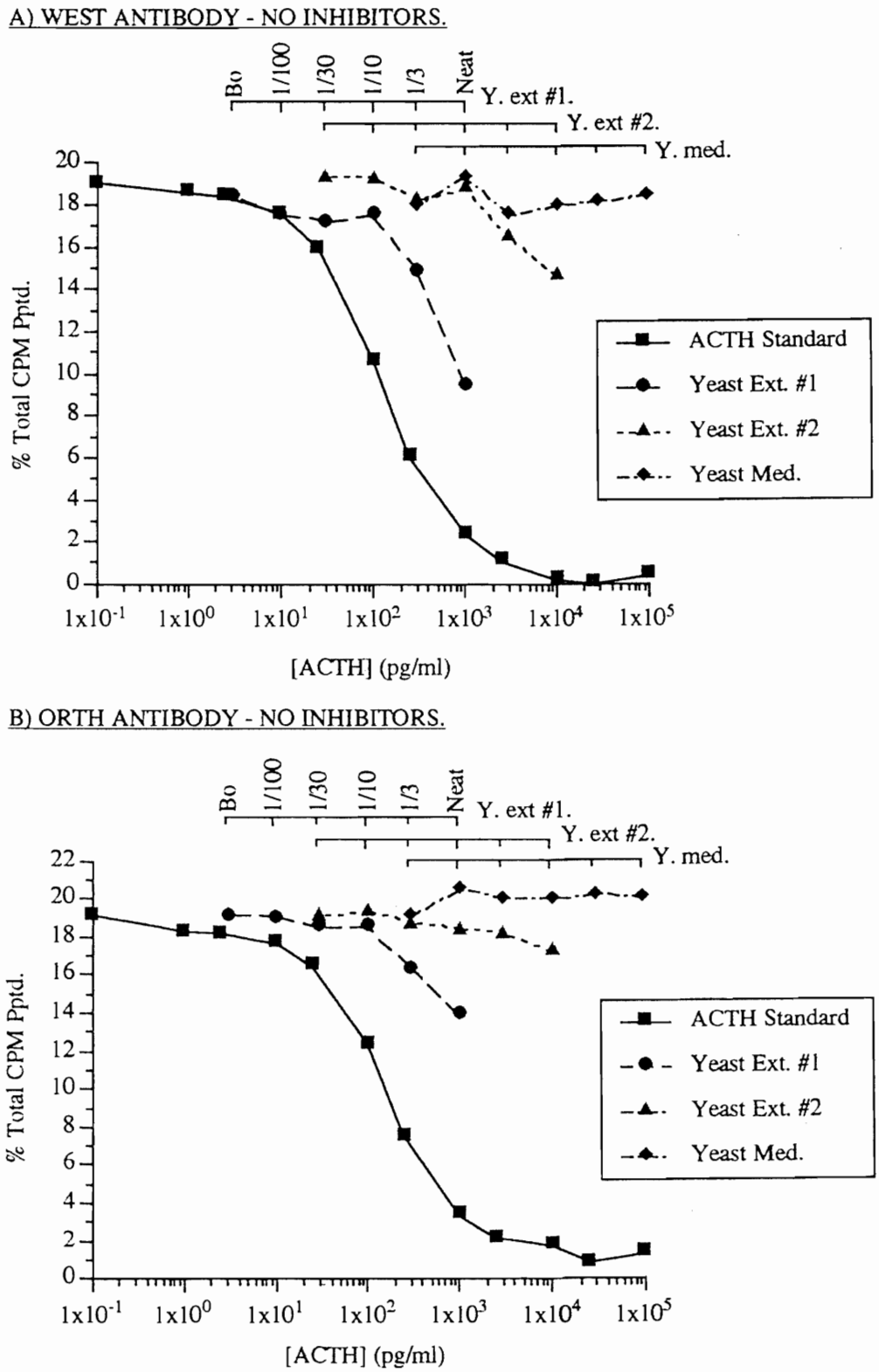
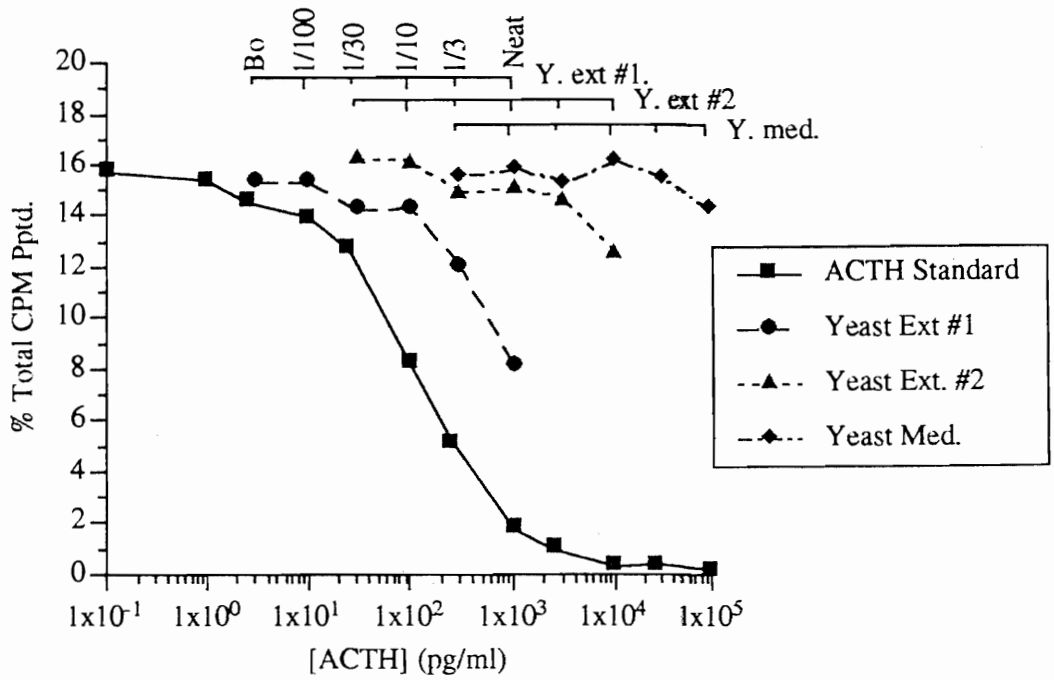


Fig. 6. Polyclonal assay of yeast extract/ medium - no protease inhibitors. Dose-response curves using antibodies directed mainly against the mid-portion (A) and amino-terminal portion (B) of vertebrate ACTH 1-39 are shown. Assay of serial dilutions of extracts and medium is also shown.

A) WEST ANTIBODY



B) ORTH ANTIBODY - WITH INHIBITORS.

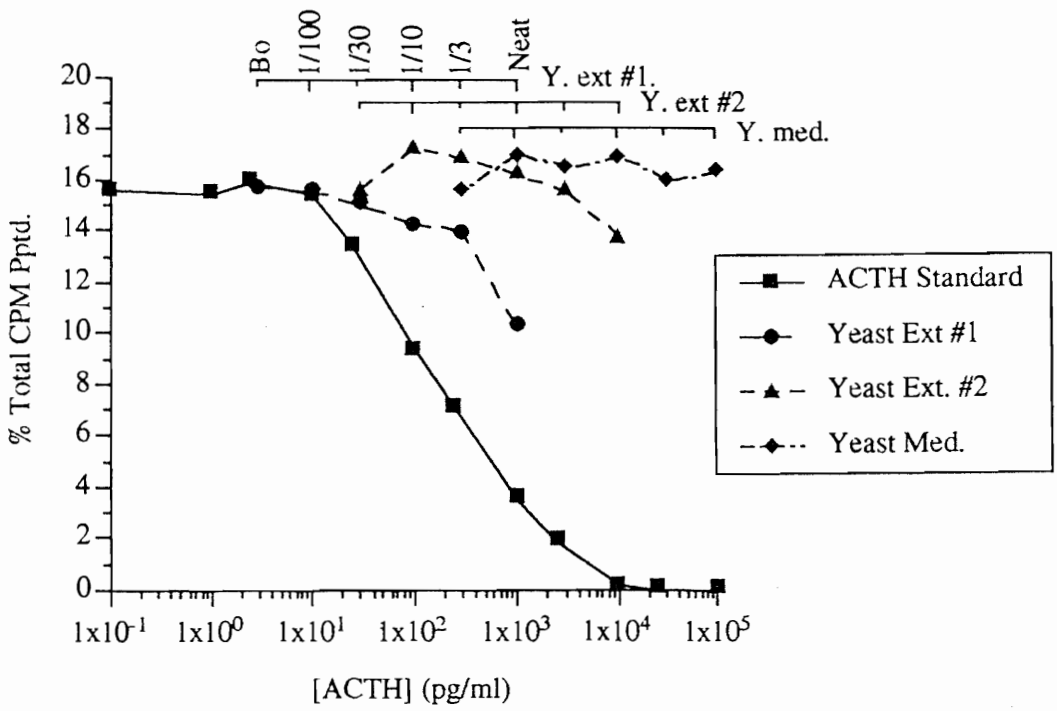


Fig. 7. Polyclonal assay of yeast extract/ medium - with inhibitors (1mM EDTA; 1mM PMSF; 2% v/v aprotonin; 1µg/ml leupeptin; 1µg/ml pepstatin). Dose-response curves using antibodies mainly against the mid-portion (A) and amino-terminal portion (B) of vertebrate ACTH1-39 are shown. Assay of serial dilutions of extracts and medium is also shown.

2.3.2 Monoclonal immunoassays

The same Tetrahymena and yeast extracts and media were also assayed using the two-monoclonal-antibody sandwich assay (immunoradiometric assay, or IRMA) previously described under methods (Section 2.2.2). Both assays were set up simultaneously to avoid the need to freeze and thaw the extracts. The results of this experiment are shown in Figure 8. In this antibody system neither extracts nor media showed any reaction whatever, indicating that whatever substances had reacted in the polyclonal assays did not have enough similarity in primary sequence or in overall conformation to react in this more stringent monoclonal sandwich assay system. No attempt was made to utilize the monoclonal antibody component separately to assay the immunoreactive material, because the generally low affinity of monoclonal antibodies for the ligand when used alone (and the resulting decrease in sensitivity; 137) would have almost certainly resulted in failure to detect the very small amounts of antigen indicated by the polyclonal assay. However, the antigenic material was definitely not retained by an anti-ACTH(1-10) immunocolumn utilizing a different monoclonal antibody (Medix Biotec) conjugated to Sepharose. In this case affinity of the antibody should not be a factor because the antibody was in significant excess, indicating that primary sequence differences as well as differences in overall conformation were almost certainly responsible for lack of binding.

2.4 Experiments to Exclude Artifacts

Several potential sources of artifact were addressed. Possible contamination of cell extracts with exogenous ACTH was minimized by the use of new, siliconized and carefully washed glassware. All chemicals were from previously unopened containers, set aside for such use. The effects of pH and osmolality were excluded by measuring these parameters in the extracts and

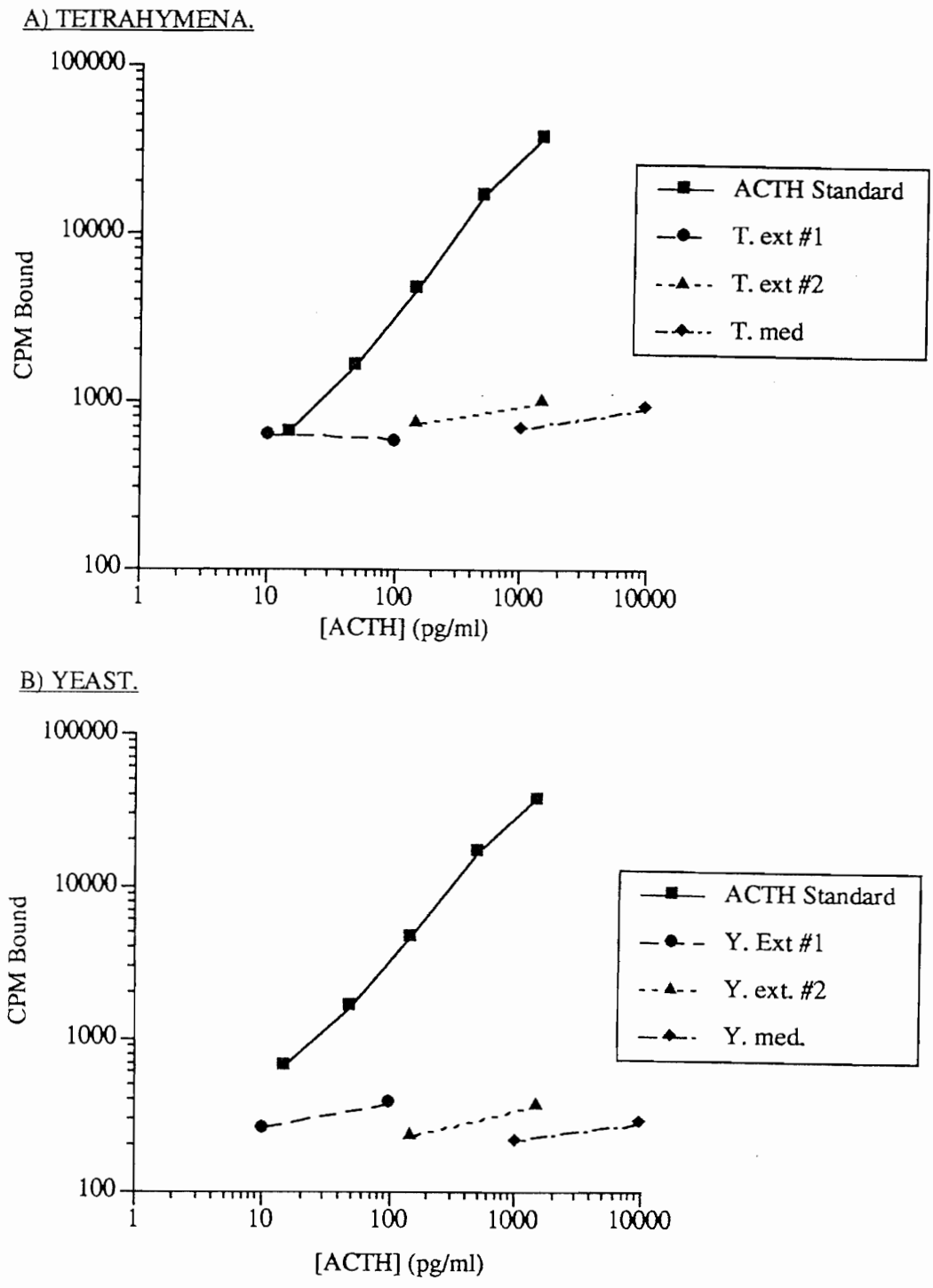


Fig. 8. Monoclonal immunoassay of extracts and medium. A) Tetrahymena extracts; B) Yeast extracts. Graphs show total counts/minute bound to polystyrene bead (y axis) versus final ACTH standard concentration (pg/ml). Neat (right) and a ten-fold dilution (left point) of extracts and medium are shown. None showed measurable activity

adjusting appropriately if necessary. Extraction and immunoassay in the presence of maximal concentrations of inhibitors failed to abolish the extract dose response curves, indicating proteases were unlikely to be responsible. Linear transformation of the extract dose-response curves by log-logit transformation (Figures 9 and 10) showed slopes within the 95% confidence limits of the standard curves, although in some cases the low amount of immunoactivity resulted in inaccuracies because of the exaggerated effects of small variations close to the B_0 value.

Further exclusion of the possibility of proteases causing artifactual reaction in the ACTH immunoassay was achieved by the following. First, ^{125}I-labelled ACTH was incubated with extract for 2 days at 4°C, either in the presence or the absence of an excess (1/3000 final dilution) of West polyclonal antibody. After 2 days incubation, excess antibody was added to the tubes previously without antibody and the incubation was continued for another 2 days in the cold (4-7°C), followed by a second antibody incubation step in all tubes (100µl of a 1:5 dilution of goat anti-rabbit antibody), and precipitation of bound ^{125}I-ACTH as described above. The results of this experiment are shown in Figure 11A. Pre-exposure of the ^{125}I-labelled ACTH to extract or medium did not reduce the number of counts precipitated as compared with control tubes in which the label was "protected" by antibody throughout; thus no significant degradation of the label occurred under conditions identical to the assay. In a second similar experiment, the elution profile of ^{125}I-labelled ACTH when run through a 25 cm Sephadex G10 column was compared before and after two days of incubation with extract (Figure 11B). No significant difference in the elution profile was seen, with specifically no increase in fractions corresponding to smaller breakdown products of the 4.5 kD peptide (data not shown).

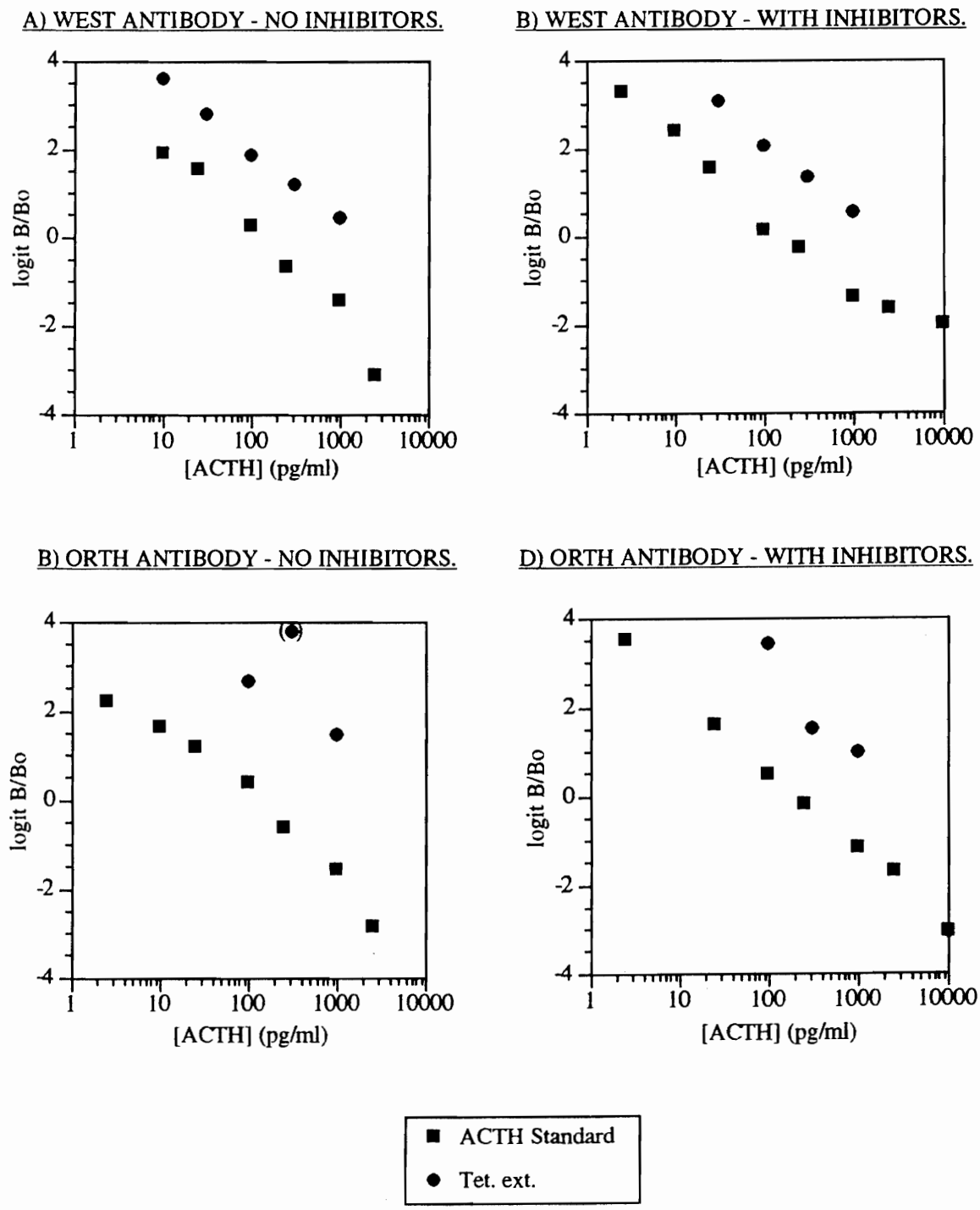


Fig. 9. Log-logit plots of Tetrahymena dose-response curves. Dose-response curves from Figures 4 and 5 were modified by logittransformation (78). Assays using both mid-portion ACTH antibody (West) and N-terminal ACTH antibody (Orth) are shown, with and without the presence of protease inhibitors.

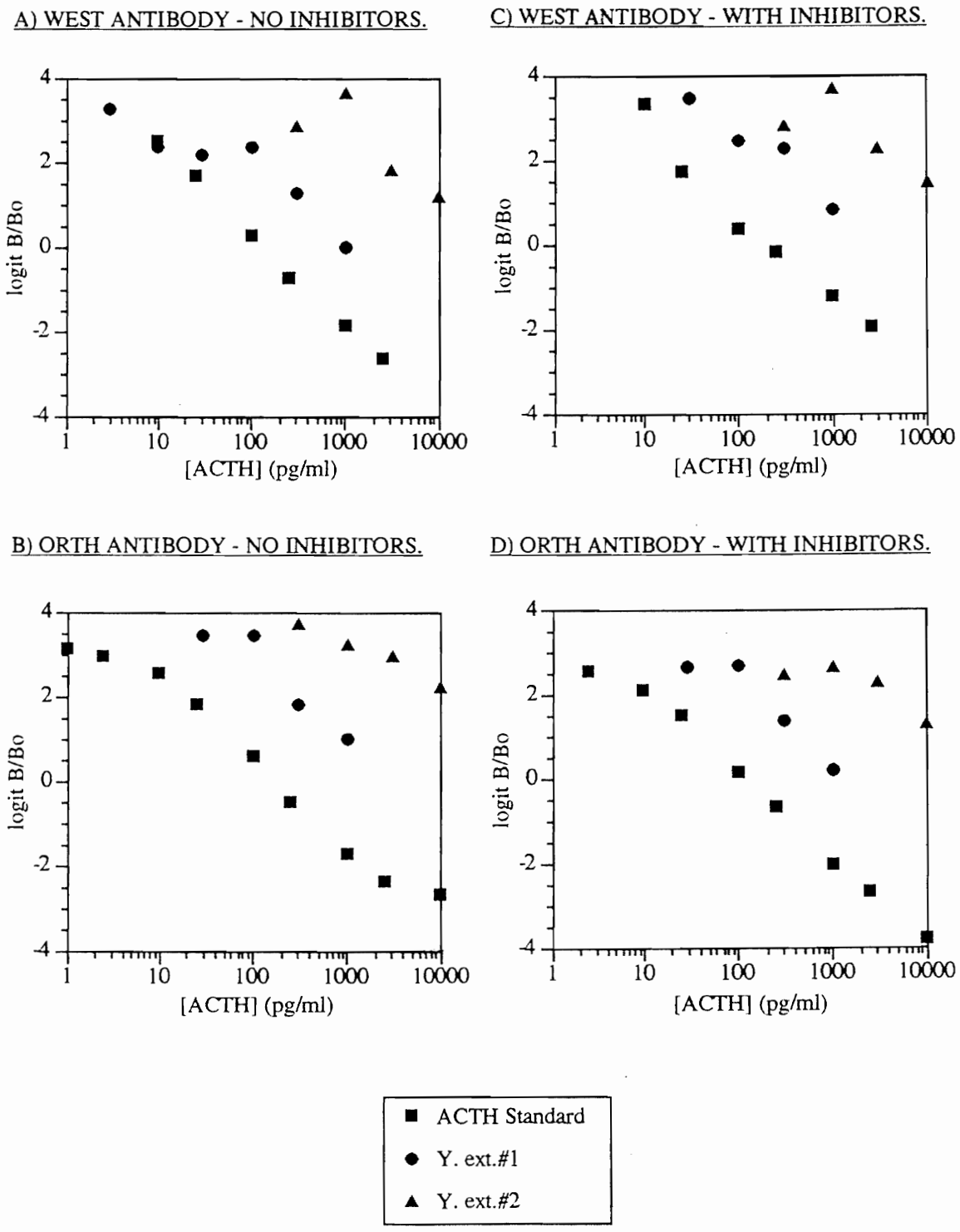
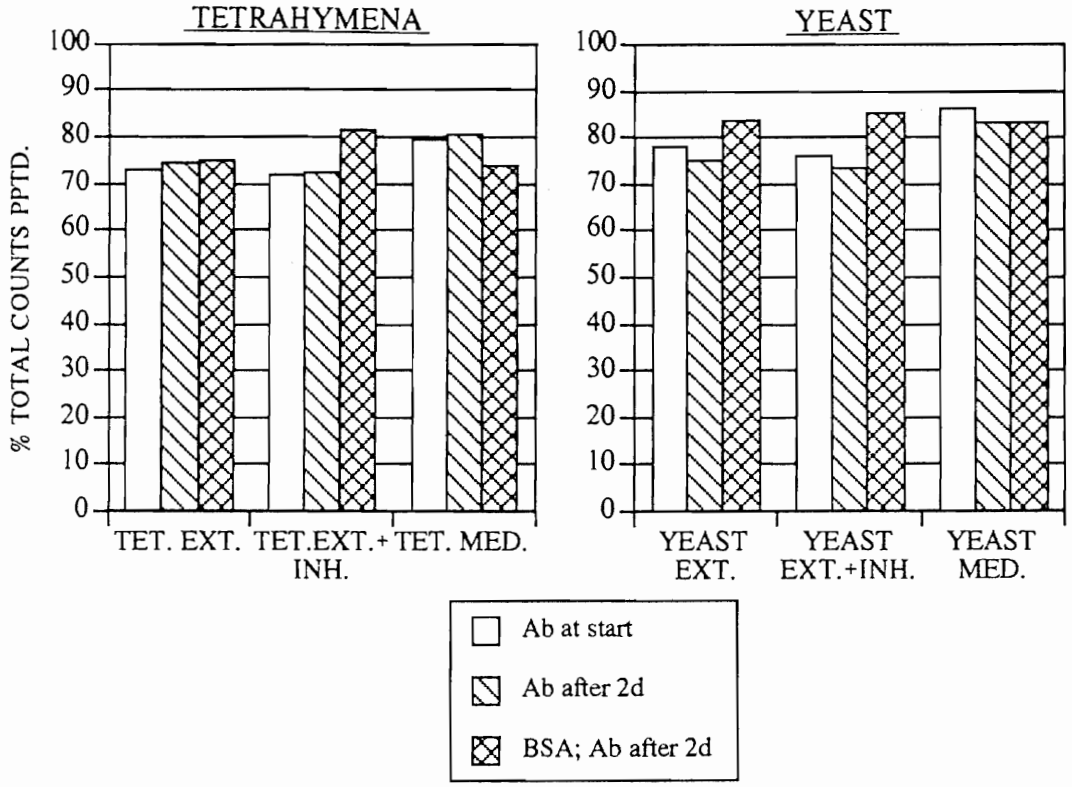


Fig. 10. Log-logit plots of yeast dose-response curves. Curves from Figures 6 and 7 were modified by logit transformation of the data (78). Assays using both the mid-portion ACTH antibody (West), and N-terminal ACTH antibody (Orth) are shown, with and without the presence of protease inhibitors.

A) ANTIBODY PROTECTION EXPERIMENTS.



B) ^{125}I-ACTH ELUTION PROFILE AFTER 2 DAY INCUBATION.

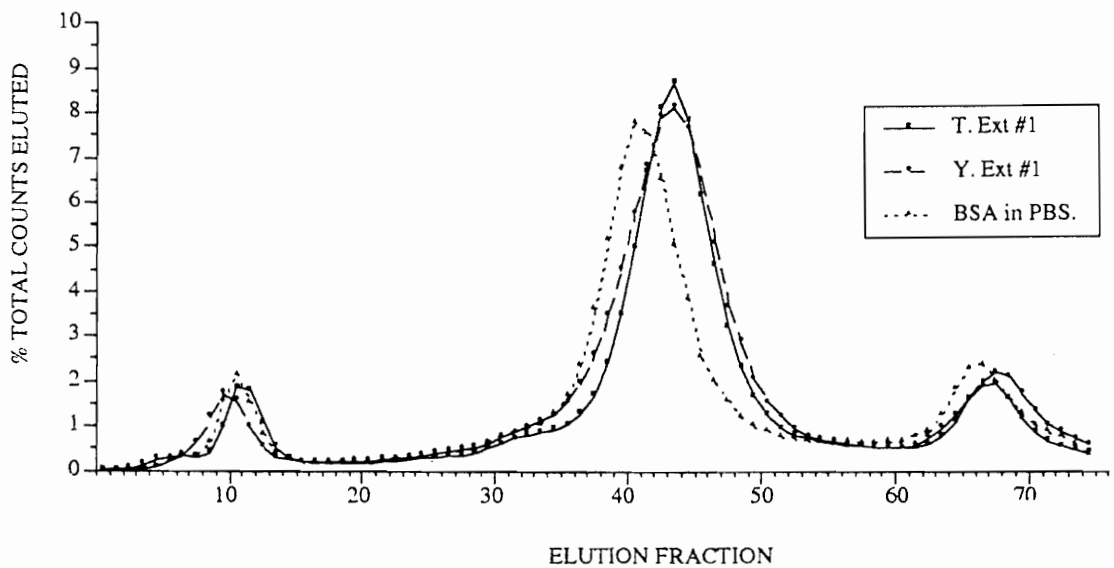


Fig. 11. Experiments to exclude protease artifacts. A) Antibody protection experiments. ACTH label was incubated for 2days at 4-7C in the presence of extract, extract plus excess West antibody (1:3000f.d.), and bovine serum albumin (BSA); immunoreactivity was then measured. B) G75 elution profile was compared after incubation with extracts, vs BSA control.

2.5 Conclusions from Immunoassay Experiments

2.5.1 Structural similarity to vertebrate ACTH

Polyclonal immunoassay experiments suggested that one or more substances resembling vertebrate ACTH1-39 were present in Tetrahymena and budding yeast, whereas no reaction was found in more stringent two-monoclonal antibody sandwich assays. These results suggested molecules were present with at least a similarity in surface profile with parts of the vertebrate ACTH molecule. The most important recognition regions for both polyclonal antibodies were known and suggested this shape resemblance included both the more N-terminal α-MSH region (residues 1-14 of ACTH) and the mid-portion (residues 11-24). Although this is potentially useful in oligonucleotide design, it is noteworthy that primary sequence homology is not necessary for antibody cross-reaction (80, 81; see section 2.2.1.1), and that many, if not most, antibodies recognize a surface epitope formed by several spacially adjacent but linearly separated parts of the molecule (79). Indeed, the lack of reaction with either a single monoclonal antibody directed at the α-MSH region, or with the two-monoclonal-antibody sandwich assay suggested that a) any primary sequence homology in the α-MSH region may be relatively weak and, b) overall similarity in shape was also probably weak, compared to the vertebrate 39 amino acid molecule.

2.5.2 Function in the cell

The small amount of immunoactivity present was consistent with previous findings of other authors in Tetrahymena (61) and indicated that this material probably constitutes only a very small fraction of total cellular protein, assuming reactivity with the antibodies on an equimolar basis as the vertebrate peptides. However, the insignificant amount (undetectable in yeast) of ACTH in the medium also suggested that whatever intracellular immunoactivity was present

was not released into the medium in a manner similar to vertebrate ACTH. Nevertheless, this conclusion relied on the efficiency (sensitivity) of the extraction technique, and also that the reacting substances were not so labile that they would be broken down during extraction. Whereas microfine silica is a very efficient way of extracting many small vertebrate peptides, the relative extraction properties of the immunoactivity in unicellular organisms, or even whether it represented a small peptide, are not known.

The low concentration suggested that the protein is more likely to fulfill a messenger or metabolic role rather than a structural one. Unfortunately, it also meant that isolation of sufficient quantities of the protein for sequence analysis would require large scale purification procedures and be technically difficult; in the absence of microanalysis techniques, the very large amounts of material required would sway one away from large scale protein purification as an intermediate step. To quote one prominent worker in the field faced with the same problem, "Since... levels in microorganisms are low, extraction and purification to homogeneity of sufficient for amino acid sequencing seems a formidable task. Recombinant DNA technology seems the most promising route" (54).

CHAPTER 3

SOUTHERN ANALYSIS OF GENOMIC DNA

3.1. Preparation of High Molecular Weight Genomic DNA

3.1.1 DNA extraction

3.1.1.1 Rat

High molecular weight rat genomic DNA was prepared as previously described, with some minor modifications (12). Livers from adult Sprague-Dawley rats were snap frozen in liquid nitrogen immediately after sacrifice, and stored at -70°C until use. Weighed frozen liver (generally about 4 g) was ground to a fine powder with pestle and mortar under liquid nitrogen, then mixed slowly with 10 volumes of lysis buffer (100 mM TrisHCl pH 8.0, 100 mM NaCl, 1 mM EDTA). Sarcosyl (Mallinkrodt) was added to 10% v/v and RNAase A (Promega Biotec) to 100 µg/ml (f.c.), and the suspension was stirred slowly overnight in the cold (4-7°C). It was then extracted with phenol (x2), 50:50v/v phenol: chloroform (x2), and with chloroform alone (x2), and the DNA remaining in the aqueous phase was then precipitated in 70% (f.c.) ethanol. The precipitated DNA mix was spooled out, resolubilized in 20 mM TrisHCl, 1 mM EDTA pH 8.0 (this often took gentle agitation overnight at 4°C), and examined for both size and purity by gel electrophoresis in 0.7% SeaKem agarose and 260/280 nm absorbance ratio. Any residual RNA and protein contamination was removed by a further round of RNAase A digestion (concentrtion as before), re-extraction in phenol: chloroform (50:50 v/v) and chloroform, and finally ethanol precipitation.

3.1.1.2 Tetrahymena

Cells were incubated in enriched medium to the end of log phase, harvested and whole cell genomic DNA (macro- and micronuclear) was extracted as previously described (138). Precipitated cells were resuspended in 10 volumes of lysis buffer (containing 100 µg/ml proteinase K; Boehringer-Mannheim; Indianapolis, IN) and incubated for 30 min at 65°C. The mixture was extracted with phenol (x2), phenol:chloroform (50:50 v/v)(x2) and chloroform (x2), and precipitated with 70% ethanol. The DNA/RNA precipitate was spooled out, and after redissolving in 20 mM TrisHCl, 1 mM EDTA, it was digested with 100 µg/ml RNAase A for 60 min at 37°C. After a further round of phenol:chloroform (50:50 v/v) and chloroform extraction, the DNA was precipitated with an equal volume of isopropanol at room temperature and redissolved in TE.

3.1.1.3 Yeast

DNA was prepared according to the protocol of Philippsen et al. (130) with minor modifications. Cells were grown to the end of log phase, harvested, and digested with Yeast Lytic Enzyme as previously described (section 2.1.2.1). Spheroplasts were lysed at 65°C for 30 min in 50 mM Tris-HCl, 50 mM EDTA, pH 8.0 containing 1% sodium dodecyl sulfate (f.c.). Most protein was initially removed by precipitation in 1.5 M potassium acetate for 60 min on ice, followed by centrifugation at 11,000 *g* for 30 min. The DNA and RNA in the supernatant were then precipitated with 70% ethanol, resuspended in TE, and digested with RNAase A (100µg/ml f.c.) at 37°C for 30 min. After further phenol:chloroform (50:50 v/v) and chloroform extraction, the genomic DNA was precipitated with isopropanol (50% f.c.) and resuspended in 10 mM Tris-HCl, 1 mM EDTA pH 8.0. Each genomic DNA preparation was checked for its purity and molecular size, by both gel electrophoresis and measurement of relative absorbance at 260/280 nm.

3.1.2 Enzyme digestion and size separation

Southern blotting was carried out as described in Sambrook (125), with minor modifications. Restriction enzymes (Eco R1, Hind III, Xba 1, Sac 1, and Hinf 1) were purchased from Boehringer-Mannheim. Genomic DNA (final concentration 100-200 μg/ml with average size over 20 kb) was digested with a 5 fold excess (5 U/μg DNA) of restriction enzyme for at least 6 h at the recommended incubation temperature (generally 37°C). Volume of enzyme was kept below 10% of final reaction volume to avoid star activity of some enzymes, such as Eco R1. Digests were terminated by incubation for 15 min at 65°C and subsequently by phenol: chloroform (50:50 v/v) and chloroform extraction. Digested DNA was precipitated in an Eppendorf microfuge for 15 min at 4°C in 70% ethanol (f.c.) and 0.3 M sodium acetate (f.c.). Precipitates were washed once with cold 70% ethanol and resuspended in 10 mM TrisHCl, 1 mM EDTA pH 8.0. In some cases incubation for several hours at 37°C was necessary to resolubilize the DNA.

Digested genomic DNA samples were routinely electrophoresed overnight at approximately 50 V (2.5-3 V/cm), 25 mA, in 0.8% agarose (Sea Kem; FMC Bioproducts) using TrisHCl/sodium acetate/EDTA (TAE) buffer, as previously described(125). To visualize DNA, ethidium bromide (1 μg/ml, f.c; Sigma) was added to the agarose gel prior to electrophoresis, as well as dispersed through the running buffer on the anodal side of the gel. DNA (10 μg) was loaded in 5:1 v/v ratio with loading dye (0.25% bromophenol blue, 0.25% xylene cyanol, 30% glycerol). Restriction digests (Hind III and Bst EII) of wild type phage lambda DNA (Bethesda Research Laboratories, Baltimore, MD) were used as size markers, giving fragment sizes ranging from 23 kb to 0.5 kb. Size markers were labelled with ^{32}P by filling in the recessed 3' end with Klenow DNA polymerase (Boehringer-Mannheim) using ^{32}P-dATP, as described (125).

3.2 Transfer and Immobilization

Following electrophoresis and staining, DNA in agarose gels was treated according to the following steps (130, 139):

1. Incubation for 15 min in 0.2 M HCl with gentle agitation.
2. Incubation in 0.5 M NaOH/1.5 M NaCl (2 x 30 min) to denature DNA.
3. Neutralization by incubation in 1 M TrisHCl (2 x 30 min)
4. Equilibration in the buffer used to transfer DNA onto membranes.

DNA was transferred onto hybridization membrane by a wick process from a sponge saturated in transfer buffer through the gel and into a stack of blotting paper (Intermountain Scientific). Although other transfer methods were tried, this process resulted in the most efficient transfer of DNA (data not shown).

3.2.1 Optimization of transfer buffer and membrane

To determine sensitivity of the Southern blotting procedure and optimize DNA transfer, several different transfer parameters were compared. Sensitivity of the system was assessed by the ability to detect different concentrations (1 ng - 1 pg) of the cloned, linearized (Eco R1-digested) rat POMC gene (courtesy J. Drouin, Montreal) against a background of 10 μg of similarly digested rat genomic DNA in each lane of a Southern blot. A long (145 base) completely homologous nondegenerate synthetic oligonucleotide corresponding to the β-endorphin region of the rat gene (140) was used as a probe. This single stranded probe was 3' end-labelled with α^{32}P-dATP using terminal deoxynucleotidyl transferase (Pharmacia), to a specific activity of at least $3x10^8$ cpm/μg, and purified from unincorporated ^{32}P by size fractionation (using the Push column technique; Stratagene). Experiments showed approximately 20-30 extra dATP molecules were added to the 3' end of the probe during the labelling process, effectively increasing the background nonspecific binding.

Preliminary experiments were aimed at establishing the optimal transfer protocol for maximizing sensitivity, with the following transfer parameters being compared:

a) Transfer buffer

i) 10X SSC, pH 8.0 (1X SSC = 150 mM sodium chloride, 15 mM sodium citrate)

ii) 1X SSC, pH 8.0, and

iii) 1M ammonium acetate; 0.02 M NaOH, pH 8.3.

b) Hybridization membrane

i) Charged nylon (Hybond-N+; Amersham)

ii) Uncharged nylon (MSI), and

iii) nitrocellulose (Hybond-C; Amersham).

Charged nylon gave the strongest hybridization signal under most transfer conditions, although the benefit over uncharged nylon was generally small. Low salt (1X SSC) buffer gave the optimal transfer, as found previously by Reed and Mann (141), and contrasting with Sambrook (125). Transfer in 1X SSC onto charged nylon was therefore used in all subsequent Southern blot procedures. Although signal strength was lower using nitrocellulose, the background was also significantly lower, particularly at low hybridization stringencies when oligonucleotides were used as probes (142). For screening of phage plaques, therefore, nitrocellulose filters (Schleicher and Schuell; Keene, NH) were routinely used since the DNA in individual phage plaques was in high concentration and sensitivity was less critical; filters were generally screened two or three times, making the greater fragility of nitrocellulose less important.

3.2.2 Optimization of crosslinking technique

Several different methods of fixing DNA onto membranes were compared. These included i) drying in air at room temperature, ii) baking at 80°C for 2 h, iii)

thymidine dimer crosslinking to the wet membrane by ultraviolet light, and iv) short exposure of the membrane to alkali (0.4 M NaOH). Two different strengths of UV exposure were compared (1200 and 2400 $\mu J/cm^2$), because previous literature suggested that overexposure disrupted the DNA structure, resulting in a decrease in hybridization signal (143). All methods, including drying at room temperature produced a measurable hybridization signal, but UV crosslinking and baking was optimal, with sensitivity of detection of 30 pg, equivalent to approximately 100,000 copies of the cloned rat POMC gene (data not shown). Exposure of the membrane to twice the recommended amount of UV light resulted in an unexpected fuzziness rather than a diminution of signal. For subsequent experiments, DNA was crosslinked by 1200 $\mu J/cm^2$ of UV exposure, followed by baking.

3.3 Construction and Labelling of DNA Probes

3.3.1 Oligonucleotides

Tables 5 and 6 give details of synthetic probes used in the Southern blots. Tables of codon bias in Tetrahymena and yeast were used to design biased oligonucleotide probes (144, 145). The nonspecific base inosine was also used in later probes in order to reduce degeneracy in codons with three or more codon choices (145, 146). Oligonucleotides were synthesized by the phosphoramidite method (147) and purified by butanol extraction and electrophoresis in a 20% polyacrylamide gel. Oligonucleotides were labelled by either i) 5'-end labelling with T4 polynucleotide kinase (Promega Biotec, Madison, WI), using γ^{32}P-dATP (NEN) in molar ratio of 3:1 or, ii) 3' tailing by terminal deoxytransferase (Pharmacia), using α^{32}P-dATP in molar ratio of 5-10:1. Specific activities of 0.5-$3x10^8$ cpm/μg and 3-$10x10^8$ cpm/μg respectively were routinely obtained with the two different methods, depending on the probe, specific activity of the ^{32}P

Table 5. *Design of oligonucleotides for probing ACTH-like sequences in unicellulars*

A) SHORT: BASED ON UNIVERSAL GENETIC CODE.

	Ser	Tyr	Ser	Met	Glu	His	Phe	Arg	Trp	Gly	Lys	Pro	Val /Ile	Gly	Lys /Arg	Lys	Arg	Arg	Pro	Val /Ile	Lys	Val /Ile	Tyr	Pro	Lys /Asn
5' coding	TCN AGY	TAY	TCN AGY	ATG	GAR	CAY	TTY	CGN AGR	TGG	GGN	AAR	CCN	GTN	GGN	AAR	AAR	CGN AGR	CGN AGR	CCN	GTN	AAR	GTN	TAY	CCN	AAR AAY
	6	2	6	1	2		2	6	1	4	2	4	4	4	2	2	6	6	4	4	2	4	2	4	4

5' ATG GAR CAY TTY $\frac{\text{CGN}}{\text{AGR}}$ TGG GG 3' 20mer/64 fold degenerate. Tm 58C.

B) LONG: BASED ON CODON PREFERENCE.

5' coding	TCY	TAC	TCY	ATG	GAA	CAY	TTC	AGA	TGG	GGT	AAG	CCY	GTY	GGT	AAG	AAG	AGA	AGA	CCY	GTY	AAR	GTY	TAC	CCY	AA
Tetrahymen.	0.9	0.8	0.9	1	1	0.9	1	1	1	1	0.9	1	1	1	0.9	0.9	1	1	1	1	0.9	1	0.8	1	1
Yeast.	0.6	0.6	0.6	1	0.8		0.5	0.6	1	0.7	0.6	0.4	0.8	0.7	0.6	0.6	0.6	0.6	0.4	0.8	0.6	0.8	0.6	0.4	1

5' CY TAC TCY ATG GAA CAY TTC AGA TGG GGT AAR CCY GTY GGT AAG AAG AGA AGA CC #22 57mer/64 fold degerate. 40%GC/ Tm 50C.

#23 56mer/64 fold 43%GC/ Tm c49C. 5' TTC AGA TGG GGT AAG CCY GTY GGT AAG AAG AGA AGA CCY GTY AAG GTY TAC CCY AA 3'

C) LONG: CODON PREFERENCE (WEAK) , PROTEIN "DEGENERACY", INOSINE, NORTHERNS

non-coding 3' ATG AGI TAC CTT GTG AAG $\frac{\text{ICT}}{\text{G}}$ ACC CCI TTT GGI TAI CCI TTI TTT $\frac{\text{ICT}}{\text{G}}$ $\frac{\text{ICT}}{\text{G}}$ GGI TAI TTT TAI ATG G 5'

Table 6. *Design of oligonucleotides for probing β-endorphin-like sequences in unicellulars*

A) SHORT: BASED ON UNIVERSAL GENETIC CODE.

	Lys	Arg	Tyr	Gly	Gly	Phe	Met	Thr /Lys	Ser /Pro	Glu /Tyr	Lys	Ser	Gln	Thr	Pro
3' NON-coding	TTY	GCN TCY	ATR	CCN	CCN	AAR	TAC	TGN T	AGN TCR	CTY	TTY	AGN TCR	GTY	TGN	GGN
	2	6	2	4	4	2	1	4	6	2	2	6	2	4	4
		3'	**ATR**	**CCN**	**CCN**	**AAR**	**TAC**	**T**	**5'**	16mer/64fold degenerate. Tm 42C					

B) LONG: BASED ON CODON PREFERENCE

3'NON-coding	TTC	TCT	ATG	CCA	CCA	AAG	TAC	TGR	AGR	CTT	TTC	AGR	GTT	TGR	GGR
Tetrahymena	0.85	0.98	0.81	0.96	0.96	1	1	0.98	0.86	0.96	0.85	0.86	1	0.98	0.81
Yeast	0.56	0.6	0.56	0.7	0.7	0.51	1	0.7	0.58	0.78	0.56	0.58	0.78	0.7	0.36
3'	**TTC**	**TCT**	**ATG**	**CCA**	**CCA**	**AAG**	**TAC**	**TGR**	**AGR**	**CTT**	**TTC**	**AGR**	**GTT**	**TGR A**	**GG**

5' #24 44mer/64fold. 39%GC Tm 45C

and activity of the enzyme. Probes differed in the rate at which terminal transferase added extra nucleotides and in specific activity obtained. In general a maximum of about 30 nucleotides were added after 60 min, although in some cases this was largely achieved by 15 min. Labelled probe was separated from unincorporated nucleotides by a short Sephadex G50 column purification (Stratagene Push Columns, San Diego, CA).

3.3.2 Polymerase chain reaction (PCR) labelling of rat POMC probe

Probes were labelled by incorporation of α^{32}P-dATP into products of the PCR reaction during amplification (148). One nanogram of linear plasmid DNA (10 ng for asymmetric PCR; 149) containing the region of interest was amplified over 30 cycles, with denaturation at 95°C, annealing at 42°C, and extension at 72°C. An initial denaturation step at 96°C for 3 min was included prior to addition of enzyme (2.5 U Amplitaq; Perkin Elmer-Cetus, Emeryville, CA). Primer concentrations (1 μM) and buffer composition (50 mM KCl; 10 mM TrisHCl, pH 8.3; 2 mM $MgCl_2$) were as recommended by the manufacturer. Comparison of different concentrations of "cold" dATP in the reaction showed that approximately equivalent amounts of label were reliably incorporated into product, regardless of the amount of cold above 5 μM (f.c.) dATP, below which the polymerase enzyme would probably cease to function efficiently (127, 148) (Figure 12). However, 10 μM (f.c.) cold dATP was chosen, since specific activity would be highest; other nucleotides were at concentration of 50 μM. Amplified products were purified by Sephadex G50 separation (Push column; Stratagene). Specific activity of over $1x10^9$ cpm/μg was routinely obtained. Between 1-1.5 million cpm/ml was added in each experiment. Double stranded probes were denatured by boiling for 10 min and rapidly cooled on ice prior to use.

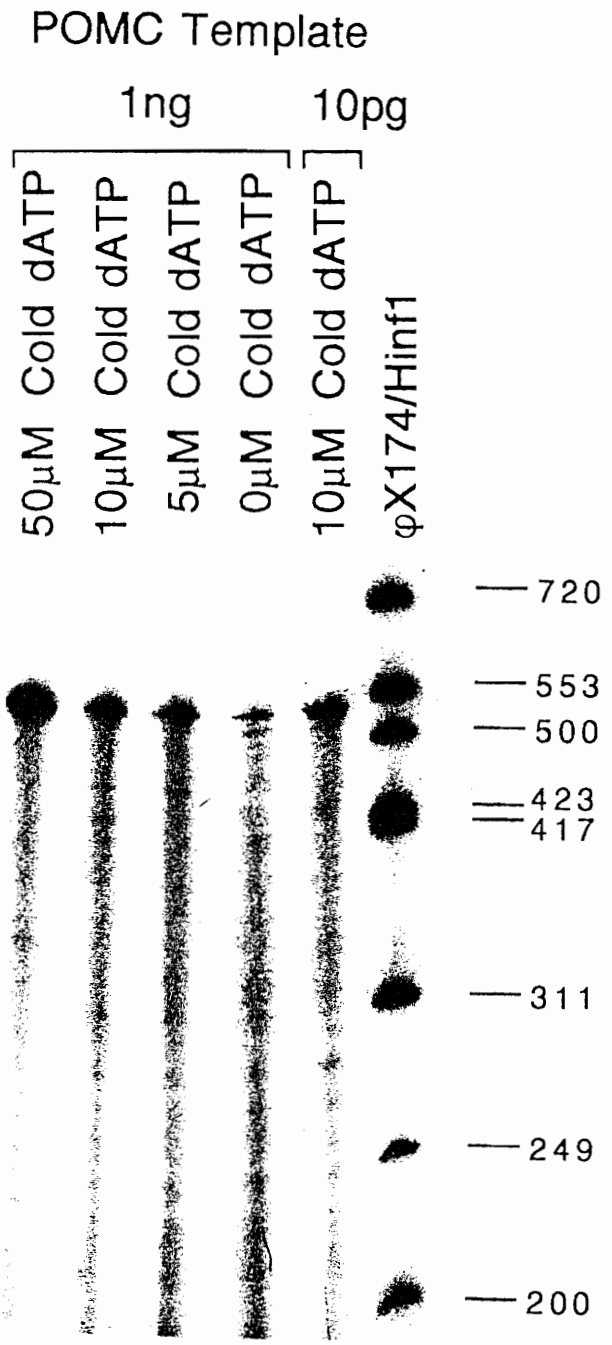


Fig. 12. Labelling of rat POMC gene (Exon 3) by PCR. Varying amounts of cold dATP in the reaction (from 50-5μM f.c.) are as shown, together with two different amounts of starting template (1ng, 1pg). PCR reaction conditions are given in the text.

3.4 Hybridization Strategy

This was carried out in several stages, according to Sambrook (125).

3.4.1 Prehybridization/ hybridization

Membranes were first prehybridized for 6-24 h (usually overnight) at a temperature similar to that used for subsequent probe hybridization (varying between 37°C and 65°C). Duration of subsequent hybridization varied (18-48 h) depending on the nature of the probe added; short single stranded oligonucleotide probes hybridizing at low stringency required significantly less time to reach equilibrium than long denatured double stranded probes at higher stringency (125). Prehybridization and hybridization buffers were identical, apart from the absence of labelled probe. Several components of the medium were varied to determine empirically the optimal composition. Denhardt's reagent was chosen as superior to nonfat dried milk ("Blotto"; 125) because it was more effective at reducing background at low hybridization stringency. Formamide concentrations were varied between 0-50% (150, 151). However, it was impossible to find a formamide concentration that gave hybridization conditions suitable for both yeast and rat DNA simultaneously. The prehybridization/ hybridization buffers used for most experiments are as described in Sambrook (129), and are shown in Table 7.

Optimal visualization of yeast bands was generally seen using the low stringency conditions; for both oligonucleotide and PCR-labelled probes, hybridization conditions generally included 20% formamide, 0.5% SDS, and 100 µg/ml *E. coli* genomic DNA as carrier. *E. coli* genomic DNA (Sigma) was generally used since yeast tRNA was considered of inadequate complexity to adequately block genomic DNA and fish (salmon sperm) DNA was avoided because it was considered unwise to use a blocking DNA of closer homology

Table 7. *Components of hybridization buffer for Southern blotting*

1.	6X SSC (1X SSC = 150 mM sodium chloride, 15 mM sodium citrate).
2.	5X Denhardt's solution (1% Ficoll 400; 1% Polyvinylpyrollidine; 1% bovine serum albumin - Fraction V; all from Sigma).
3.	50 mM (f.c.) sodium phosphate, pH 7.4.
4.	0.5% (f.c.) Sodium dodecyl sulfate (SDS).
5.	0-50% formamide (Redistilled; EM Science); unless otherwise stipulated, a final concentration of 20% was routinely used.
6.	100 µg/ml sonicated denatured *E. coli* carrier DNA (Sigma).

to the probe than the sequences being screened. The *E. coli* carrier DNA was denatured by boiling for 10 min, before addition to the buffer. However, for rat genomic DNA, single bands were generally best visualized using higher stringency - 50% formamide, 1% SDS, and 100 µg/ml Salmon sperm DNA (Sigma), hybridizing at a temperature of 42°C, and with a final wash in 0.2X SSC/0.1% SDS, at 58°C or higher.

3.4.2 Washing conditions

All membranes were washed initially at low stringency in high salt (2X SSC, 0.1% SDS) at room temperature (2 x 15 min). Subsequent washes depended on the stringency of hybridization sought, with final washing generally being either 37°, 42°, 50° or 65°C, in 0.2X SSC, 0.1% SDS. Final washing conditions varied depending on the particular probe used, and were determined by estimated annealing temperature (125), or empirically by detecting the amount of background using a Geiger counter and/ or rewashing if an initial short (e.g., overnight) exposure showed an unacceptably high level of background activity (143).

3.5 Results of Southern Blotting

3.5.1 Southern blots using oligonucleotides

Despite repeated Southern blots, no bands were seen in Tetrahymena genomic DNA digests under conditions where both rat and yeast DNA showed positive signals. Labelling of an Eco R1 digest of Tetrahymena genomic DNA by the random primer technique prior to electrophoresis and transfer showed efficient transfer of the DNA onto the hybridization membrane (data not shown).

Figure 13 shows results of parallel Southern blots for rat and yeast DNA, using degenerate oligonucleotides coding for conserved portions of the α-MSH and β-endorphin regions of vertebrate POMC, and incorporating codon bias into their design, based on the distinct preference of yeast and Tetrahymena for certain codons (144, 145). The use of terminal transferase labelling of oligonucleotides resulted in significantly higher background, but major band patterns were unchanged.

3.5.2 Southern blots using PCR-labelled rat gene

Figure 14 shows Southern blots of rat and yeast DNA identical in all respects other than the stringency of hybridization and washing and probed with the PCR-labelled rat gene in the 530 bp region between γMSH and β-endorphin. The left blot (a) was hybridized and washed at high stringency, whereas the right (b) was at much lower stringency. At high stringency the rat DNA showed a 13 kb band for the Eco R1 digest, a 3. 7 kb band for Hind III, a 6.7 kb band for Xba 1, a 10 kb band for Sac I, and a 0.7 kb band for Hinf I. These rat bands corresponded in size to those previously described (33); although Drouin and Goodman (33) initially described two bands in rat teratocarcinoma cells, these authors later describe the presence of only one gene in the rat (34), with restriction fragment sizes similar to those given above.

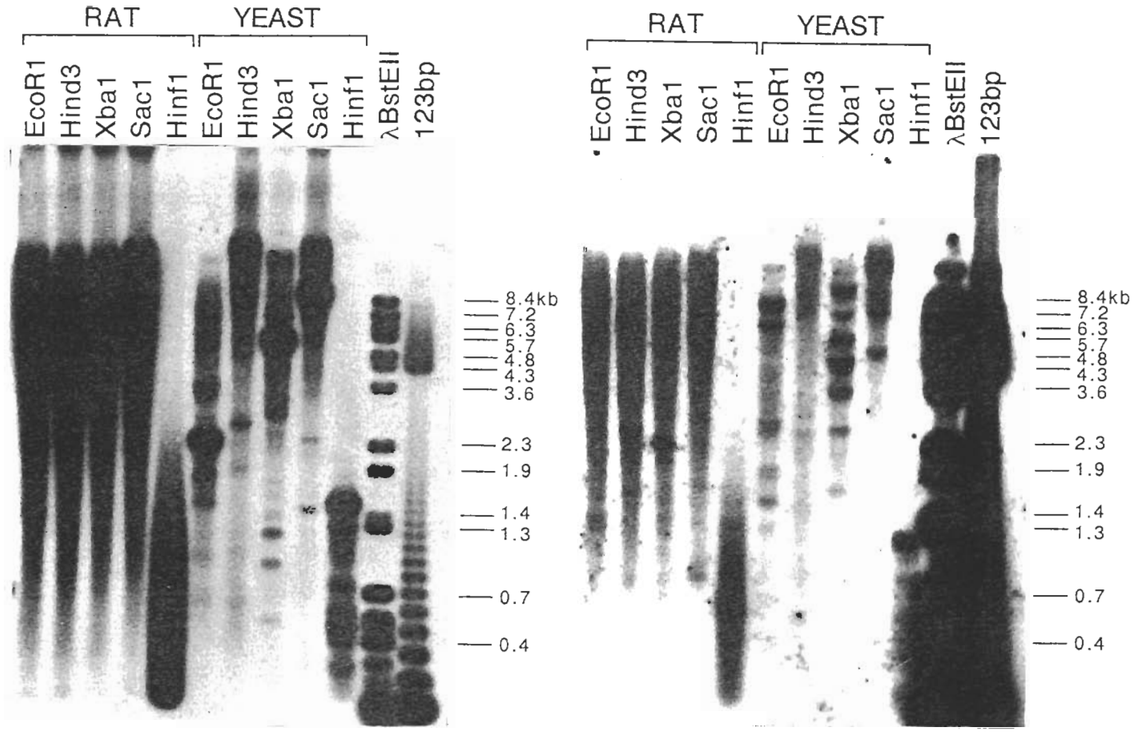


Fig. 13. Southern blots using oligonucleotide probes. Rat and yeast genomic DNA was digested with enzymes as shown and electrophoresed in 0.7% agarose in 1X TAE buffer, together with labelled DNA size markers, then transferred to nylon. The left was probed with oligonucleotide #23 (ACTH) and the right with #24 (β-endorphin) (Tables 5 & 6), at 37C in hybridization buffer (Table 7). Probes were labelled using terminal transferase (#23) or T4 polynucleotide kinase, as described in text. Blots were washed at final stringency of 1X SSC, 0.1% SDS at 37C, and were exposed to X-ray film for 2.5 days and 6 days.

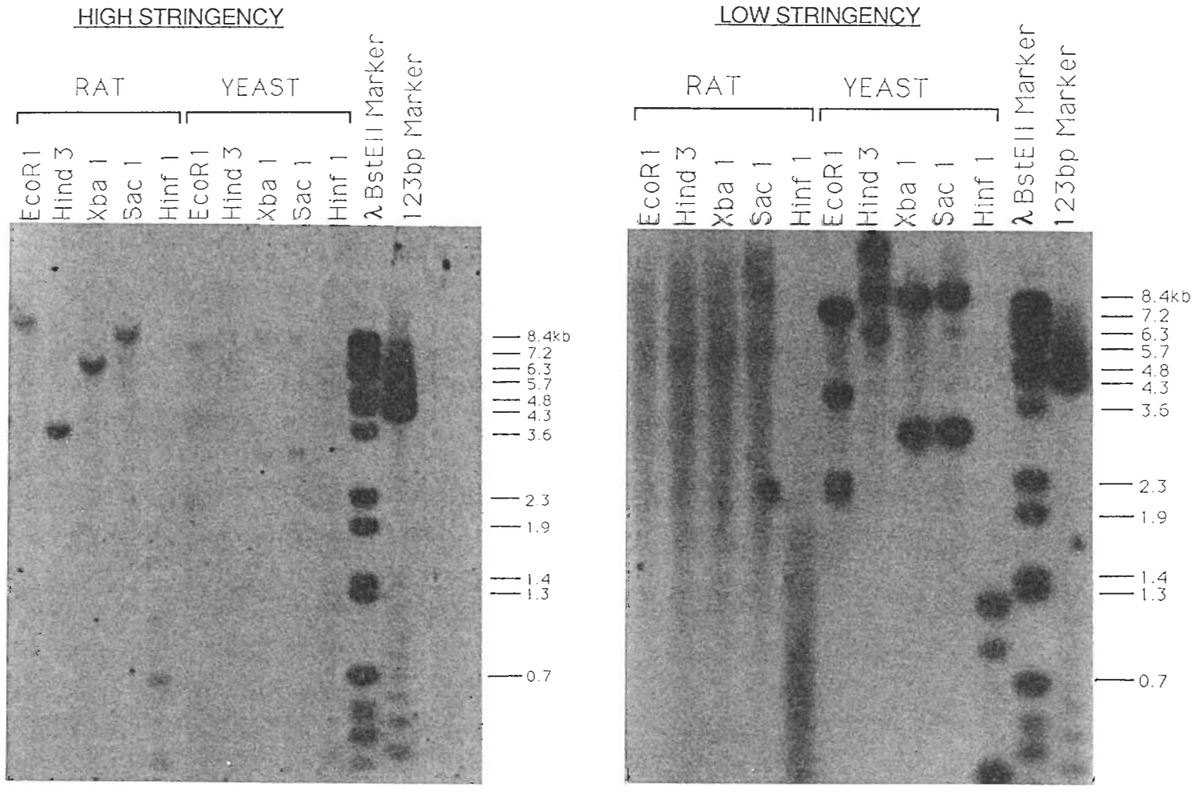


Fig. 14. Southern blots using rat POMC gene probe. Rat and yeast genomic DNA was digested with enzymes as shown and electrophoresed in 0.7% agarose in 1X TAE buffer, together with DNA size markers. Both blots were probed with PCR-labelled rat POMC gene (530 base fragment from conserved third exon) at high (left), and low (right) stringencies. Final washes were 0.2X SSC, 0.1% SDS at 56° and 42°C respectively.

At high stringency two very faint bands could be seen in several yeast DNA lanes. When hybridization was performed at the lower stringency seen at right (Figure 14B), nonspecific binding to the rat DNA obscured the single copy specifically hybridizing sequences. In the yeast lanes, however, two bands were readily identified. These restriction fragment lengths are shown for both rat and yeast in Table 8.

Although the hybridizing restriction fragments in yeast DNA were of different sizes to rat, careful attention was paid to the possibility that these yeast bands were not due to contamination of the yeast genomic DNA preparation with the cloned rat POMC gene. To answer this, separate experiments involving digestion and analysis of the cloned rat gene was performed, generating a different set of restriction fragments, and showing this was indeed not the cause of these bands.

The strength of the hybridizing bands in yeast was initially very impressive, but it is important to note that the yeast genome is only 1/300th the size of the rat, and therefore 10 µg of yeast genomic DNA contains three hundred times as many copies of each potential hybridizing sequence. Thus, when 30 ng

Table 8. *Restriction fragment lengths (kb) of rat and yeast DNA (Rat POMC Exon 3 PCR probe)*

Enzyme	Species.	
	Rat.	Yeast.
Eco R1	13kb	8.0, 4.3, (2.4, 2.2)
Hind III	3.7	9.0, 6.7
Xba 1	6.7	8.4, 3.0
Sac 1	10	8.5, 3.0
Hinf 1	0.3, 0.7	1.25, 0.9, 0.35

of yeast genomic DNA (containing an equal number of copies of each gene to 10 μg of rat DNA) were digested, electrophoresed, transferred and hybridized under identical conditions, no bands were seen (data not shown), indicating the level of homology was not as great as one might first think.

CHAPTER 4

SCREENING OF YEAST GENOMIC LIBRARIES

4.1 Library Screening Methods

4.1.1 Plating phage libraries on soft agar

Both a yeast genomic DNA library cloned in the expression phage lambda gt11 and strain Y1090 *E. coli* cells (used for λgt11 transfections) were kindly provided by Dr D. Stillman, Cellular, Viral and Molecular Biology Dept., University of Utah. Y1090 cells were grown at 37°C in Luria Broth (10 g Tryptone, 5 g Yeast Extract; 10 g NaCl/ Litre) with 2% maltose and 10 mM MgSO4 (f.c.). Cells were harvested in log phase (absorbance at 600 nm of about 0.5) by centrifugation at 2000 *g* for 5 min. The pellet was resuspended in 10 mM MgSO4 (to O.D.600 of 0.5) and stored on ice until use (within 24 h).

Bacteria were transfected with phage by coincubation at 37°C for 20 min. Approximately 250,000 plaques were screened at 8,000 plaques per 82 mm diameter plate, equivalent to five yeast genomes and sufficient to give greater than a 99% likelihood of including a single copy gene of interest in the primary screen. Two hundred microliters of transfected bacterial suspension were mixed with 3ml of liquid LB/Ampicillin/0.7% agarose at 48°C. After quickly vortexing, the mixture was immediately plated onto LB/Agar plates and incubated at 37°C overnight (12-16 h) until plaques were approximately 1 mm in diameter, but not confluent.

For secondary screening, a small (about 3 mm diameter) area of the primary plate was removed by Pasteur pipette and transferred into 1ml of SM buffer in an Eppendorf microfuge tube (SM buffer: 50 mM NaCl, 10 mM TrisHCl, 10 mM $MgSO_4$, 0.1% gelatin). Chloroform (20 µl) was added to completely lyse all bacteria and maximize phage titer. Phage were eluted by agitation overnight at 4°C. One microliter of eluted phage (final dilution around 1:300,000) generally resulted in approximately 100-250 plaques per plate, enabling picking of individual positive plaques for further characterization.

4.1.2 Plaque transfer and screening

Phage DNA was transferred directly onto nitrocellulose filters by laying the dry, labelled filters (Schleicher and Schuell) onto the bacterial lawn on the plates. Two filter lifts were performed for each plate, the first filter being applied to the plate for 1 min and the second for 3.5 min. Location of the filter on the plates was carefully and accurately marked with a fine 23-gauge needle stabbed through the filter and into the agar in several asymmetric positions. DNA was denatured and immobilized on the filters according to Sambrook (125). Filters were placed (DNA side up) on Whatman 3 mm paper saturated with 0.5 N NaOH for 7 min, neutralized by two washing steps in 1 M TrisHCl/1.5 M NaCl, pH 7.4, then air dried and baked for 2 h at 80°C.

Filters were screened for positive plaques by hybridization as described earlier. After prewashing in 6X SSC, they were prehybridized and hybridized in 6X SSC, 0.5%SDS, 50 mM sodium phosphate, pH 7.4, 20% formamide and 100 µg/ml of sonicated denatured *E. coli* genomic DNA. Total hybridization volume was approximately 2 ml per filter, with the hybridization being done in a small round Tupperware container. Other parameters of the hybridization, including amount and type of probe, are as described previously (section 3.4).

4.1.3 Purification of insert DNA

Single positive plaques were amplified initially by growth to confluence on soft agar overnight with phage then eluted by gentle agitation into SM buffer (100 mM NaCl, 50mM TrisHCl, pH 7.5, 2% $MgSO_4$-$7H_2O$, 0.01% gelatin; 5 ml added per plate). This phage suspension was then used to infect a 500 ml culture of Y1090 *E. coli* cells in log phase (125), followed by another 3-5 h of culture at 37°C until stringy debris indicated lysis of cells. Chloroform (10 ml) was added and the incubation continued for another 10 min to lyse any remaining cells.

DNase and RNase A (1 µg/ml final; Promega Biotec) were added to the culture lysate for 30 min at 25°C; after addition of NaCl to 1 M final concentration, the culture was allowed to sit on ice for 1 h, then centrifuged at 11,000 *g* for 10 min at 4°C. The supernatant was mixed with 40% polyethylene glycol 8000 (PEG; Sigma) to bring final PEG concentration to 8%, then allowed to sit once more at 4°C for 1 h, and again centrifuged at 11,000 *g* for 10 min. The precipitate was dried in air and then resuspended in SM buffer, mixed with an exactly equal volume of CsCl-saturated SM buffer and centrifuged overnight at 45,000 rpm in a Ti70 fixed angle rotor at 20°C .

Phage were removed by syringe and dialyzed in SM buffer to remove CsCl. The protein capsid was removed by digestion for 30 min at 37°C with 100 µg/ml Proteinase K (Boehringer-Mannheim), in the presence of 20 mM EDTA and 0.5% SDS. The digest was then extracted once with redistilled phenol, once with a phenol:chloroform (50:50 v/v) mix and once with chloroform alone. Isolated high molecular weight phage DNA in the aqueous phase was then precipitated using two volumes of absolute ethanol. The DNA was spooled off using a glass rod, washed twice in 70% ethanol, and redissolved slowly in TE (10 mM TrisHCl, 1 mM EDTA, pH 7.4).

4.2 Results of Library Screening

4.2.1 Screening using oligonucleotide probes

The library was initially probed with biased oligonucleotides #23 and 24, which were 3' end-labelled using terminal transferase. When hybridization was performed at low stringency (20% formamide; 37°C), with final wash at 0.2X SSC/0.1% SDS at 42°C, seven positive clones were chosen for large scale amplification, after several rounds of secondary and tertiary screening. Purified phage DNA from each clone was digested with Eco R1 to release the insert; the digests were then electrophoresed and subsequently probed with terminal transferase-labelled oligonucleotides #23 and #24.

Hybridizing bands were identified varying between 6.7 and 0.5 kb in size (data not shown), with several containing inserts of similar sizes which subsequently proved to be identical in sequence. These positive bands were subcloned into the PUC-derived plasmid vector pTZ18, and positive inserts identified by color selection and subsequent hybridization. The size of these inserts meant direct sequencing of the inserts was impractical. The insert DNA in each case was therefore purified away from the vector and digested with several different restriction enzymes with four base recognition sequences.

Electrophoresis of these digests and subsequent Southern analysis again using oligonucleotides #23 and #24 (directed at ACTH and β-endorphin regions) further narrowed the regions of DNA homology to easily sequencible sizes of several hundred base pairs. These fragments were "blunt-ended" with the Klenow fragment of DNA polymerase and ligated into the Bluescript sequencing plasmid, pKS-. The positive clones containing inserts were sequenced from both ends, using the dideoxy sequencing method (125).

Several of the sequences obtained are shown in Figure 15. Probe sequence in two of these is compared to the sequence identified from yeast DNA

#22/23 (Clone JJ1)

Probe
1 CYT---ACTCYATGGAACAYTTC-AGATGGGGTAAGCCYGTYGGT-AAG
161 CTTATAAATCAAAAGAATAGACCGAGATAGGGT-TGAGTGTTGTTCCAG
Clone

#24(Clone JJ1)

Probe
1
GGRGAT---TRRGACTTTTCT--GATGTCAT-------GAAACCACCGTATCT
GCGTTTCCGTTAAACTATCGGTTGCGGCCATATCTACCAGAAAGCACCGTTTCC
22
Clone

Sequence within Clone J3.

```
5' GTG GGT GAC CAT ACG CGA AAC TCA GGT GCT GCA ATG
   TTT ATT TCT TTT TTT TTT TTT TTT TTT TTT TTT TTT
   TCT AGT TTC TTG GCT TCC TAT GCT AAA TCC CAT AAC  3'
```

Fig. 15. Yeast sequences hybridizing to oligonucleotide probes. Regions of closest nucleotide homology to probes are as shown, with probe being the uppermost of each pair and yeast DNA sequence beneath. A sequence showing the presence of long runs of dA-dT nucleotide pairs is also shown.

and showing highest nucleotide homology. Additionally, several sequences showed a long run of poly-dT residues, which appeared to be responsible for the hybridization to the dATP-terminal transferase-labelled probes. None of the sequences contained an open reading frame.

The fact that no open reading frame was found in any clone, together with the apparent spurious hybridization to terminal transferase prompted me to

reprobe the Southern blot of phage DNA from the positive clones with the same oligonucleotides, but this time 5' end-labelled by T4 polynucleotide kinase to avoid the confounding feature of the labelling process itself; no hybridizing bands were seen at a wash stringency of 0.2X SSC/0.1% SDS at 37°C, the stringency at which bands were present in Southern blots. Overall, these results indicated not only that a number of the "positive" clones were an artifact of the labelling procedure, but that the sequences identified by positive bands on the Southern blots had not been found. This result was surprising in that Southern analysis of DY150 genomic DNA had shown very similar band patterns regardless of the method used to label the probe (5' polynucleotide kinase or 3' terminal transferase tailing), even though the background was significantly higher using the terminal transferase method of labelling.

When the library was reprobed twice with the same oligonucleotides labelled with T4 polynucleotide kinase and washed at 0.2X SSC at 37°C, a number of potentially positive clones were identified on initial screening. However, signal to background ratio was low and none of the clones survived subsequent secondary and tertiary screening. As a third approach, I constructed a shelf library using the Lambda Zap II system (Stratagene); Eco R1-digested yeast DNA was electrophoresed on 0.7% NuSeive agarose (FMC) and the region corresponding to the size of the positive band on Southern digests (2.4 kb) was ligated into the Eco R1-digested phage vector. Although the ligation was efficient and very few nonrecombitants were obtained, one round of screening with each of several probes at low stringency showed no positive clones. The number of recombitant plaques should have been adequate to identify at least five positive clones based on the size distribution of fragments and assuming all DNA sequences were inserted with equal frequency. Time was unavailable to pursue this avenue further.

4.2.2 Screening using rat POMC PCR probe

Screening using a PCR-labelled 530 base pair fragment of the rat POMC gene was more successful. It was also performed at the low stringency described above (20% formamide; 37°C), with final wash in 0.2X SSC/0.1% SDS at 50°C. Approximately five genomes were screened in each hybridization experiment, giving a statistical likelihood of over 99% of identifying at least one clone containing a single copy gene of interest. However, because on average one would have expected five positive clones for each separate single copy hybridizing species whereas the number of positive clones identified on each primary screening of this library was never more than two (one, one and two respectively on three separate occassions), this suggested that the library was significantly underrepresented in some sequences.

All positive phage clones were isolated by subsequent secondary and tertiary hybridization screenings, and large amounts of phage DNA were prepared. Subsequent restriction digestion of this phage DNA showed all the positive clones to be identical. The fact that only one hybridizing species was consistently identified, rather than two as would be expected from the Southern blot, together with the relatively low frequency of this sequence in primary screens, suggested that the library may well have been significantly under-represented in one of the sequences hybridizing in the Southern blot, and deficient in the other, possibly through repeated amplifications.

4.2.3 Mapping of hybridizing fragments

4.2.3.1 Subcloning of yeast insert DNA

DNA prepared from the positive clones was digested with Eco R1 to cleave out the yeast DNA insert. Size fractionation of the digest showed two small DNA fragments in addition to the very large lambda DNA vector arms;

these were 0.85 kb and 0.4 kb in size. However, Southern blotting of this restriction digest and subsequent probing with the same PCR-generated probe showed that neither small fragment hybridized; rather the high molecular weight DNA, including the phage (vector) itself, showed positive hybridization (data not shown). These results indicated i) that the hybridizing region of yeast DNA had not been spliced out of the vector, presumably because at least one arm of the cloning site had been lost during library construction, and ii) assuming no star activity of the enzyme had occurred (i.e., there were no additional potential restriction sites in the vector) the two small fragments were from another non-hybridizing, possibly adjacent region of yeast DNA, rather than the hybridizing region.

In order to identify an insert fragment containing the region of homology and small enough to be subcloned, the phage DNA was digested with additional restriction endonucleases, both singly and in pairs (double restriction digests). In addition to Eco R1, the enzymes Kpn 1, Pvu 1 and Sac 1 were chosen because of their relatively low overall frequency of cutting in the vector, but with the presence of at least one restriction site close to the Eco R1 cloning site, thereby minimizing the length of unwanted vector sequence in any hybridizing fragment. Subsequent transfer and probing identified smaller hybridizing bands separate from vector sequences. The most potentially useful were a 3 kb Sac 1 fragment and an 11 kb Pvu 1 fragment (data not shown). Although the 3 kb Sac 1 fragment was smaller, it nevertheless appeared to contain all the hybridizing region of DNA, because no other hybridizing bands were also present in phage DNA digested with this enzyme and was likely to be much easier to subclone. This fragment was subcloned into the multicloning region of the pUC-derived plasmid pTZ18U, and amplified in *E. coli* (strain 71-18) using the maxiprep procedure and cesium chloride ultracentrifugation (125).

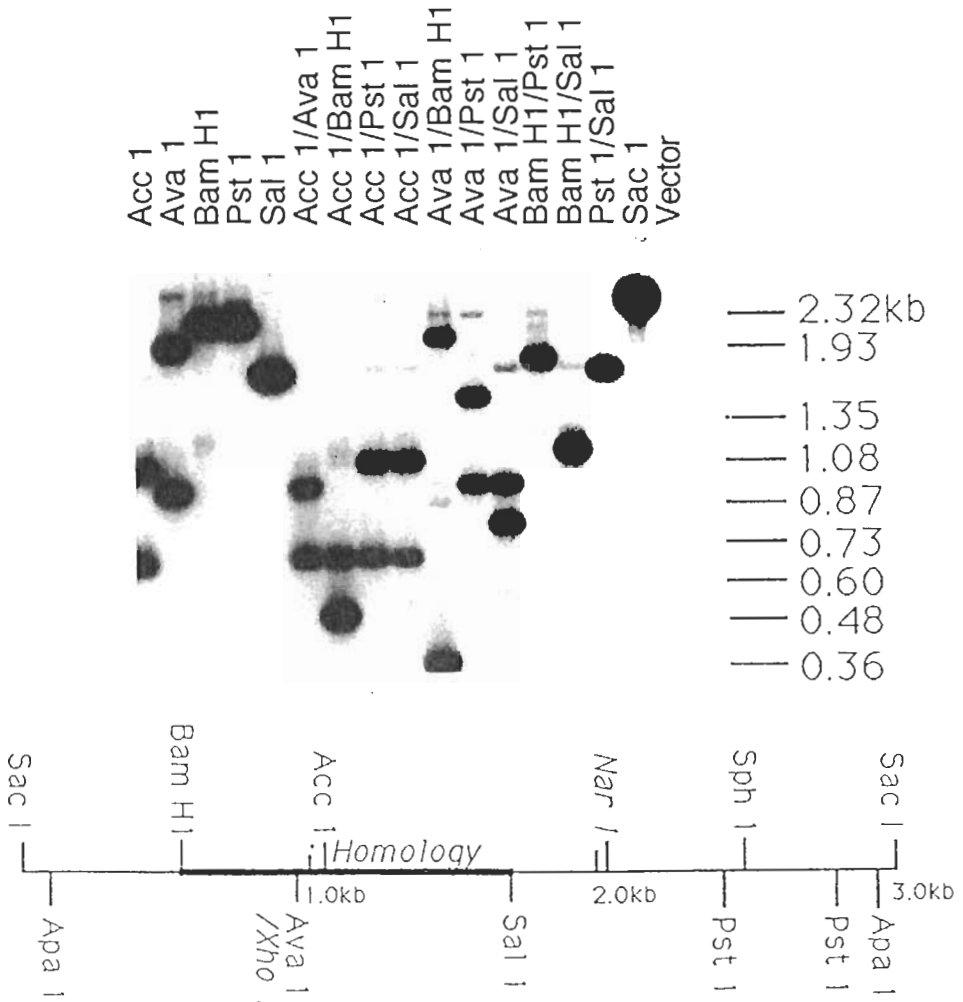
Further analysis of the insert DNA was initiated by identifying a smaller region of DNA containing the homology to the rat gene within the 3 kb insert. This was achieved by generating a restriction map using double enzyme digests of insert DNA. The plasmid (pTZ19h) was initially digested with Sac 1 to isolate the insert, which was then digested overnight with the following enzymes, both alone and in pairs: Acc 1, Ava 1, Bam H1, Sal 1, and Pst 1. These enzymes were chosen in part because the vector lacked restriction sites for them other than within the multicloning region, making the restriction pattern easier to interpret. The enzyme digests were then electrophoresed, transferred to nylon, and probed using the rat PCR-labelled probe, as was used previously. The hybridization pattern, together with the restriction map generated and region of potential homology are shown in Figure 16. This experiment identified the hybridizing region as lying largely if not entirely in a 1.1 kb region between internal Bam H1 and Sal 1 sites.

4.2.4 Back-hybridization using yeast sequences

The 1.1 kb Bam H1/Sal 1 fragment identified above was purified from the adjacent vector and insert sequences. It was labelled by the random primer method (125) using a commercial kit (Boehringer-Mannheim) under the conditions specified by the manufacturer. Labelled probe was purified from unincorporated α^{32}P-dATP by passage through a short Sephadex G50 column (Stratagene "Push Column"). A specific activity of at least $3x10^8$ dpm/μg was obtained, with less than 1% contamination by unlabelled nucleotide.

Southern blots containing digests of both yeast and rat genomic DNA were probed with this probe under conditions of low (37°C; 20% formamide) and high (42°C; 50% formamide) stringency. The stringency of washing was also increased progressively and the blots exposed at each step to identify

A) RESTRICTION MAP.



B) SEQUENCING PROTOCOL.

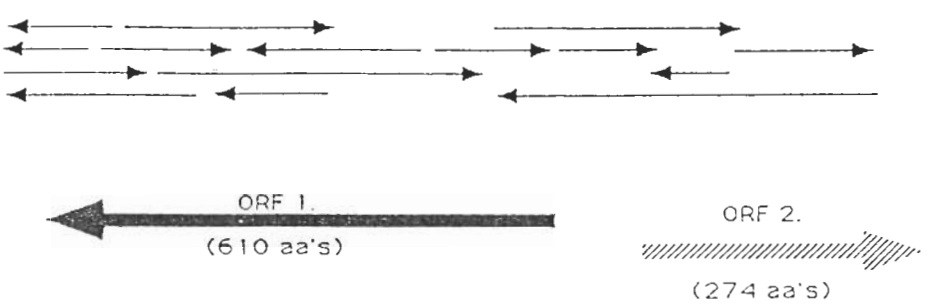


Fig. 16. Restriction map and sequencing protocol of 3kb Sac 1 fragment. A) Southern blot of single and double digests of 3kb Sac 1 fragment with resulting restriction map and region of DNA homology to rat probe being as shown. B) DNA sequencing protocol for 3kb Sac 1 fragment, with open reading frames as shown.

regions of homology within yeast and rat DNA. The results of this experiment are shown largely in Figure 17. Identical blots are compared; that probed with the PCR-labelled rat gene is seen on the left and that with the 1.1 kb yeast DNA fragment on the right. The data indicated that at least one of the sequences that had been cloned that was responsible for the bands hybridizing to the PCR-labelled rat gene on previous Southern blots. However, the data also indicated that the other hybridizing sequence(s) had not yet been identified.

As previously indicated, several attempts to identify other hybridizing clones using the existing λgt11 yeast library were unsuccessful, suggesting the other sequences were no longer adequately represented. I therefore tried to identify and clone these sequences by constructing a shelf library of yeast DNA in the size range of 3.8 to 4.6 kb after Eco R1 digestion (corresponding to the second Eco R1 band), and in the size range of 3.0-3.6 kb after Xba 1 digestion. DNA of this size range was purified from a 0.7% NuSeive agarose gel and ligated into the commercial phage vector Lambda Zap II (Stratagene). A high titer of recombitants was obtained (122) with less than 5% nonrecombitants, and approximately 30,000 plaques were plated and screened by hybridization to the PCR-labelled rat gene. However, no positives were obtained after a single round of screening, and time was unavailable for further investigation.

Several other interesting conclusions were possible from the pattern of bands in Figures 14 and 17. First, the size of the Eco R1 band hybridizing to the 1.1 kb yeast sequence (approximately 12 kb) was too large for the λgt11 vector, which has a maximum insert size of approximately 7-8 kb. This might explain why all the positive clones isolated lacked one of the Eco R1 sites, since an *intact* 12 kb restriction fragment would not be able to be packaged in this vector and therefore almost certainly not be represented in the library. Second, the strength of the hybridization signal to the 1.1 kb yeast fragment was *less* than the

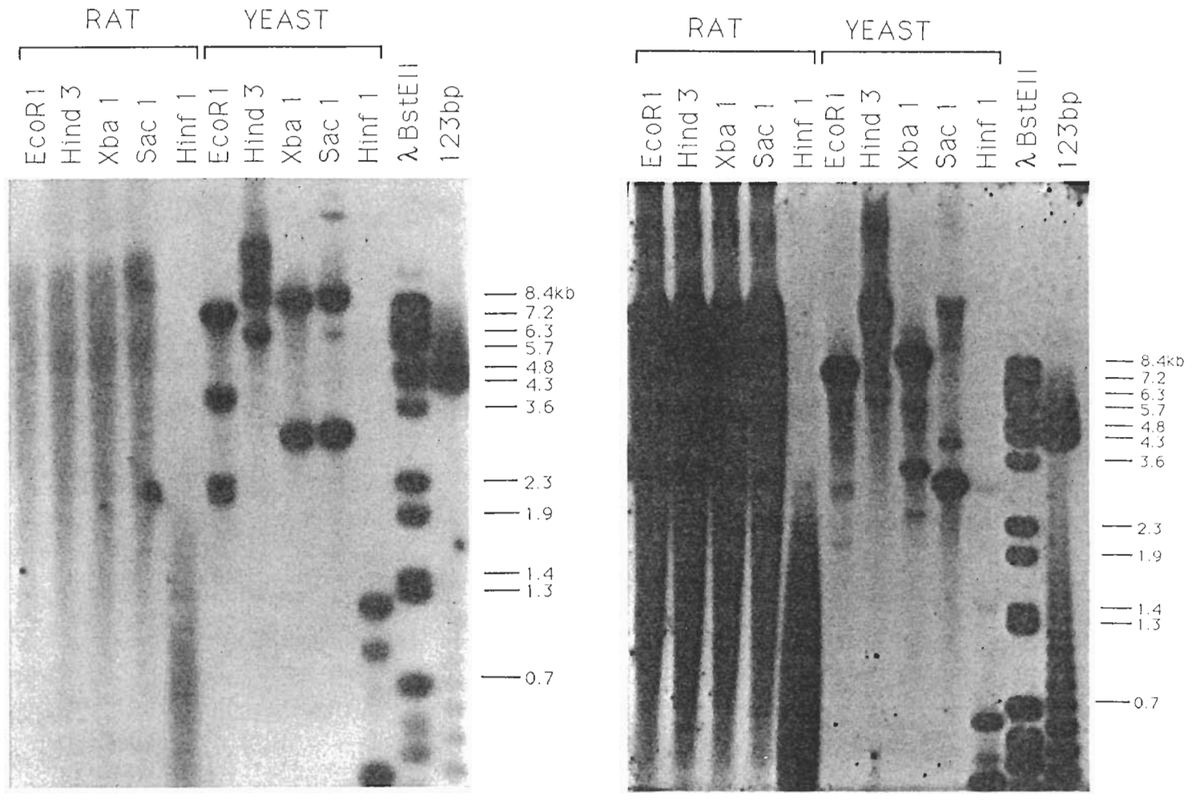


Fig. 17. Southern blots using rat POMC gene and yeast 1.1 kb Bam H1-Sal 1 fragment. Rat and yeast genomic DNA was digested with enzymes as shown and electrophoresed as described in text. The left was probed under low string-ency with PCR-labelled rat POMC gene as described in text. The right was probed with yeast 1.1 kb Bam H1-Sal 1 fragment (random primer-labelled), and washed in 0.2X SSC, 0.1% SDS at 42°C.

signal seen with the rat POMC gene and not more, as one might have expected for a probe derived from the same species. This would suggest that the homology of each of the different yeast sequences with the rat POMC gene was greater than their homology between themselves, and that the yeast sequences were unlikely to represent potential members of a family of related genes. Finally, although the "reverse" hybridization pattern using yeast sequence to probe rat DNA did not clearly show single bands corresponding to the POMC gene, it indicated that there were possibly other sequences related to the yeast gene, probably at a relatively low level of homology.

CHAPTER 5

ANALYSIS OF HYBRIDIZING YEAST FRAGMENTS

5.1 Methods

5.1.1 Overview of subcloning and sequencing protocols

The 3 kb of phage insert DNA was initially cloned into the multiple cloning site within the plasmid vector pTZ18 (Promega), a high copy vector with a selectable ampicillin resistance marker. Smaller fragments were later subcloned into the Bluescript II plasmid vector (KS- orientation; Stratagene), a vector in which the multicloning site is designed specifically to allow sequencing by Exonuclease III/ S1 nuclease digestion.

Initial sequencing of the 1.1 kb Bam H1/ Sal 1 fragment showed the presence of a long, previously uncharacterized, open reading frame extending through the entire 1.1 kb in the direction from Sal site to Bam site (Figure 16B). To characterize this potential gene further, this prompted me to sequence the rest of the 3 kb insert by a combination the following: a) Subcloning and sequencing individual restriction fragments within the 3 kb, b) "Primer walking," and c) Exonuclease III/ S1 nuclease sequencing of two long DNA fragments partly overlapping the 1.1 kb fragment. These other two fragments were i) a 1 kb Acc 1/ Eco R1 fragment extending from the Acc 1 site 500 bases internal to the Bam H1 site, out to an Eco R1 site just outside the Sac 1 cloning site (this overlapped with the initial 1.1 kb fragment by 500 bases and included the downstream C-terminal end of the open reading frame), and ii) a different Acc 1/Eco R1

fragment including the entire 1.3 kb external to the Sal site (this fragment contained the 5' end of the open reading frame and also included a second gene on the opposite strand). A diagram showing the different sequencing steps is shown in Figure 16B.

Further sequencing of yeast DNA was required beyond the Sac 1 site at the end containing the second (smaller) open reading frame. This was achieved by redigesting the intact phage DNA with another pair of enzymes (Eco R1/Sph 1) and identifying a hybridizing restriction fragment extending outside the 3 kb Sac 1 fragment. A 30 base synthetic oligonucleotide lying just inside of and directed toward the Sac 1 site at this end was 5'-labelled with T4 polynucleotide kinase and used as a probe under high stringency. The probe hybridized uniquely to a 2 kb Eco R1/Sph 1 fragment which was subcloned into Bluescript, and this same synthetic DNA oligomer was then also used as a sequencing primer to sequence up to and beyond the 3' end of the second open reading frame.

5.1.2 Exonuclease III digestion strategy

The 1.1 kb Bam H1/Sal 1 fragment was directionally cloned into the multicloning site in Bluescript. Bam H1 and Sac 1 cuts then enabled exonuclease III deletions from the 5' overhang of the Bam H1 site into the insert (and subsequent sequencing of one strand), while Kpn 1 and Sal 1 cuts enabled deletions from the 5' Sal 1 overhang and sequencing of the other strand. By this method, overlapping deletions of varying sizes from 1.1 kb down to less than 0.1 kb were generated in both DNA strands, enabling sequencing of the entire 1.1 kb from both directions. A similar strategy was employed for the two Eco R1/Acc 1 fragments (1.3 kb and 1 kb) on either side of this 1.1 kb fragment, enabling sequencing of almost all of the original 3 kb Sac 1 fragment. After cloning into identical sites on the vector, subsequent Bst XI and Eco R1 (5' overhang) cuts

enabled sequencing of one strand, while Kpn 1 and Acc 1 (5' overhang) cuts allowed sequencing of the other. Deletions of various sizes were again generated within the insert sequence by digestion with Exonuclease III (Boehringer-Mannheim). Plasmid DNA (10 µg) was digested at 34°C with 175 U of Exonuclease III, with aliquots of the digest being removed every 15 sec into ice cold S1 Nuclease buffer. This ratio of DNA to enzyme resulted in digestion of about 100 bases/min. The single stranded overhang was then digested with 30 units of S1 Nuclease at 30°C for 30 min. After ethanol precipitation and washing, the DNA was blunt-end ligated for 4 h at room temperature (T4 DNA Ligase), resulting in circular plasmids with varying size deletions in the insert. These were transformed into *E. coli* 71-18 strain and grown overnight on LB/Ampicillin plates.

Size of the insert DNA in individual colonies was estimated by PCR amplification of DNA from intact bacteria taken from individual colonies. Oligonucleotides binding to vector sequences on either side of the multiple cloning region ("T3" and "T7" primers) were used as primers. The PCR protocol consisted of 30 cycles, each of 1 min at 94°C (to initially lyse cells and denature DNA), 1.5 min at 42°C (for primer annealing) and finally 3 min at 72°C (for extension). Electrophoresis in a 1.5% agarose gel showed products corresponding to inserts of different length, which could then be ordered by size, and individual clones differing by about 150 bases in length could be sequenced.

5.1.3 Preparation and sequencing of DNA

Plasmid DNA was purified for double stranded sequencing by the alkaline lysis method (125). Overnight bacterial cultures (5 ml) were pelleted by brief centrifugation and cell pellets resuspended in 200 µl of Lysis buffer (10 mM TrisHCl, pH 8.0; 1 mM EDTA; 0.9% glucose) for 10 min on ice and cell lysis was

achieved by adding another 400 µl of 0.2N NaOH/ 1%SDS on ice for 15 min. Concentrated potassium acetate solution (360 µl) was added for another 15 min on ice to precipitate proteins and genomic DNA, tubes were centrifuged for 30 min, and supernatants were digested at 37°C for 30 min with 25 µl of 10 mg/ml RNase A. They were then phenol:chloroform (50:50 v/v) extracted, precipitated, washed in 70% ethanol, and resuspended in 0.2 N NaOH for 10 min to denature. This solution was neutralized with 1/10th volume of 2 M ammonium acetate, pH 4.5, and precipitated with 70% ethanol (f.c.). After centrifugation at room temperature for 5 min, washing the pellet twice with 70% ethanol and drying under vaccuum, the DNA pellets were ready for sequencing.

Sequencing was carried out by the dideoxy chain termination procedure as described (125), with some modifications. The above plasmid preparations were generally not sufficiently pure to use labelled α^{32}P-dATP in the extension reaction, so synthetic oligonucleotide primers (5 ng/sequencing reaction) were 5'-labelled with T4 Polynucleotide kinase (Promega), using a threefold molar excess of gamma γ^{32}P-dATP (New England Nuclear) at 37°C for 1 h. Primers were annealed to denatured double stranded DNA template at 37°C for 10 min in 10 µl volume, after which 2.4 µl of a 1/10 dilution of T7 DNA polymerase (Sequenase; U.S. Biochemical Co.) in water was added; 2.4 µl of this mix were then immediately added to an equal volume of the appropriate mix of deoxy- and dideoxy-nucleotide triphosphates for each base in four separate reactions. Extension was allowed to proceed at 37°C for at least 5 min, after which the reaction was terminated by addition of denaturing buffer containing formamide, glycerol and marker dyes. Sequencing reactions were electrophoresed on 50 cm long 6% polyacrylamide denaturing gels (6 M urea in Tris-Borate-EDTA buffer), at 2000-3000 V. Extra resolution was obtained either by use of wedge gels (0.2-1.2 mm) or by double loading of samples, with addition of extra running buffer prior to

the second loading. Gels were fixed in 10% acetic acid for 30 min, dried in a 65°C drying oven and exposed to X-ray film overnight or longer if necessary.

5.2 Sequence of First Open Reading Frame (ORF1)

5.2.1 Nucleotide sequence analysis

Sequencing showed the presence of two long potential open reading frames (ORFs) directed away from one another on opposite strands (Figure 16). The larger one is completely encoded within the 3 kb Sac 1 restriction fragment and extends from base 1938 to base 106 (total length 1832 bases in Sal-Bam orientation), coding for a putative protein of 610 amino acids, as shown in Figure 18. The larger open reading frame (ORF1) includes all the hybridizing region to the rat POMC probe, including the 1.1 kb Bam-Sal fragment, and will therefore be considered at some length.

A number of features of the sequence suggest that this ORF indeed codes for an intact and functional yeast gene, including all the surrounding canonical sequence motifs (152). The sequence context strongly suggests the indicated methionine as the first amino acid, with upstream promoter, enhancer and transcription initiation sequences present at appropriate distances for yeast. The A nucleotide three bases upstream corresponds to Kozak's rules (153) for the ATG to function as an efficient translation start signal. No upstream ATG sequences are present in any reading frame between this codon and the concensus promoter (TATA box) sequence (154). Stop codons are present in all three reading frames before the next upstream methionine is reached and the first downstream ATG occurs 67 codons later. The sequences GGTGG and AACAG, beginning 51 and 14 nucleotides upstream of the ATG translation start codon, conform to known consensus transcription initiation sequences in yeast - the

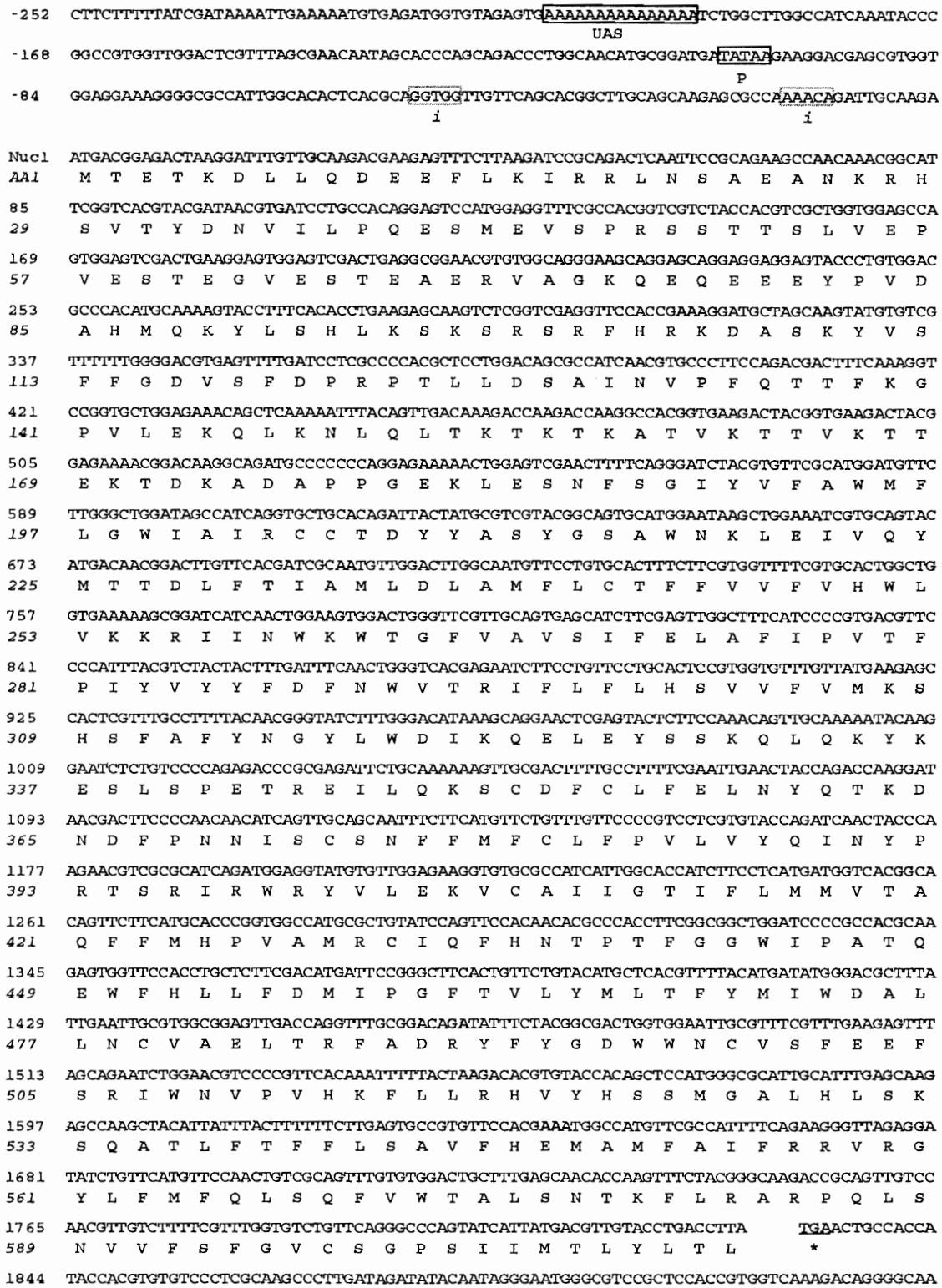


Fig. 18 Sequence of putative 610 amino acid protein (ORF1). Consensus start and stop codons are underlined. Nucleotide and amino acid numbers from likely start site are shown at left. Likely promoter sites (P), transcription initiation sites (*i*), and upstream activating sequences (UAS) are shown.

sequences RRYRR (where R represents either A or G, and Y represents C or T) and TC(A/G)A appear to act as initiation sites in about half of all yeast genes (155). However, initiation motifs in the other 50% of cases are at present poorly defined and it is therefore possible that other sites may also be used. The short distance between transcription and translation start sites is typical of yeast (155, 156). A "TATAA" motif lying 54 and 91 bases upstream respectively of the two transcription initiation sites also conforms well to the concensus promoter sequence ("Pribnow box") for eukaryotes (157, 158) and is almost certainly the promoter site. This distance is within the well described 40-120 base separation between promoter and transcription initiation sites usually seen in yeast (158, 159).

A concensus potential intron 5' splice site (CGTAYGT) is present at nucleotides 329-334 of ORF1 (160), raising the question of whether the nucleotide sequence is interrupted by an intron. This is important in assessing whether a nucleotide region with homology to the probe actually encodes protein (exon) sequence, or "irrelevant" intron sequence. Although final proof would rest on sequencing across this region in the mRNA, the presence of an intron is highly unlikely for the following reasons:

i) Introns are very rarely seen in yeast, particularly in nuclear genes coding for protein sequences (161).

ii) The chance likelihood of finding a particular six base sequence in a 2kb stretch of DNA is approximately 50%, whereas concensus splice sequences can often be found within both exons and introns and not used (160).

iii) Virtually all of the few known introns in yeast begin very close to the 5' end of the gene (within about 100 nucleotides), apart from special cases where the gene is alternately spliced and the intron includes coding sequences within a second transcript (161, 162). Here the intron splice site would be at codon 111.

iv) Other essential concensus sequences are missing, particularly the highly conserved branch point sequence (TACTAAC) (161). Further, most yeast intron sequences are extremely A + T rich, a feature not seen downstream of the potential splice site.

5.2.2 Homology between ORF1 and rat POMC

5.2.2.1 Nucleotide homology

Restriction mapping showed the region hybridizing to the rat POMC probe to lie over most of the large open reading frame, from the Bam H1 site at 609 through to the Sal 1 site at 1763. The nucleotide sequence was therefore compared with that of the rat POMC probe by the Search program for detecting sequence homologies in Intelligenetics. Allowing for potential mismatch gaps of three bases and a minimum alignment length of 30 bases, there were 32 matches at a 70% nucleotide homology level, 60 matches at a 65% level, and over 200 matches at a 55% homology level. Some of the best matches are shown in Figure 19. These all showed overall homology greater than 65%, and in some regions over 80%, quite sufficient to explain the binding to the probe at relatively high washing stringencies (0.2X/0.1%, 50°C). Further, almost all regions of the POMC probe showed significant nucleotide homology with at least one part of the long open reading frame; the homology was not just confined to one particular part of the probe.

When the rat POMC gene and the ORF1 nucleotide sequences were analyzed for overall sequence alignment by another Intelligenetics program (Genalign), the homology between the two nucleotide sequences was much less convincing, however. Although there were short regions of homology, there were also large gaps between these regions and they generally did not correspond to those regions of POMC which are most conserved in vertebrates.

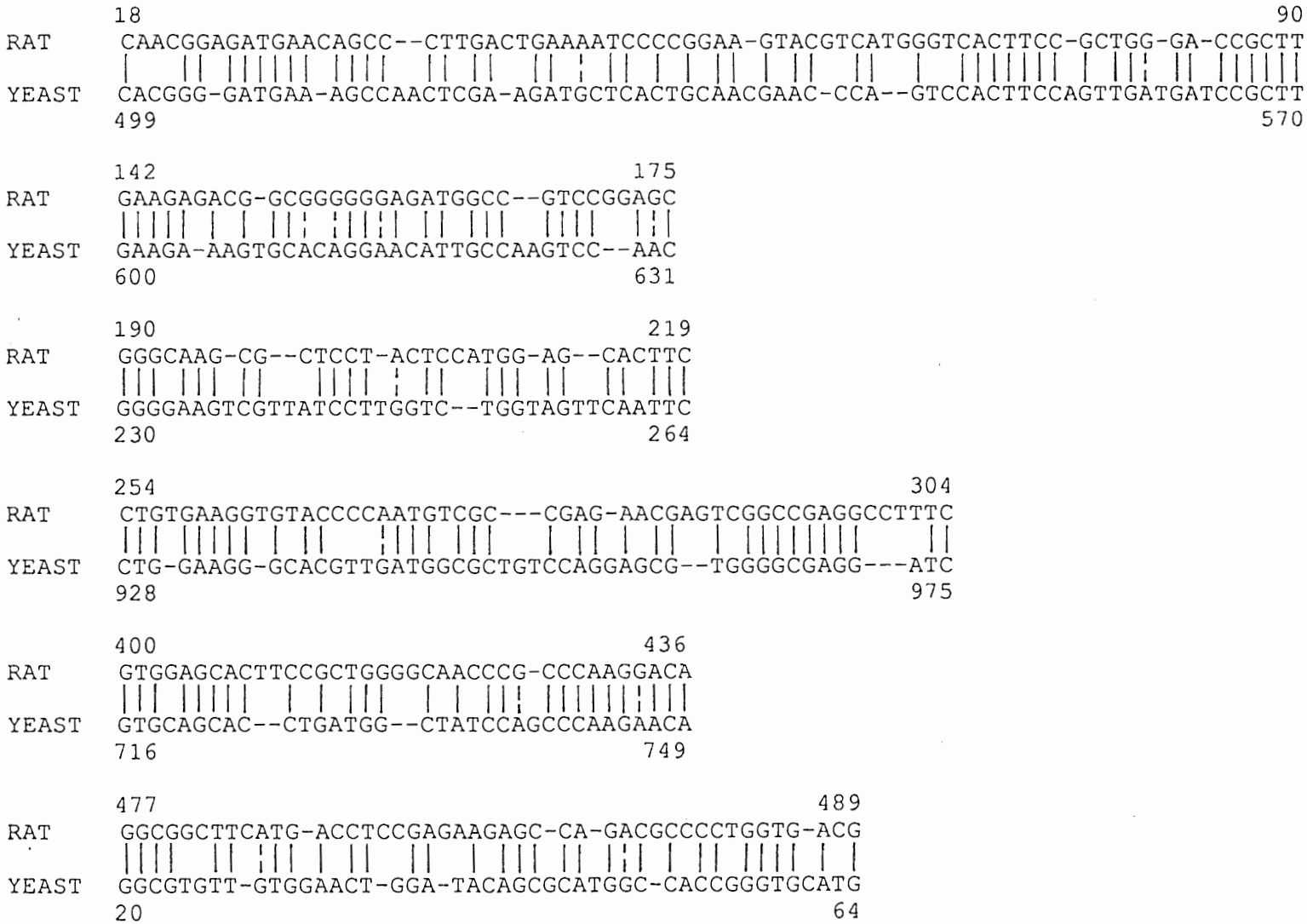


Fig. 19. Strongest regional nucleotide homologies between ORF1 and rat POMC gene.

To summarise, an overall level of homology which would suggest an *organizational* relationship between the two genes was not obvious (data not shown). I therefore looked for shorter regions of significant amino acid homology within the ORF1 and POMC genes.

5.2.2.2 Peptide homology

When short regions of the two peptide sequences were compared (using the Search homology program in Intelligenetics) the most significant region of homology was in the region of the vertebrate ACTH molecule (Figure 20). Here the two peptide sequences showed a 30% homology over a 50 amino acid region, although the highly conserved core sequence of vertebrate α-MSH (MEHFRW) was missing from ORF1. When the level of homology was analyzed by setting the window over still shorter regions, with only single gaps allowed in the analysis, the highest level of homology to the vertebrate α-MSH region was much further downstream towards the C-terminal end of the protein (residues 500 515) where 50% identity was present over a 16 amino acid stretch, although again two single gaps within the conserved α-MSH core sequence were needed. The level of homology was disappointing when no gaps were permitted, however, with no region having significant homology to either the α-MSH or Met-enkephalin sequences, even over short stretches. The best fit for α-MSH was a short sequence of nine amino acids very close to the C-terminal end of the yeast protein (residues 587-596), whereas the best fit to the vertebrate Met-enkephalin sequence was an even shorter insignificant sequence of only seven amino acids directly upstream (residues 558-564).

When the two peptide sequences were aligned using the Wisconsin GCG Bestfit program, the overall level of identity was 18.5%, with 45.1% chemical similarity (Figure 20). This was as high as any of the sequences identified from

Fig. 20. Alignment of amino acid sequence from ORF1 with rat POMC. Best regional matches for ACTH, α-MSH, and Met-enkephalin sequences are shown. Overall alignment using the Wisconsin GCG bestfit program is shown beneath.

the NIH (Genbank) and Swiss (EMBL) protein databases, (even though POMC was not listed; see below). Nevertheless most of the regions of identity were relatively short (3-4 amino acids) and did not correspond to the most conserved regions of the vertebrate POMC molecule, nor to the matches described above when shorter sequences were compared. Sequence comparison between the putative ORF1 protein and Aplysia sequences also failed to identify any significant matches.

5.2.3 Further analysis of ORF1

The nucleotide sequence of the 610 amino acid open reading frame codes for a potential protein of calculated molecular weight of 71.6 kD, with a calculated pI of 8.33. Further analysis of the sequence of this large open reading frame was performed to identify functional motifs and homology with other proteins, and thereby possibly gain insight into functional similarity with any vertebrate POMC-derived peptides.

5.2.3.1 Functional motifs

5.3.2.1.1 Dibasic cleavage sites. The vertebrate POMC precursor is spliced by enzymes recognizing pairs of dibasic residues spaced at the ends of product peptides. Because there appears to be a fundamental conservation of processing mechanisms and enzymes capable of correctly processing vertebrate POMC are present in yeast, it was therefore logical to ask whether pairs of basic residues were present and spaced at regular intervals in this protein. Although a number of pairs of basic residues were present, they did not appear to have a regular spacing to suggest cleavage into product peptides (Figure 21). There were, however, two regions in the N-terminal half of the ORF that were particularly basic in nature (residues 88-105 and 155-173). In each case over one third of the

```
1     MTETKDLLQD EEFLKIRRLN SAEANKRHSV TYDNVILPQE SMEVSPRSST   50
51    TSLVEPVEST EGVESTEAER VAGKQEQEEE YPVDAHMQKY LSHLKSKSRS  100
101   RFHRKDASKY VSFFGDVSFD PRPTLLDSAI NVPFQTTFKG PVLEKQLKNL  150
151   QLTKTKTKAT VKTTVKTTEK TDKADAPPGE KLESNFSGIY VFAWMFLGWI  200
201   AIRCCTDYYA SYGSAWNKLE IVQYMTTDLF TIAMLDLAMF LCTFFVVFVH  250
251   WLVKKRIINW KWTGFVAVSI FELAFIPVTF PIYVYYFDFN WVTRIFLFLH  300
301   SVVFVMKSHS FAFYNGYLWD IKQEIEYSSK QLQKYKESLS PETREILQKS  350
351   CDFCLFELNY QTKDNDFPNN ISCSNFFMFC LFPVLVYQIN YPRTSRIRWR  400
401   YVLEKVCAII GTIFLMMVTA QFFMHRVAMR CIQFHNTPTF GGWIPATQEW  450
451   FHLLFDMIPG FTVLYMLTFY MIWDALLNCV AELTRFADRY FYGDWWNCVS  500
501   FEEFSRIWNV PVHKFLLRHV YHSSMGALHL SKSQATLFTF FLSAVFHEMA  550
551   MFAIFRRVRG YLFMFQLSQF VWTALSNTKF LRARPQLSNV VFSFGVCSGP  600
601   SIIMTLYLTL 610
```

KEY:

- L L Leucine Zipper.
- G Asn glycosylation
- P Creatine kinase II phosphorylation site.
- PEST PEST Sequence.
- P Tyrosine kinase phosphorylation site.
- P Protein kinase C phosphorylation site.
- P cAMP phosphorylation site.

Fig. 21. Derived amino acid sequence and functional motifs of ORF1. Numbers represent amino acids from initial methionine residue.

residues were either lysine or arginine; the first region also contained two histidine residues, which are positively charged at neutral pH, adding to the overall net positive charge of this region. The possible significance of these regions relates to their potential to facilitate specific DNA binding by interacting with the negatively charged phosphate backbone of the DNA helix (163). These basic regions can be seen by their overall hydrophilic character in the hydrophobicity profile of the protein, and may be important in binding of this protein to the DNA helix, as discussed below.

5.3.2.1.2 Phosphorylation sites. There are consensus sites for phosphorylation by several protein kinases, as shown in Figure 21. While some of these sites may be unavailable for modification in the intact protein, they suggest that the tertiary structure and functions of this protein may be regulated by the reversible phosphorylation of one or more residues (164). In this light it is interesting that four potential protein kinase C phosphorylation sites lie in a highly basic part of the protein (residues 155-173) and may therefore directly influence the affinity for, and ability of this region to interact with, the negatively charged DNA helix (see below).

5.3.2.1.3 Leucine zipper motif. Perhaps the most interesting structural motif in this protein is the repeat of four leucine residues, regularly spaced every seventh amino acid between positions 318 and 339, approximately halfway between the two ends of the protein. This exact repetition suggests a leucine zipper motif (Figure 22). Numerous studies in other proteins show that the leucine zipper motif is involved in dimerization of a protein, either with an identical protein (homodimer) or with a different protein also having a leucine repeat (heterodimer) (165, 166). Although the exact secondary structure of a leucine zipper is still debated (167), a potential α-helix in the region is an important feature because it enables the leucine residues to align on one side of

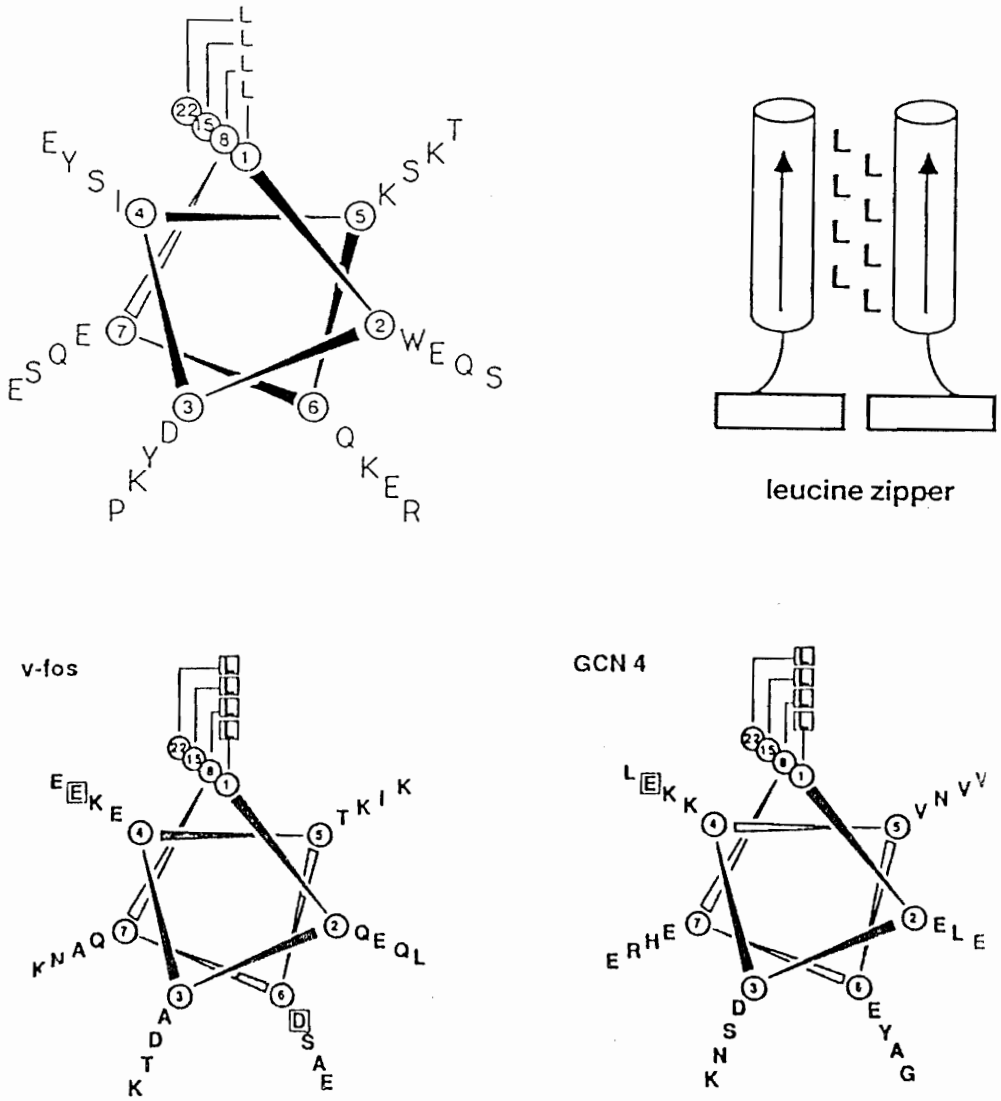


Fig. 22. Helical wheel of potential leucine ziper in 610A.A. protein compared with known leucine zippers. All show a large number of charged amino acids, around the repeated leucine backbone. A mechanism by which such a structure enables protein dimerization is also shown.

the helix (every second turn) and hence to interdigitate with the α-helix of an apposing leucine zipper. Secondary structure prediction by the Chou-Fasman algorithm indeed indicates an α-helical structure as most likely in this region (Figure 23). Additionally, leucine zipper interaction is stabilized by two other factors: i) an amphipathic profile of amino acids, with hydrophobic amino acids on one side of the helix and polar or charged amino acids on the other (165), and ii) a high proportion of acidic and basic residues that help confer stability on such a dimer through ion pairing between opposing charged residues. While the profile of the amino acid sequence is not strongly amphipathic in the leucine zipper region of ORF1, there are nevertheless a large number of charged residues, and the profile is quite convincing when compared to several other leucine zipper proteins known to dimerize (165; Figure 22).

Does this motif reveal anything about the function of this protein? Most, but not all, leucine zipper proteins appear to bind DNA (165, 166); those that do not appear to function as enzymes. The DNA binding proteins can be divided into two subsets. Those contacting the DNA helix immediately upstream of the leucine zipper have a high frequency of basic residues (lysine and arginine) in this immediate upstream region (scissors grip model; 165), whereas a second group contact DNA more distantly at basic regions generally upstream from the zipper motif. The two basic regions well upstream from the zipper motif in this protein suggest that it binds DNA and that it therefore probably belongs to the latter group.

5.3.2.1.4 PEST sequence. Rogers et al. (168) described the presence of regions rich in the amino acids proline (P), glutamic acid (E), serine (S) and threonine (T) in proteins that have very short intracellular half lives (two hours or less). These regions are usually flanked (but always uninterrupted) by positively charged amino acids and often contain sites for phosphorylation by casein

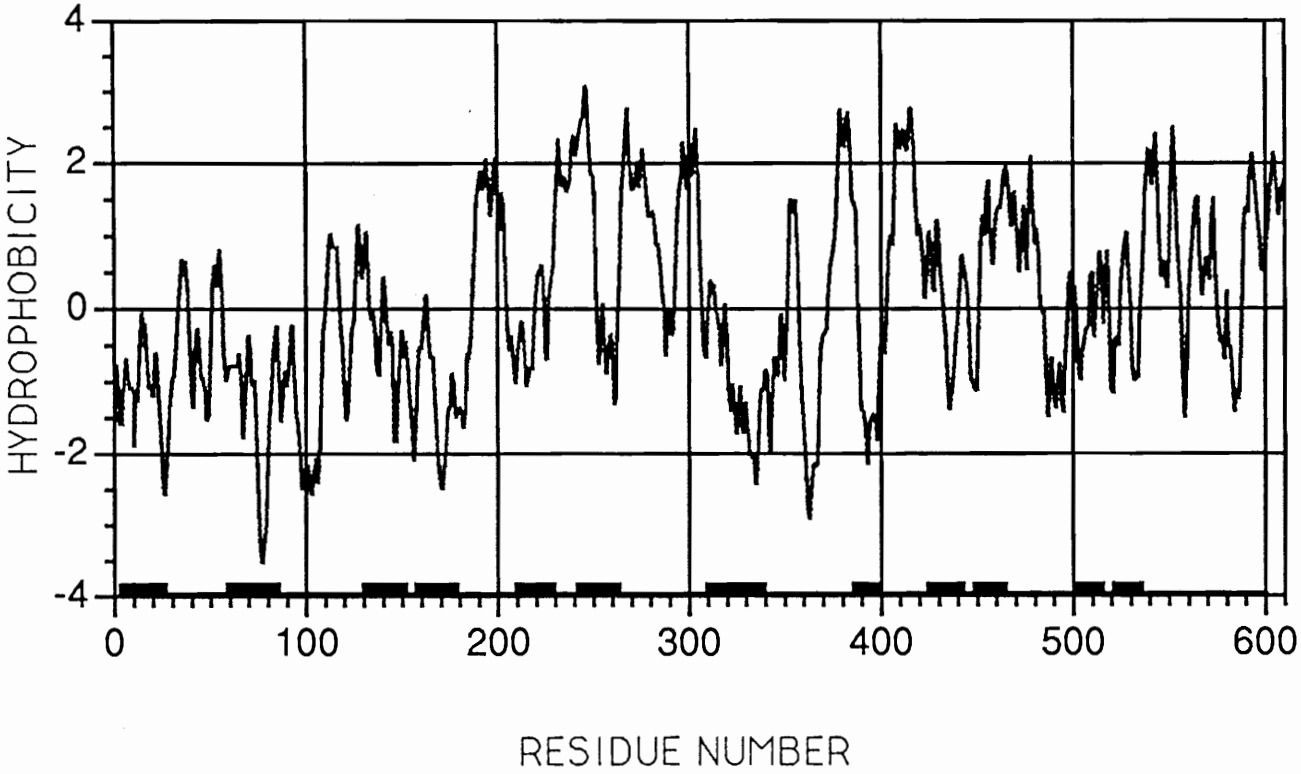


Fig. 23. Hydrophobicity profile and secondary structure analysis of the putative protein from ORF1. Regions of potential alpha helix (as determined by the Chou-Fasman algorithm) are shown as heavy bars along the x axis. The potential leucine zipper motif occurs between residues 310-340 in one of these alpha helical regions.

kinase II (169). This large open reading frame indeed contains a region in its N-terminal end containing two such stretches of amino acids. Between amino acids 47 and 70, 16 residues are of these four types (73%), a figure far exceeding the likelihood of probability (Figure 21). A second short stretch rich in acidic residues (particularly Glu) is also present shortly downstream between residues 74 and 86. This region also contains a high percentage (45%) of these four amino acids. Many rapidly degraded proteins have important regulatory functions, and the group includes such important proteins as *myc*, *myb*, *fos* and p53 (168). Such regions have also been postulated as potential calcium binding regions, once phosphorylated by casein kinase II (168). While most Ca^{2+}-binding proteins (calpains) reside in the cytoplasm (170), a number of the other PEST proteins, such as the oncogenes listed above, are also known to bind DNA, and a potential regulatory DNA-binding role for ORF1, as suggested by the leucine zipper and basic regions, is therefore quite consistent with the data.

5.2.3.2 Significant homologies. The predicted peptide sequence of ORF1 was analyzed for significant homology with other proteins in the current NIH and Swiss protein resource databanks. No protein showed 100% homology, nor was there greater than approximately 21% identity with any known protein; the best matches are listed in Table 9 (171). Of the best matches, at least two proteins have regulatory effects on transcription and translation, several are enzymes, and three are of presently unknown function. However, none contains a potential leucine zipper motif, as described above. The highest degree of homology among other yeast proteins is with the DNA-binding protein, rad9 (Figure 24; 181), and with an mRNA-binding protein, cytochrome B translational activator (177). The rad9 protein of yeast appears to delay progression of the yeast cell cycle from G2 to mitosis while DNA is repaired, and its activity can be dramatically increased by prolonged exposure of yeast cells to ultraviolet light. It is

Table 9. *Proteins with closest homology to 610A.A. protein*

Acc #	Name	Bestfit identity /similarity	# of 70% matches	Best (610) Matches (other)	Function (refs.)
S05638	Cuticle protein 8 (locust)	17.8% 40.4%	1	slvepvest slvepdgst	Structural protein of locust cuticle. Contains multiple tetrapeptide repeats (172).
A33513	Hypoth. protein-b (HIV virus)	43.6% 19.8%	6	See Fig 24.	ORF on minus strand of HIVvirus, transcribed in infected T-cells. Function uncertain (173).
A31305	ghf-1 transcription factor (bovine).	37.5% 19.6%	2	See Fig 24.	Pituitary transcription factor; activates GH and PRL genes. Homeobox protein- POU family (174,175).
A33489	Hypothetical 36k protein is891(anabaena)	42.5% 15.9%	6	hl-ks-ks ekqlknlq hlrkseks ekslkrlq	Insertion sequence in cyanobacterium (176).
cbs2$yeast	Cytochrome B translational activator (cbs2 gene; yeast).	41.7% 16.8%	3	slspetreil pr-sst slrpet-nil prmsst	Activates translation of cytochrome B mRNA binding to leader sequence (177).
A31916	cytochrome p450 aromatase (chicken)	46.6% 20.7%	6	sksrsrfh svvfvmk sksssvfh svfhvmk	Catalyzes androgen to estrogen conversion (178)
fdhf$ecoli	formate dehydrogenase EC1.2.1.2(E.coli)	42.5% 16.4%	4	wd-allncvae rssttsl wdeal-nyvae rs-ltsl	Formate dehydrogenase protein containing selenocysteine coded at UGA codon (179).
A05027	hypothetical chloroplast protein ORF1068 (liverwort)	43.9% 16.1%	12	lesnfsgi lfelnyqtkd legnfsgi lf-lnylqkd	Long hypothetical protein in Liverwort chloroplasts. Calc MW 127kD. Function unknown (180).
A32789	rad9 DNA repair protein (yeast)	42.9% 17.9%	10	See Figure 24.	DNA-binding. Arrests transition from G2 to mitosis in yeast, while DNA is repaired (181).
A33106	mastermind protein (D. melanogaster)	34.1% 16.3%	3	dikqeleyss kflrarp dikqelfyss kfl-krp	Neuronal differentiation protein (182)
Y9K8BPP22	Hypoth. 9.8kD protein (phage P22).	41.7% 17.9%	2	snfsgiy srsrfhr snfsgiy srs--hr	Hypothetical protein between RAL and GP17, Phage P22 (183)

A) Yeast rad9 DNA repair protein (Acc.#A32789)

```
115  GD-VSFD   P=4.6E-6        148  KNL-QL   P=9.1E-6
836  GDAVTFD   E=3.64          678  KNLMQL   E=7.15

153  TKT-KTK   P=4.6E-6        249  VHWLVKK  P=4.6E-6
717  TKISPTK   E=3.64          870  VH-LKKK  E=3.64

328  SSKQLQK   P=4.6E-6        345  EILQKS   P=9.1E-6
673  SS-ELQK   E=3.64          71   EIPQKS   E=7.15

364  DNDFPN-NI P=1.8E-7        388  QI-NYPRT P=1.6E-6
96   DNDRPNANI E=0.146         391  QIVNNPRT E=1.27

588  SNVVFSF   P=4.6E-6
1145 SNNIFSF   E=3.64
```

Bestfit (Wisconsin GCG): Identity 16.8%
Similarity 42.9%

B) Pituitary-specific transcription factor (PIT-1/GHF-1) (Acc#A31305)

```
12   EPLKIRR   P=4.9E-6        19   LNS-AEA  P=4.9E-6
138  EP-KVRR   E=0.849         15   LNSDASA  E=0.849

340  SPETREILQK P=1.33E-7
126  SPEIRE-LEK E=0.023
```

Bestfit (Wisconsin GCG): Identity 20.7%
Similarity 37.1%

C) Hypothetical protein- human T cell leukemia virus (Acc#33513).

```
 6   DLIQDEE   P=4.1E-6         61  EG-VESIEAER P=8.4E-10
76   DLLV-EE   E=0.639         228  EGEVESIEAER E=000

71   VA-GKQE-QE P=9.8E-8       176  APP-GE   P=8.0E-6
183  VARRKQEEQE E=0.015        126  APPRGE   E=1.26

177  PPGEK     P=9.0E-7
121  PPGEK     E=0.141
```

Bestfit (Wisconsin GCG): Identity 19.8%
Similarity 43.6%

Fig. 24. Homology of 610 amino acid protein (ORF1) with other proteins. A) rad9 DNA repair protein; B) Pituitary specific transcription factor PIT-1; C) hypothetical T cell leukemia virus protein. In all cases, yeast protein is uppermost. Identical residues have solid boxes; chemically similar residues have hatched boxes.

not, however, directly involved in mutational repair. The overall level of identity between rad9 and ORF1 by bestfit analysis (Wisconsin GCG) is only 17.9%, although by a more strict regional alignment there are at least seven regions of over 70% homology (Figure 24). Also in yeast, the cytochrome B translational activator protein is a mitochondrial protein binding to the upstream regulatory region of the mRNA message for the cytochrome P450 gene. Although this protein shows low overall homology (16.8% identity), stricter regional alignment also shows three regions of over 70% homology.

The transcription factor ghf-1 (pit-1) is a vertebrate pituitary-specific transcription factor binding to the upstream regulatory elements of growth hormone and prolactin genes (174, 175)(Figure 24). It also appears to be a developmental regulatory protein, containing two sequence motifs belonging to the 'homeobox' and 'pou' protein families (184). However, the greatest homology with ORF1 does not appear to lie within either of these conserved developmentally important regions. Put another way, ORF1 does not contain a recognizable 'homeobox' or 'pou' domain. The longest and most significant region of homology is also seen in rad9 (see above), but the functional significance, if any, of this is unknown.

Four of the most homologous proteins are hypothetical ones from open reading frames in relatively primitive organisms, including two viruses, a cyanobacterium, and chloroplast DNA of the liverwort Marchantia. Potentially most interesting among these is an open reading frame on the negative strand of the HIV virus (173), (Figure 24). Although this hypothetical protein is of currently unknown function, it is tempting to speculate that it too may have a role in transcriptional regulation.

5.2.4 Overall analysis of ORF1

The data do not indicate that ORF1 has a prohormone-type structure, in that there is no canonical signal peptide apparent and there are no regular dibasic cleavage sites. Rather, the data are more consistent with a regulatory and potential DNA-binding role, particularly in view of the leucine zipper motif and regions enriched in basic residues. Overall, the evidence also indicates that the protein product is probably present at low concentration in the cell, consistent with such a function.

How can these results best be interpreted in this light, and are such features able to exclude an evolutionary relationship between this protein and vertebrate POMC? The levels of homology found between POMC and ORF1 are generally low and the fact that one can change the sequence alignments depending on the parameters built into the computer program is a potential cause for skepticism in any analysis, particularly when the degrees of homology are low (15, 171). However, given the evolutionary distance involved, it is intriguing that the overall level of protein homology with POMC is as good as any of the proteins identified from the protein database by current sequence analysis programs. It is also interesting that when only short sequences are considered, the highest level of homology in the yeast protein is to ACTH (one of the most conserved regions of vertebrate POMC), and with insertion of gaps up to a 50% identity can be found for a 16 residue stretch and a 30% homology over a 50 residue stretch. Using the scrambled sequence approach of Doolittle (171), the 15 matches between the region of ORF corresponding to ACTH (shown in Figure 21) is significantly greater than would be expected by chance. By this method the mean number of random matches is approximately 6.4 with standard deviation of 1.4. This would place the 15 matches seen in ORF1 at over six standard deviations from the number of matches expected purely by chance.

The lack of overall sequence homology might seem disappointing, but is not unexpected considering the extreme evolutionary distance from vertebrates, and is also not inconsistent with the concept of protein modules being shuffled during evolution (see Discussion, Chapter 7).

Finally, no significant homology was identified between ORF1 and Aplysia sequences, nor was any homology to vertebrate corticotropin releasing factor (CRF) found. This suggests that the POMC-like sequences within the Aplysia FMRF gene are not on a direct evolutionary line between vertebrates and yeast, although the means by which sequences related to several quite different vertebrate genes became integrated within the FMRF gene remains an intriguing and rather disturbing mystery (91).

5.3 Sequence of Second Open Reading Frame (ORF2)

5.3.1 Nucleotide sequence analysis

The second potential open reading frame extends in the opposite direction on the alternate strand. There are two potential protein initiation sites, at either base 2326 or 102 bases (34 codons) downstream at base 2428, resulting in a potential protein of 275 or 241 amino acids respectively (Figure 25). This assumes the first stop codon is as shown; the last few bases beyond the Sac 1 site in this protein were only sequenced in one direction, so the final proof of this site as the termination codon would require sequencing of the alternate strand.

According to Kozak's rules (153) for efficient translation initiation, in both potential ATG translation start sites the 5'-sequence "A.n.n-" favors initiation, but neither ATG is immediately followed by the optimal G residue. Although (other things being equal) the most upstream ATG is generally favored for initiation (154), the context of the upstream ATG suggests that the more downstream ATG is actually the preferred start to the protein (158, 159). Two TATA (Pribnow

```
                    UAS
-300  GCCAGATTTTTTTTTTTTTTTTCACTCTACACCATCTCACATTTTTCAATTTTATCGATAAAAAGAAGAGATGAGC
-225  AGCACATACATAGTGTGATAAAGAGGAAAAAAGCAGAAGAAGCAGAAGAAGAAGAGCATAGTAAAAAAGTAGTGC
-150  TAAAAAGGTGCGGTACGGAAACACAGGAAATATATCAGAGAAGAAATAATGTCACGTCCTGAGGAGTTGGCACCA
                                   P                 i
-75   CCGGAGATTTTCTATAATGATAGCGAAGCACACAAGTACACGGGTTCGACCAGAGTGCAGCATATCCAGGCGAAG
                      P                              i                        i
Nucl                                                                      AA 20
ATG ACG CTG AGG GCG TTG GAG CTT TTG AAT CTG CAG CCG TGC AGT TTC ATT CTG GAT ATC
M   T   L   R   A   L   E   L   L   N   L   Q   P   C   S   F   I   L   D   I
61                                                                          40
GGG TGC GGG TCC GGA CTG TCT GGG GAG ATT TTG ACG CAG GAG GGA GAC CAT GTG TGG TGT
G   C   G   S   G   L   S   G   E   I   L   T   Q   E   G   D   H   V   W   C
121                                                                         60
GGT TTG GAT ATA TCG CCC AGC ATG CTT GCG ACC GGT CTT AGT AGA GAG CTG GAG GGC GAC
G   L   D   I   S   P   S   M   L   A   T   G   L   S   R   E   L   E   G   D
181                                                                         80
TTG ATG TTG CAG GAT ATG GGC ACC GGG ATA CCG TTC CGG GCG GGC TCG TTT GAC GCG GCT
L   M   L   Q   D   M   G   T   G   I   P   F   R   A   G   S   F   D   A   A
241                                                                        100
ATT AGT ATC AGT GCG ATC CAA TGG CTG TGC AAT GCG GAC ACT TCA TAC AAC GAT CCT AAA
I   S   I   S   A   I   Q   W   L   C   N   A   D   T   S   Y   N   D   P   K
301                                                                        120
CAG CGG TTG ATG AGG TTT TTC AAC ACA TTC TAT GCT GCA CTG AAG AAG GGA GGG AAA TTT
Q   R   L   M   R   F   F   N   T   L   Y   A   A   L   K   K   G   G   K   F
361                                                                        140
GTG GCC CAG TTC TAC CCG AAA AAC GAC GAC CAG GTG GAC GAC ATA CTG CAG TCT GCC AAG
V   A   Q   F   Y   P   K   N   D   D   Q   V   D   D   I   L   Q   S   A   K
421                                                                        160
GTG GCA GGG TTC AGT GGC GGG CTT GTG GTG GAC GAC CCA GAG TCT AAA AAG AAT AAG AAG
V   A   G   F   S   G   G   L   V   V   D   D   P   E   S   K   K   N   K   K
481                                                                        180
TAC TAC CTT GTG TTG AGC AGT GGG GCC CCA CCG CAG GGG GAG GAG CAG GTG AAT TTG GAC
Y   Y   L   V   L   S   S   G   A   P   P   Q   G   E   E   Q   V   N   L   D
541                                                                        200
GGT GTG ACC ATG GAC GAG GAG AAC GTC AAC TTG AAG AAA CAA CTG CGC CAG CGC TTG AAG
G   V   T   M   D   E   E   N   V   N   L   K   K   Q   L   R   Q   R   L   K
601                                                                        220
GGA GGC AAA GAC AAG GAG TCT GCC AAG AGT TTC ATT CTA AGA AAG AAG GAG CTC ATG AAA
G   G   K   D   K   E   S   A   K   S   F   I   L   R   K   K   E   L   M   K
661                                                                        240
AGA CGT GGG AGG AAA GTT GCG AAG GAC TCC AAG TTC ACC GGG AGG AAA AGA AGA CAC AGG
R   R   G   R   K   V   A   K   D   S   K   F   T   G   R   K   R   R   H   R
721
TTC
F
     *
724  TAGAAGAAACACACATCCTCACACGCATATTCTTTTGTATACTATAAAAATACACACTTATACATTTCCCTCTGT
799  CTCTGTCTCCTTTCTACATGTATACTTTTTTTGTATATGTATACTTGTATATATACACTTATATCTGTATGTGAGTG
874  GTCAATGGCTCCGCGCACTTTTATGTAGTGATGTATTTTCCTTTTTTTTCTTGGTTTTCAATTTTGGTGCTTTGGA
949  ACGCCCAGTAAAGAGAAGAAGAAGAATCGAGTCTG
```

Fig. 25. Sequence of putative protein from ORF2. The putative start and stop codons are underlined. Nucleotide numbers from the putative ATG start site are shown on the left, while amino acid numbers are shown at right. Potential upstream activating sequences (UAS) and promoter sites (p) are boxed in solid lines. Potential transcription initiation sites (i) are shown in hatched boxes.

box) motifs are present at bases 2308 and 2364, and if the downstream ATG is used these would lie at 40-50 and 100-110 bases upstream from the putative yeast transcription initiation sequences, TC(G/A)A and RRYRR (Figure 25; 159). Such distances lie within the published range for distances between TATA elements and transcription initiation sites in yeast (40-120 nucleotides; 158). However, only 18 nucleotides separate the upstream TATA box and the ATG, and a putative transcription initiation site would need to lie between these two. Such close proximity would strain the bounds of credibility, assuming current ideas in yeast are correct (158, 159). Nevertheless, somewhat shorter distances can be forced to work experimentally (185), TATA promoter sites are apparently not essential in all cases (186), and the concensus transcription initiation motifs listed above are relatively poorly defined and only found in about 50% of cases (155, 158). Therefore, while the more upstream ATG cannot unequivocally be excluded as the protein start site, the contextual sequence evidence weighs overall in favor of the more downstream ATG. It is, of course, also possible that the more upstream TATA motif acts as a promoter additively or synergistically with its downstream TATA counterpart to promote transcription initiation from the downstream sites. Finally, the presence of two such elements also suggests that they may possibly have different biological roles, such as promoting constitutive versus regulated expression, as already found for the *HIS3* gene in yeast (158, 159).

5.3.2 Further analysis of ORF2

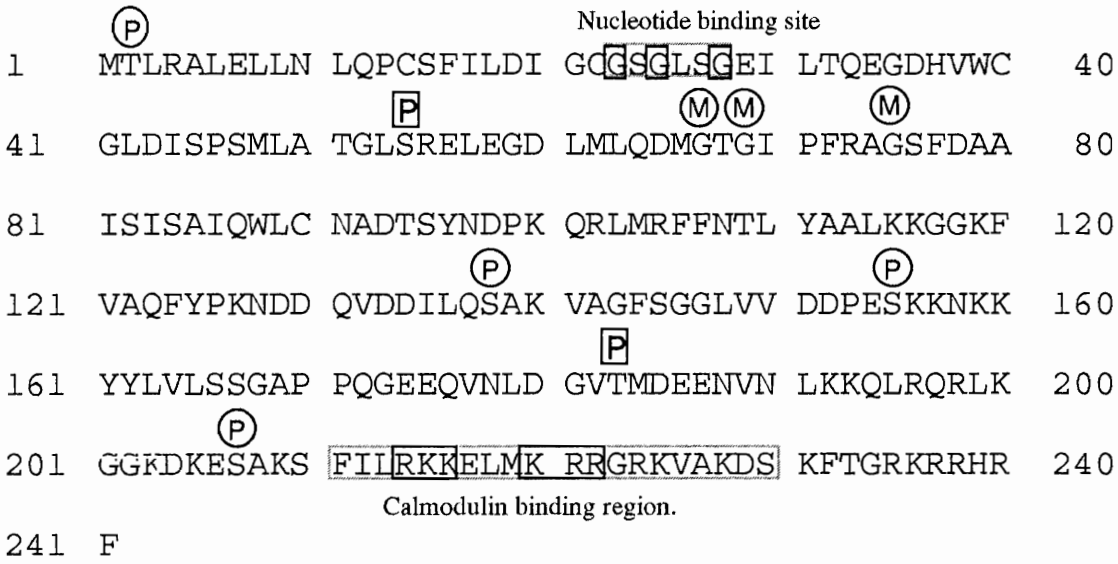
The second putative open reading frame did not lie in the region hybridizing to the POMC gene and is therefore of only indirect interest. It will therefore be considered only briefly in the following section.

5.3.2.1 Concensus motifs

This protein contains two interesting concensus motifs. The sequence G-X-G-X-X-G (residues 63-68) is a potential nucleotide binding site, in typical N-terminal location within the protein (187). Second, the basic region at the C-terminal end of the molecule strongly resembles a concensus calmodulin binding site (188). Potential phosphorylation sites in the primary sequence of ORF2 also indicate that the protein may be posttranslationally modified by phosphorylation to regulate its tertiary structure and function . Potential myristylation sites (189) also indicate that it may also be anchored to cell membranes, although myristylation is not obligatory if this sequence is present, nor is membrane attachment a requirement for activation by protein kinase C (190) (Figure 26).

5.3.2.2 Homology to other proteins

A search of the protein databanks showed two particularly interesting matches (Figure 27). The first was to skeletal myosin light chain kinase (MLCK), a protein catalyzing the reversible phosphorylation of myosin. However, overall (Bestfit) homology was low and the most significant region was the calmodulin binding site at the tail end of MLCK (177, 178). Although the two other matches both lay within the catalytic region of the molecule, neither is known to have any important function and concensus sequences for protein kinases were absent from ORF2 (191), making this protein unlikely to have a kinase function. In this light, the interesting homologies in its midportion with the *E. coli* gyrA protein (192) and with a hypothetical protein from the *Staphylococcus aureus* transposon tn554 suggest that it has a function related to DNA unwinding, which would be ATP-dependent and possibly regulated by the Ca^{2+}/Calmodulin system (188).



KEY:

Ⓟ Protein kinase C phosphorylation site

P Casein kinase II phosphorylation site.

Ⓜ Myristylation site.

Fig. 26. Sequence and functional motifs of putative ORF2 protein. Numbers represent amino acids from initial methionine residue.

A) RABBIT MYOSIN LIGHT CHAIN KINASE (MLCK; SKELETAL). (ACCESSION NUMBER A25830)

```
48      MLATGLS                 P=6.2E-6
486     MLLSGLS                 E=0.613

113     ALKKGGKF                P=2.2E-6
301     AL-GGGKF                E=0.222

209     KSFILRKKELMKRR          P=5.9E-11
567     KSQILLKKYLMKRR          E=0.000
```

BESTFIT WITH MLCK (WISCONSIN GCG) IDENTITY 27.0%
SIMILARITY 52.2%

B) HOMOLOGY WITH POTENTIAL DNA GYRASES:

Gyr A (E. coli; #S03757) & tn554 (S. aureus; #F24584).

```
Gyr A(E.C)  59  KKVLDVGCGGGILAESMAREGATV-TGLDMGFEPLQVA   95
ORF2 241AA  50  SFILDIGCGSGLSGEILIQEGDHVWCGLDISPSMLATG   87
tn554(S.A)  37  GRALDIGCGSGLLVEKLASYYDEV-VGIDISNQMLDLA   73

Gyr A(E.C)  96  KLHALESGIQVDYVQETVEEHAAKHAGQYDVVTCMEML   133
ORF2 241AA  88  LSRELEGDLMLQDMGTGI----PFRAGSFDAAISISAI   121
tn554(S.A)  74  KSKRQLTNTVYLNMNAEQ----LNFNEKFDFIVSRTTF   107

Gyr A(E.C) 134  QWLCNADTSYNDPKQRLMRFFNTLYAALKKGGKFVAQF   171
ORF2 241AA 122  EHVPDPQSVVRACAQLVKPGGDVFFSTLNRNGKSWLMA   159
tn554(S.A) 108  HHLDDIASVIQQMKE----------LLNEEGRIV---   131
```

Fig. 27. Homology of putative ORF2 protein with other proteins. A) Rabbit myosin light chain kinase (MLCK). Yeast putative 241 amino acid protein is uppermost. B) *E. coli* gyr A protein and *S. aureus* tn554 transposon protein. In both cases, identical residues have solid boxes; chemically similar residues have hatched boxes.

Further evidence to support a possible regulatory role for both proteins is the low degree to which they conform to the normal codon usage pattern seen in yeast, with generally more even usage of synonymous codons than is the norm (Table 10). Even allowing for overall interspecies differences in codon preference, there is a strong relationship in any one species between codon composition and mRNA expressivity of different genes. Those with the highest translation rates tend to have highly biased codon frequencies, whereas regulatory proteins are expressed in far lower amounts in the cell and therefore do not have the selective pressure to contain codons with abundant tRNA's (193, 194). The fact that both genes have a similar deviation from the norm may also indicate that they are expressed in a fairly similar rate (ignoring other regulatory influences), and may therefore somehow function in a functionally coordinated fashion.

5.4 Poly(dA-dT) Enhancer Sequence

Upstream concensus binding sequences for several known yeast transcriptional regulatory proteins (such as GAL4, GCN4, and HAP1 binding proteins) appear to be absent from the 5' regions of both ORF1 and ORF2(155). However, the presence of a long 15 nucleotide stretch of A-T residues 200 bases upstream from ORF1, and 300 bases upstream from ORF2, is extremely interesting (Figures 18 and 25). Although this sequence motif is common in yeast and appears to function as an upstream activating sequence (UAS) or enhancer (158, 195), even more interesting is its presence at an "appropriate" (80-600 bp; 156) distance from the start of two different open reading frames in opposite orientation, and its ability to function as an enhancer in either orientation (195) through the binding of an activator protein (196). Such a situation has already been described for the *his3* and *pet56* genes in yeast, two

Table 10. *Codon preferences in yeast open reading frames*

Amino Acid	Codon	Percentage of Total. Database	ORF 1	ORF 2	Amino Acid	Codon	Percentage of Total. Database	ORF 1	ORF 2
Met	ATG	100	100	100	Gly	GGT	70	10	13
						GGC	14	42	22
Trp	TGG	100	100	100		GGA	9	16	17
						GGG	7	31	49
Asn	AAT	46	27	44					
	AAC	54	73	56	Pro	CCT	26	9	13
						CCC	10	48	12
Asp	GAT	59	36	22		CCA	57	26	25
	GAC	41	64	78		CCG	7	17	50
Cys	TGT	73	25	25	Thr	ACT	41	25	13
	TGC	27	75	75		ACC	29	25	50
						ACA	22	9	12
Gln	CAA	78	32	17		ACG	8	41	25
	CAG	22	68	83					
					Val	GTT	47	19	9
Glu	GAA	78	32	0		GTC	28	17	9
	GAG	22	68	100		GTA	11	0	0
						GTG	14	64	82
His	CAT	55	13	50					
	CAC	45	80	50	Arg	CGT	18	4	8
						CGC	3	18	13
Lys	AAA	44	31	39		CGA	4	7	0
	AAG	56	69	61		CGG	2	14	13
						AGA	60	30	33
Phe	TTT	49	36	27		AGG	13	18	33
	TTC	51	64	73					
					Leu	TTA	24	9	0
Tyr	TAT	44	22	20		TTG	42	35	42
	TAC	56	78	80		CTT	40	7	18
						CTC	4	14	4
Ile	ATT	52	20	50		CTA	12	4	4
	ATC	33	70	30		CTG	8	31	32
	ATA	15	10	20					
					Ser	TCT	37	9	22
Ala	GCT	49	15	13		TCC	21	15	11
	GCC	26	40	26		TCA	16	6	6
	GCA	19	33	13		TCG	6	33	11
	GCG	6	12	48		AGT	13	13	39
						AGC	7	24	11

unrelated genes in which a common 17 base pair poly(dA-dT) element serves as the upstream promoter element for constitutive expression of both (195). It is tempting to speculate that this is indeed the case for these two unknown open reading frames. Consistent with this hypothesis are the codon frequencies in ORF1 and ORF2, which resemble one another in not showing strong bias, indicating neither gene is likely to be expressed at a high level. However, this sequence may also act by excluding nucleosomes, rather than by binding specific proteins, thereby enabling transcription factors more ready access to DNA (195).

5.5 Summary

Even though the second open reading frame showed no homology to the vertebrate POMC probe, there are a number of reasons to indicate that the proteins coded by the two open reading frames may well have an interrelated function. These include the following: a) By homology and sequence criteria, both have a function which seems most compatible with a DNA binding activity, the larger (ORF1) because of its leucine zipper motif and upstream regions enriched in basic residues, and the smaller (ORF2) because of its sequence similarity to other proteins having putative DNA gyrase activity. b) Both have a similar and unusual spectrum of codons, and ORF1 has a strong PEST sequence (168). These features indicate that neither is likely to be expressed in large amounts in the cell and they may have short half lives, features that are indeed consistent with a regulatory function. c) They both appear to be potentially under the control of a long 15 residue enhancer sequence of poly(dA-dT) residues, an organization already described for at least one other yeast system (195), and possibly used more widely than currently appreciated.

CHAPTER 6

STUDIES TO DETERMINE SIGNIFICANCE OF ORF1

6.1 Expression: Northern Blot Analysis

Yeast total RNA was prepared by the phenol extraction method (125). All solutions were treated with diethylpyrocarbamate (DEPC; Sigma), sterile plastic containers and pipettes were used, and gloves were worn throughout. Cells were harvested and rapidly cooled in mid log phase by centrifugation in plastic tubes half filled with ice. After resuspension in a small volume of aqueous LETS buffer, they were lysed by repeated vortexing with acid washed glass beads (0.45 μm; Sigma) in the presence of cold buffered phenol. The upper aqueous phase was removed after centrifugation, carefully leaving the interface. After reextraction twice with phenol:chloroform (50:50 v/v) and twice with chloroform, the RNA/DNA mix was precipitated in 0.1 M LiCl/ 70% ethanol at -20°C overnight, and resuspended in 10 mM TrisHCl, 1 mM EDTA, pH 7.6 (TE). Purified mRNA was prepared by passage of this preparation through an oligo-dT cellulose column as previously described (126).

Either 40 μg of yeast total RNA or 200 ng of mRNA were separated by electrophoresis overnight in 1% NuSeive agarose (RNase-free; FMC Biologicals) in 2X MOPS/Formaldehyde buffer as described (126). RNA was denatured by heating to 65°C for 5 min in formamide/formaldehyde loading buffer and prestained for 5 min on ice with ethidium bromide (100 μg/ml f.c.) prior to loading. The gel was washed repeatedly with 1X SSC to remove formaldehyde prior to

transfer onto charged Nylon (Hybond N) by the previously described method (125). The yeast DNA was probed with the entire 3 kb insert from clone pTZ19h using (20% formamide, 37°C). Results are seen in Figure 28. The yeast RNA showed two hybridizing bands of 2.4 kb and 1.6 kb. Potentially these could represent two separate length transcripts from one open reading frame; however the smaller transcript is too short for ORF1 and a far more likely explanation is that the larger is the primary transcript from ORF1 and the smaller that from ORF2. This experiment therefore indicated that ORF1 and ORF2 are expressed during log phase and that neither is a pseudogene (as might be expected, in view of the rarity of pseudogenes in yeast; 161).

6.2. Genetic Manipulation Experiments

6.2.1 ORF1 overexpression protocol

A restriction fragment containing the large open reading frame was directionally subcloned into the vector pYES2.0 (Invitrogen), a high copy number expression vector using the replication origin from the endogenous yeast 2 μm plasmid (197). The protocol was as follows (198, 199). A 2011 bp fragment was excised between Eco R1 and Nar 1 restriction sites; the 5' Nar 1 restriction site was filled in with Klenow enzyme to make a blunt end prior to Eco R1 digestion. This fragment was then directionally cloned into Eco R1 and blunted Hind III sites within the multicloning region of the pYES2.0 vector.

The subcloned fragment did not include the upstream promoter sequence, being placed entirely under the control of a strongly galactose-inducible or glucose-repressible promoter (Gal promoter) in the vector. No upstream AUG sequences were present in any reading frame between the gal promoter and the 5' translation initiation codon, thereby avoiding extraneous (and possibly out of frame) initiation sites. This plasmid construct was transformed into *E. coli* (strain

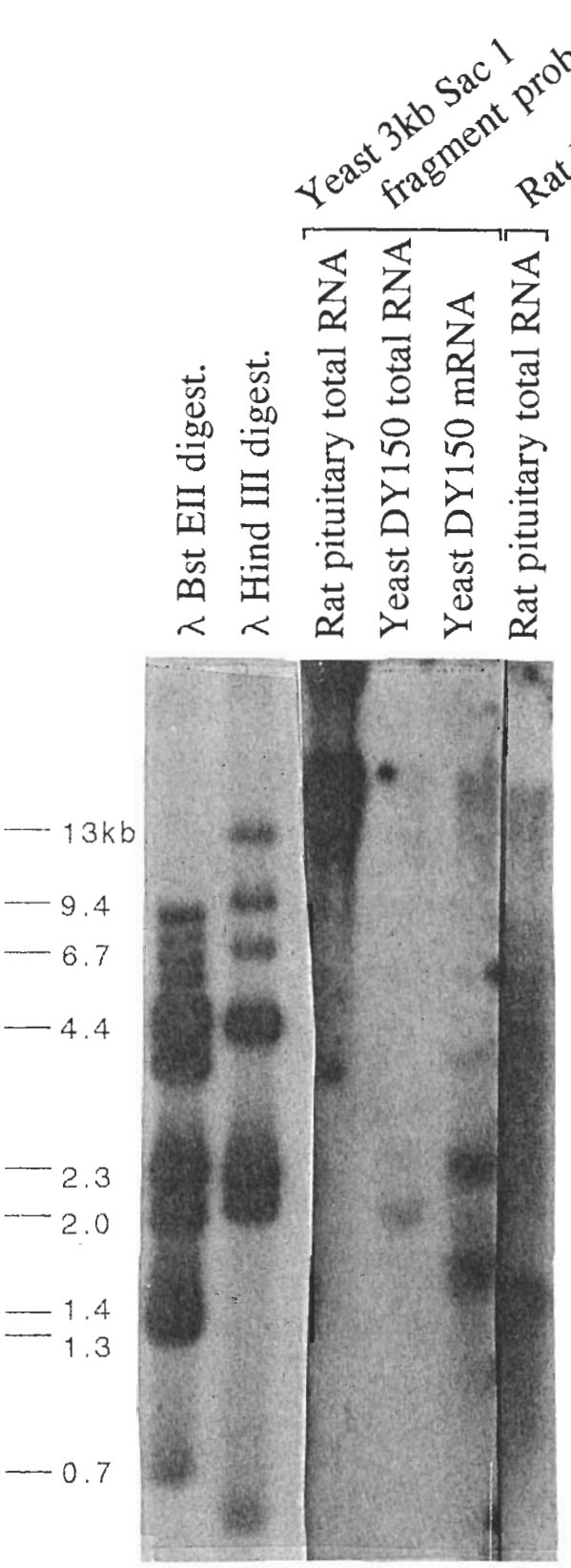


Fig. 28. Northern blot of yeast RNA probed with 3kb Sac 1 fragment. Size markers and control probe of rat pituitary RNA with rat POMC are shown.

71-18), and clones were selected using ampicillin resistance. Correct size and orientation of the insert DNA was verified in several ampicillin-resistant clones by a) restriction digests of plasmid minipreps, and b) sequencing of the plasmid minipreps through the ligation sites and into the insert. One hundred nanograms of plasmid DNA was then used to transform yeast cells of the original haploid strain (DY150). Transformants were selected using the URA3 marker on the vector, which conferred the ability to grow on uracil-deficient plates. Yeast colonies showed no difference in size, color or apparent growth rate on selectable solid (agar) medium from colonies transformed with vector alone (without ORF1 insert). Two other control "overexpressed"plasmid constructs were also made and these are detailed below.

6.2.1.1 Rat POMC Exon 3

A construct containing the terminal 530 bases of the rat POMC gene, including sequences coding for the α-, β- and γ-MSH peptides, ACTH and β-endorphin; this sequence was identical to that used as a hybridization probe and was generated by PCR of the rat gene using DNA primers with appropriate external restriction sites for convenient cloning into the expression vector (3' Eco R1 and blunted 5' Xba 1 sites were ligated to Eco R1 and blunted Hind III sites in the vector). Positive clones were again selected by ampicillin resistance and the size, orientation and sequence of the inserts were verified by restriction digestion and sequencing from vector into the insert from both sides. It should be noted that the primary translation product should begin with the methionine residue (#53) in γ-MSH; because it does not, however, include a signal peptide sequence, it is therefore not likely to be processed through the rough endoplasmic reticulum, and secreted into the medium.

6.2.1.2 β-galactosidase (Lac Z) gene

A third construct was made in which the lac Z (β-galactosidase) gene was inserted into the expression vector (200). A 4 kb Dra 1(blunted) /Sal 1 fragment containing all of the cloned β-galactosidase gene including the initial methionine but excluding the promoter region, was again directionally cloned into pYES2.0, using Xho 1 and blunted Xba 1 sites within the multicloning region of the vector, and transformed into *E.coli* 71-18 cells. Positive transformants showed a dark blue color when grown on plates selecting for both ampicillin resistance and the presence of β-galactosidase activity (X-gal/IPTG plates). Plasmid DNA was then transformed into DY150 cells using the lithium acetate procedure (201), with positives selected by restoration of URA3 activity on uracil-deficient plates.

6.2.2 ORF1 deletion protocol

A large section of the coding region of ORF1 was deleted and replaced with the LEU2 gene by homologous recombination. A 1 kb fragment between Nar 1 (base#2011) and Ava 1 (964) sites was excised from the vector and the Nar 1 restriction end was filled in (blunt-ended) with Klenow enzyme. In its place was inserted a 2 kb Sal 1/blunted Hpa 1 fragment, containing a functional LEU2 gene which carried its own promoter, and therefore was not dependent on upstream sequences for its transcription. Plasmid constructs were transformed into *E. coli* (71-18) and positive clones selected using ampicillin resistance. Clones containing the insert were identified by Sac 1 restriction digests of plasmid miniprep DNA which were 4 kb in size, rather than the 3 kb in those not containing the new insertion. Approximately 100 ng of this 4 kb linear DNA construct (between Sac 1 sites excluding vector) was transformed into both haploid (YD150) and diploid yeast (M226) strains by the lithium acetate

procedure (126). The diploid strain was a kind gift of Dr M. Sandbarken (Howard Hughes Medical Institute, University of Utah), with genotype as shown in Table 11.

Positive transformants from each strain were selected for their ability to synthesize leucine by growth on leucine-deficient plates (i.e., possess functional LEU2 gene). This required integration of the linear DNA into the yeast genome, since the linear fragments themselves were unable to autonomously replicate. A total of 14 clones were obtained for the haploid transformation and 8 clones for the diploid. Southern blots were performed to verify both deletion of the endogenous genomic ORF1 sequence and correct integration of the LEU2 gene in its place. This was needed to prove that integration had not occurred at either the LEU2 locus or close to a yeast autonomously replicating (ARS) sequence (201) and also that duplication of ORF1 had not occurred during integration of exogenous DNA. Five separate haploid yeast colonies were grown in leucine deficient medium and genomic DNA was prepared. DNA was digested with Eco R1 and Sac 1 enzymes, electrophoresed and transferred to charged nylon in two identical blots, all as previously described. An equal amount of wild-type

Table 11. *Genotype of M226 yeast strain*

Diploid	
Homozygous	LEU 23,12
Homozygous	URA 3,52
Trp	1-7/wt
His	4-713/wt
Met	2-1/wt
Lys	2/wt

genomic DNA was also digested with Eco R1 and Sac 1, and included in the electrophoresis and transfer steps as a control, together with radioactive molecular size markers. The blots were then hybridized with two different oligonucleotide probes, both 5' end-labelled with T4 polynucleotide kinase. The first probe hybridized to a unique sequence within the 4 kb Sac 1 fragment, but lying external to the inserted LEU2 gene, while the second hybridized within the region of DNA previously deleted and replaced with LEU2 (i.e., within ORF1). Both blots were washed at high stringency in their final wash (0.2X SSC/0.1% SDS at 65°C). Results are seen in Figure 29.

The first blot (Figure 29A) shows hybridization using the probe external to the deleted region. It confirms that the Sac 1 fragment increased in size from 3 kb (seen in the wild type control) to 4 kb (indicating the new insertion), whereas the hybridizing Eco R1 fragment has decreased in size to about 3.5 kb by virtue of the introduction of a new Eco R1 site within the LEU2 gene. The second blot (Figure 29B) shows the hybridization pattern using the oligonucleotide probe recognizing a sequence within the deleted 1 kb of ORF1. Whereas the control genomic DNA showed a hybridizing band in each lane the same size as the rat PCR probe, the five transformants showed no hybridizing bands, indicating that this sequence had indeed been deleted from the genome. Collectively, these data indicate both that most (1 kb) of ORF1 had been deleted in all five transformants, and that the LEU2 gene had indeed been integrated in its place into the genome.

6.2.3. Effects of ORF1 overexpression and deletion

6.2.3.1 Effect on ACTH immunoactivity

Yeast DY150 cells transformed with the overexpression plasmid pYES2.0, containing ORF1, were grown in defined medium in the presence of glucose (in

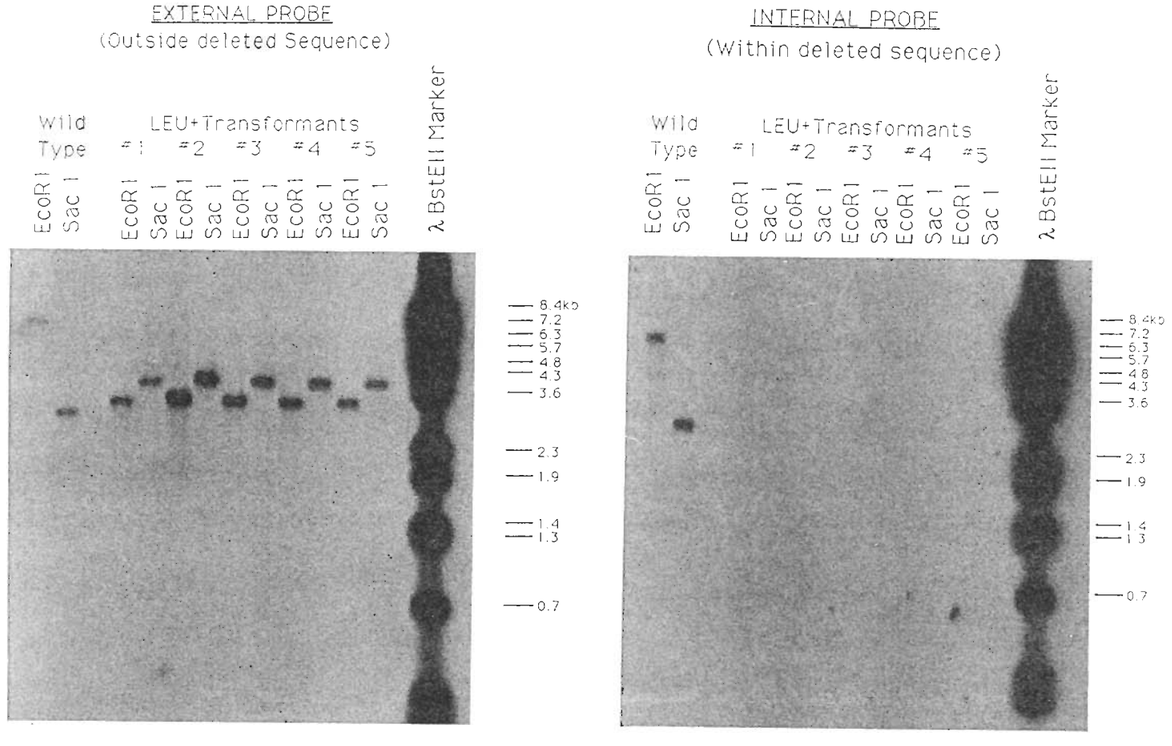


Fig. 29. Southern blots to test knockout of ORF1. Genomic DNA from wild type DY150 strain and from five colonies with potential ORF1 knockout was digested with Eco R1 and Sac 1 restriction enzymes, and electrophoresed over-night in 0.7% agarose in 1X TAE buffer, with molecular size markers, and transferred to nylon. Both were then probed at high stringency, the left with an oligonucleotide outside the deleted region of ORF1, and the right with an oligonucleotide lying just outside ORF1 and within the 3 kb Sac 1 fragment. Final washes were 0.2X SSC, 0.1% SDS at 56°C, with exposure for 6 days at -70°C under intensification.

which transcription from the gal promoter was strongly repressed) or galactose (in which transcription from the gal promoter was strongly stimulated). The cells were grown to the end of log phase, washed, harvested and snap frozen in liquid nitrogen. The cell walls were broken by homogenization in a Waring blender under liquid nitrogen for five successive one minute intervals, until a fine cream-colored powder was produced. This was resuspended in five volumes of glacial acetic acid, extracted and assayed in West and Orth polyclonal assay systems as described previously. The assays showed that the strains carrying the over-expressing plasmid had no effect on extractable ACTH-like activity over controls, whereas induced expression should have increased the amount of protein product by 20-100 fold (202); this indicated that ORF1 was unlikely to encode the protein responsible for the ACTH immunoactivity.

6.2.3.2 Effect on growth velocity

Growth rate in defined medium was checked by measurement (O.D.600 nm) of the rate of change of cell density in the following yeast strains, grown in appropriate selective media. In all cases, 10 ml of medium was inoculated with 50 µl of cells previously grown to saturation and the tubes were incubated at 30°C with vigorous shaking (250 rpm).

a) Wild type DY150 cells (all nutrients present).

b) ORF1 knockout (DY150 cells in which LEU2 had been inserted by homologous recombination; grown in leucine-deficient medium).

c) ORF1 overexpresssion (in plasmid pYES2.0 in transformed DY150 cells; uracil-deficient medium).

d) DY150 cells transformed with the pYES2.0 vector (no insert)

e) DY150 cells overexpressing the C-terminal 530 bases of the rat POMC gene cloned in pYES2.0. (uracil-deficient medium).

f) DY150 cells overexpressing the lac Z gene cloned in pYES2.0 (uracil-deficient medium).

The results of this comparison are shown in Figure 30, in which cell density is expressed at increasing times after innoculation. While all cells took longer to enter log phase when grown in the presence of galactose as a carbon source, the rate of growth in log phase showed no significant differences, either between constructs for each carbon source or for the same construct grown in different carbon sources. This suggested that the potential gene coded by ORF1, despite being in single copy, was not critical in influencing the growth rate under the conditions mentioned.

In a separate experiment performed under identical conditions, the addition of synthetic porcine ACTH(1-39) to the medium had no effect on growth rate compared to control DY150 cells (data not shown).

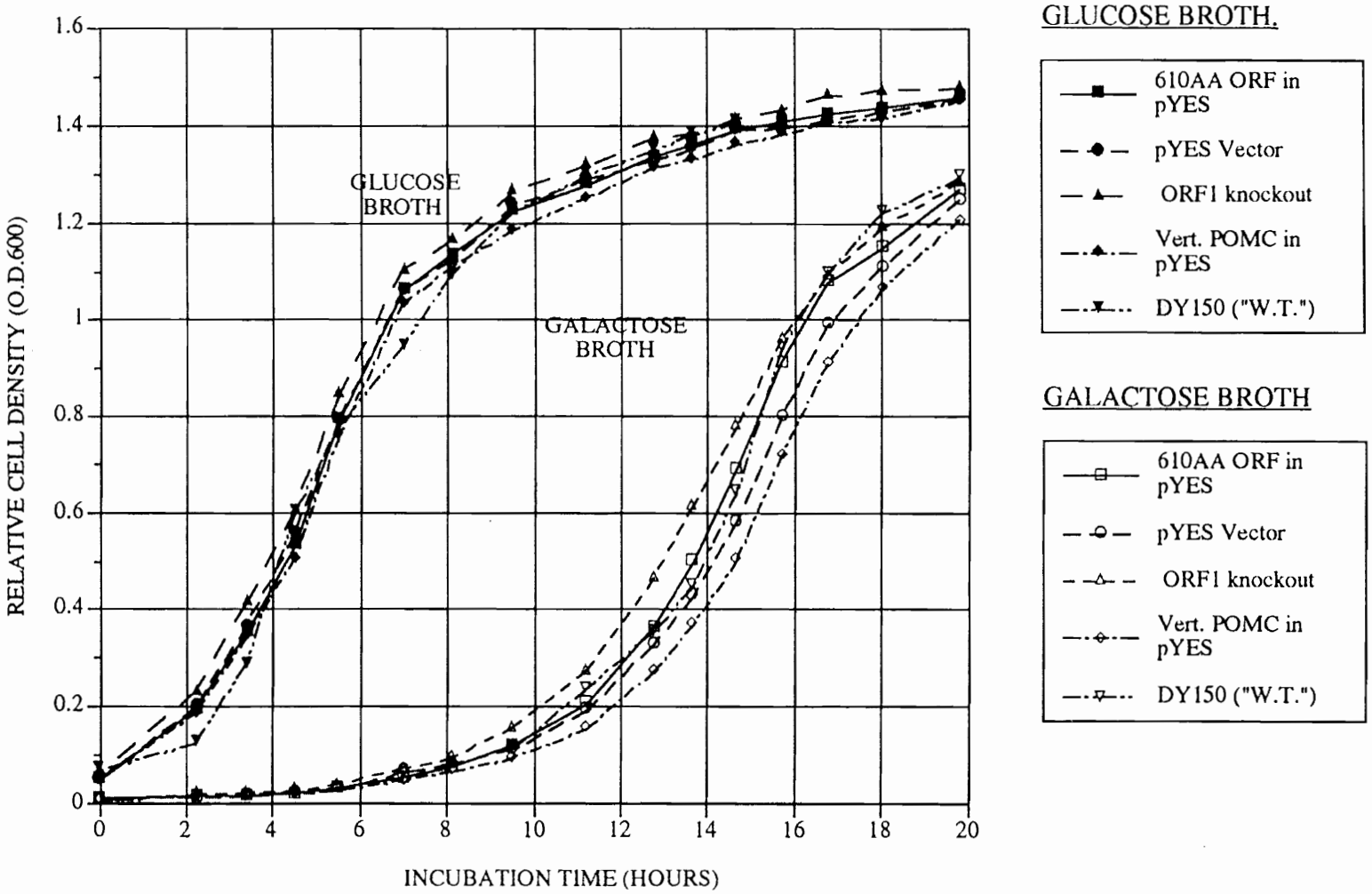


Fig. 30. Comparison of growth rate of different yeast strains in defined medium. Relative cell density (y axis) is plotted against incubation time at 30C (x axis). Yeast were grown in either glucose or galactose as a carbon source. Different strains are as described in text.

CHAPTER 7

DISCUSSION AND SIGNIFICANCE

This search for a POMC-related gene failed to identify an unmistakable homolog in unicellular organisms. There are three possible reasons for this: a) it was present but not found b) whatever homolog is present is so diverged as to be unrecognizable using current techniques, and c) it was present but not in the form anticipated.

Although none of these possibilities can be excluded with certainty until the entire yeast genome is known, both the present work and knowledge gained from other workers suggest the latter two possibilities, and particularly the last, are much more likely. As already set out in the introduction, the sequence data in invertebrates are very limited and somewhat confusing from a teleological point of view. The finding of POMC-related sequences within an unrelated gene and the associated presence of CRF-like sequences (43) suggests that a distinct gene with solely POMC-like sequence and organization is much less likely than originally thought. There is evidence elsewhere to suggest that the genes for arginine vasopressin and the neurophysins may also be distantly related to POMC peptides (203), which clouds the issue even further, and makes a unique evolutionary homolog even less likely. Other very experienced workers have failed to identify such a gene in yeast despite considerable search (204, 205), and the small amount of immunoactive material present has also hindered others from taking an easier but more indirect route of purifying the protein (54).

The second conclusion (an unrecognizably diverged homolog) can also be drawn from the present data, although this is again impossible to prove unless the entire genome is sequenced or the discovery is fortuitous, as in the case of the FMRF sequences (43); the POMC-like peptides within the Aplysia FMRF gene showed low homology and would not have been detected by the techniques used here.

The third conclusion is the most satisfying, both in terms of the current findings, and from the point of view of synthesizing these findings with our current understanding of molecular evolution. As knowledge of invertebrate phylogeny and biology increases, several generalizations appear to hold true about so-called vertebrate peptides in primitive life forms, from which I feel the present results can be best put in perspective. The initial plethora of antibody evidence for vertebrate peptides in primitive invertebrates led to conclusions which are in retrospect somewhat simplistic, and a way of thinking referred to as "vertebratism" by one author (83). At a structural level, the immunological (antibody cross-reaction) data led to the mistaken conclusion that this necessarily indicated significant conservation of primary sequence. Although the amount of sequence data in invertebrates (and in unicellulars particularly) is small, the available data indicate that such a conclusion is invalid, and that the antibody recognition involves a complex surface shape composed from several potentially widely spaced parts of the molecule (79).

At a functional level, the inherent assumption has been that whatever roles were found for the evolutionary homolog to a vertebrate peptide represent a rather simplified version of more complex roles in "higher" species, without all the extra complexities that millions of years have added. Although examples of such functional similarity can indeed be found, numerous instances throughout the invertebrate kingdoms also indicate that invertebrate proteins can have very

different and sometimes unpredictable functions bearing no apparent relation to those in vertebrates. As examples of vertebrate peptides possessing other functions in invertebrates, vasopressin-related peptides can function as potent neurotoxins in molluscs (206), and sequences within vasoactive intestinal peptide (VIP; 207) and possibly gonadotropins (208) can act as reducing agents to protect against free radical damage. Of great evolutionary significance in all this is that the critical residues, and in turn the profile of selection pressures in different parts of such molecules, may therefore potentially differ widely from those in higher vertebrates, so that it becomes very difficult to predict which residues will be conserved.

Overall, the evidence appears to fit best with the concept of "protein modules" (209), wherein certain sequences can be mixed and matched to produce hybrid molecules with functions different from either component. This appears to be the best explanation to date of the presence of some sequences distantly related to POMC peptides in unexpected proteins. These include distant sequence homologies to α-MSH in several enzymes, including RNA polymerase II (210), α-amylase (211), and carbonic anhydrase (212), as well as part of the heavy chain of at least one of the immunoglobulin molecules (213). In other examples, it explains how certain sequences, such as conserved parts of the epidermal growth factor sequence and plasminogen (210), are present within molecules having very different overall structure and function. It is therefore not implausible that the large open reading frame found here (ORF1) may indeed contain sequences with a significant (although presently obscure) evolutionary relationship to vertebrate POMC molecules, though the overall structure of the protein predicted from ORF1 does not match POMC. It may well be that similar sequences were used to form this leucine zipper protein as were later shuffled to form several enzymes and the earliest "recognizable" distinct POMC precursor.

The finding that ablation of ORF1 was not lethal and that overexpression had no obvious effect on cell growth or morphology indicates that the gene is not essential for growth and differentiation in the media tested. A nonessential function argues against significant evolutionary conservation, at least under the circumstances in which the yeast was grown. However, it may well be essential for some function not presently selected (e.g., some function in intermediary metabolism), and which other specific experiments might reveal.

Whether this leucine zipper protein is indeed the protein most homologous to vertebrate POMC peptides remains uncertain. Given the additional several hundred million years of evolutionary distance between yeast and Aplysia (90, 91), and the lack of success of other investigators in identifying a unicellular POMC homolog by similar screening techniques to my own (187, 188), it seems highly unlikely that any gene with greater than the 20-30% protein homology to vertebrate POMC (as identified in parts of ORF1) will be found with existing technology (214). It may well be that the final story will only be revealed when the yeast genome is sequenced.

REFERENCES

1. **Chretien, M., Benjannet, S., Gossard, F., Gianoulakis, C., Crine, P., Lis, M., Seidah, N.G.** From β-lipotropin to β-endorphin and proopiomelanocortin. *Can. J. Biochem.* 57: 1111-1121, 1979.

2. **Smith, A.I. and Funder, J.W.** Proopiomelanocortin processing in the pituitary, central nervous system, and peripheral tissues. *Endocrine Rev.* 9: 159-179, 1988.

3. **Roberts, J.L., Phillips, M., Rosa, P.A. and Herbert, E.** Steps involved in the processing of common precursor forms of adrenocorticotropin and endorphin in cultures of mouse pituitary cells. *Biochemistry* 17:3609-3618, 1978.

4. **Eipper, B.A. and Mains, R.E.** Structure and biosynthesis of proadrenocorticotropin/β-endorphin and related peptides. *Endo. Rev.* 1:1-26, 1980.

5. **Fenger, M.** Proopiomelanocortin: α-amidated and related peptides. *Scand. J. Clin. Lab. Invest.* 50: 229-245, 1990.

6. **Smyth, D.G. and Zakarian, S.** Selective processing of β-endorphin in regions of porcine pituitary. *Nature* 288: 613-615, 1980.

7. **Imura, H.** ACTH and related peptides: molecular biology, biochemistry and regulation of secretion. *Clin. Endo. Metab.* 14: 845-866, 1985.

8. **Martens, G.J.M., Civelli, O. and Herbert, E.** Nucleotide sequence of cloned cDNA for pro-opiomelanocortin in the amphibian *Xenopus laevis*. *J. Biol. Chem.* 260: 13685-13689, 1985.

9. **Soma, G.-K., Kitahara, N., Nishizawa, T., Nanami, H., Kotake, C., Okazaki, H. and Andoh, T.** Nucleotide sequence of a cloned cDNA for proopiomelanocortin precursor of chum salmon *Onchorynchus keta*. *Nucl. Acid Res.* 12: 8029-8041, 1984.

10. **Wallis, M., Howell, S.L. and Taylor, K.W.** The Biochemistry of the polypeptide hormones. Chapter 5, pp92-145, Wiley, 1985.

11. **Takahashi, A., Kawauchi, H., Mouri, T. and Sasaki, A.** Chemical and immunological characterization of salmon endorphins. *Gen. Comp. Endo.* 53: 381-388, 1984.

12. **Bentley, P.J.** Comparative Vertebrate Endocrinology. 2nd Edition, pp115-121, Macmillan, 1982.

13. **Li, C.H., Oosthuizen, M.M.J. and Chung, D.** Isolation and primary structures of elephant adrenocorticotropin and β-lipotropin. *J. Pep. Prot. Res.* 32: 573-578, 1988.

14. **Kawauchi, H., Kawazoe, I., Adachi, Y., Buckley, D.I. and Ramachandran, J.** Chemical and biological characterization of salmon melanocyte-stimulating hormones. *Gen. Comp. Endo.* 53: 37-48, 1984.

15. **Doolittle, R.F., Feng, D., Johnson, M. and McClure, M.** Relationships of human protein sequences to those of other organisms. *Cold Spring Harb. Symp. Quant. Biol.* 51:447-455, 1986.

16. **Rodrigues, K.T. and Sumpter, J.P.** The radioimmunoassay of α-melanocyte-stimulating hormone and endorphin in trout (*Salmo gairdneri*) and the effect of blinding on the plasma levels of these peptides. *Gen. Comp. Endo.* 54: 69-75, 1984.

17. **Kitahara,N., Nishizawa, t., Iida, K., Okazaki, H., Andoh, T. and Soma, G.-I.** Absence of a gamma-melanocyte-stimulating hormone sequence in proopiomelanocortin mRNA of chum salmon *Oncorhynchus keta. Comp. Biochem. Physiol.* 91B: 365-370, 1988.

18. **Kawauchi, H.** Chemistry of proopiocortin-related peptides in the salmon pituitary. *Arch. Biochem. Biophys.* 227: 343-350, 1984.

19. **Gispen, W.H., Wiegant, V.M., Greven, H.M. and de Wied, D.** The induction of excessive grooming in the rat by intraventricular application of peptides derived from ACTH: structure-activity studies. *Life Sci.* 17: 645-652, 1975.

20. **Bickel, U., Born, J., Fehm, H.L., Distler, M. and Voigt, K.H.** The behaviorally active peptide ACTH4-10: measurement in plasma and pharmacokinetics in man. *Eur. J. Clin. Pharmacol.* 35: 371-377, 1988.

21. **Seidah, N.G. and Chretien, M.** Complete amino acid sequence of a human pituitary glycopeptide: an important maturation product of proopiomelanocortin. *Proc. Natl. Acad. Sci. U.S.A.* 78: 4236-4240, 1981.

22. **Pedersen, R.C. and Brownie, A.C.** Adrenocortical response to corticotropin is potentiated by part of the amino-terminal region of pro-corticotropin /endorphin. *Proc. Natl. Acad. Sci. U.S.A.* 77: 2239-2243, 1980.

23. **Lis, M., Hamet, P., Gutkowska, J., Maurice, G., Seidah, M.G., Lariviere, N., Chretien, M. and Genest, J.** Effect of N-terminal portion of proopiomelanocortin on aldosterone release by human adrenal adenoma in vitro. *J. Clin. Endocrinol. Metab.* 56: 489-495, 1981.

24. **Parker, L., Lifrak, E., Shively, J.** Human adrenal gland cortical androgen-stimulating hormone (CASH) is identical with a portion of the joining peptide of pituitary pro-opiomelanocortin (POMC) [Abstract 299] *Proc. 71st Annual Meet of the Endocrine Soc.*, 1989.

25. **Drouin, J., Nemer, M. Charron, J., Gagner, J-P., Jeannotte, L., Sun, Y.L., Therrien, M. and Tremblay, Y.** Tissue-specific activity of the pro-opiomelanocortin (POMC) gene and repression by glucocorticoids. *Genome* 31: 510-529, 1989.

26. **Kawauchi, H., Adachi, Y. and Tsubokawa, M.** Identification of the mRNA coding for the ACTH-β-lipotropin precursorin a human ectopic ACTH-producing tumor. *Biochem. Biophys. Res. Comm.* 98: 533-540, 1980.

27. **Bioleau, G., Barbeau, C., Jeannotte, L., Chretien, M. and Drouin, J.** Complete structure of the porcine pro-opiomelanocortin mRNA derived from the nucleotide sequence of cloned cDNA. *Nucl. Acid Res.* 11: 8063-8071, 1983.

28. **Oates, E. and Herbert, E.** 5' sequence of porcine and rat proopiomelanocortin mRNA. *J. Biol. Chem.* 259: 7421-7425, 1984.

29. **Nakanishi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A.C.Y., Cohen, S.M. and Numa, S.** Nucleotide sequence of cloned cDNA for bovine corticotropin-beta-lipotropin precursor. *Nature* 278: 423-427, 1979.

30. **Drouin, J., Burbach, P., Charron, J. and Gagner, J.-P.** Expression of the pro-opiomelanocortin (POMC) gene. *J. Cell. Biochem.* Suppl. 7A: 107, 1983.

31. **Crine, P., Lemieux, E., Fortin, S., Seidah, N.G., Lis, M. and Chretien, M.** Expression of variant forms of proopiomelanocortin, the common precursor to corticotropin and β-lipotropin in the rat pars intermedia. *Biochemistry* 20: 2475-2481, 1981.

32. **Drouin, J. and Goodman, H.M.** Most of the coding region of rat ACTH/β-LPH precursor gene lacks intervening sequences. *Nature* 288: 610-613, 1980.

33. **Takahashi, H., Hakamata, Y., Watanabe, Y., Kikuno, R., Miyata, T. and Numa, S.** Complete nucleotide sequence of the human corticotropin-β-lipotropin precursor gene. *Nucl. Acid Res.* 11: 6847-6858, 1983.

34. **Cochet, M., Chang, A.C.Y. and Cohen, S.N.** Characterization of the structural gene and putative 5'-regulatory sequences for human proopiomelanocortin. *Nature* 297: 334-339, 1982.

35. **Takahashi, H., Teranishi, Y., Nakanishi, S. and Numa, S.** Isolation and structural organization of the human corticotropin-beta-lipotropin precursor gene. *FEBS Lett.* 135: 97-102, 1981.

36. **Martens, G.J.M., Biermans, P-P.J., Jenks, B.G. and van Overbeeke, A.P.** Biosynthesis of two structurally different proopiomelanocortins in the pars intermedia of the amphibian pituitary gland. *Eur. J. Biochem.* 126: 17-22, 1982.

37. **Uhler, M. and Herbert, E.** Complete amino acid sequence of mouse pro-opiomelanocortin derived from the nucleotide sequence of proopiomelanocortin cDNA. *J. Biol. Chem.* 258: 257-261, 1983.

38. **Notake, M. Tobimatsu, T., Watanabe, Y., Takahashi, H., Mishina, M., and Numa, S.** Isolation and characterization of the mouse corticotropin-β-lipotropin precursor gene and a related pseudogene. *FEBS Lett.* 156: 647-71, 1983.

39. **Uhler, M., Herbert, E., D'Eustachio, P.D. and Ruddle, F.D.** The mouse genome contains two nonallelic pro-opiomelanocortin genes. *J. Biol. Chem.* 258: 9444-9453, 1983.

40. **Kawauchi, H., Takahashi, A. and Abe, K-I.** Occurrence of a new corticotropin-like intermediate lobe peptide in salmon pituitary glands. *Arch. Biochem. Biophys.* 213: 680-688, 1982.

41. **Nishizawa, T., Kitahara, N., Nanami, H., Hara, N., Kotake, C., Okazaki, H., Andoh, T. and Soma, G-I.** Heterogeneity of 3' -nontranslated regions in proopiomelanocortin (POMC) precursor mRNA of chum salmon *Oncho-rynchus keta*: polymorphism of the gene. *Biochem. Biophys. Res. Comm.* 122: 556-562, 1984.

42. **Bisbee, C.A., Baker, M.A., Wilson, A.C., Hadji-azimi, I. and Fischberg, M.** Albumin phylogeny of clawed frogs. *Science* 195: 785-787, 1977.

43. **Taussig, R. and Scheller, R.H.** The Aplysia FMRFamide gene encodes sequences related to mammalian brain peptides. *DNA* 5: 454-461, 1986.

44. **Roth, J., LeRoith, D., Shiloach, J. Rosenweig, H.L., Lesniak, M.A. and Havrankova, J.** The evolutionary origins of hormones, neurotransmitters, and other extracellular chemical messengers. *New Engl. J. Med.* 306: 523-527, 1982.

45. **Dohler, K.D.** Development of hormone receptors: conclusion. *Experientia* 42: 788-794, 1986.

46. **LeRoith, D., Shiloach, J., Roth, J., and Lesniak, M.A.** Insulin or a closely related molecule is ntive to *Escherichia coli. J. Biol. Chem.* 257: 4925-4930, 1981.

47. **LeRoith, D., Shiloach, J., Roth, J., and Lesniak, M.A.** Evolutionary origins of vertebrate hormones: Substances similar to mammalian insulins are native to unicellular organisms. *Proc. Natl. Acad. Sci. U.S.A.* 77: 6184-6188, 1980.

48. **LeRoith, D., Lesniak, M.A. and Roth, J.** Insulin in insects and annelids. *Diabetes* 30: 70-76, 1981.

49. **Acevedo, M.F., Slifkin, M., Pouchet, G.R., and Pardo, M.** Immunocytochemical localization of a choriogonadotropin like protein in bacteria isolated from cancer patients. *Cancer* 41: 1217-1219, 1978.

50. **Maruo, T., Cohen, H., Segal, S.J. and Koide, S.S.** Production of choriogonadotropin-like factor by a microorganism. *Proc. Natl. Acad. Sci. U.S.A.* 76: 6622-6626, 1979.

51. **Bramley, T.A., Menzies, G.S., Willeams, R.J., Adams, D.J. and Kinsman, O.S.** Specific, high-affinity binding sites for human luteinizing hormone (hLH) and human chorionic gonadotrophin (hCG) in *Candida* species. *Biochem. Biophys. Acta.* 167: 1050-1056, 1990.

52. **Richert, N.D. and Ryan, R.J.** Specific gonadotropin binding to *Pseudomonas maltophilia. Proc. Natl. Acad. Sci. U.S.A.* 73: 878-882, 1977.

53. **Berelowitz, M., LeRoith, D., Von Schenk, H., Newgard, C., Szabo, M., Frohman, L.A., Shiloach, J., and Roth, J.** Somatostatin-like immunoactivity and biological activity is present in *T. Pyriformis*, a ciliated Protozoan. *Endocrinology* 110: 1939-1944, 1982.

54. **LeRoith, D., Pickens, W., Crosby, L.K., Berelowitz, M., Holtgrefe, M. and Shiloach, J.** Evidence for multiple molecular weight forms of somatostatin-like material in *Escherichia coli. Biochem. Biophys. Acta* 838: 335-342, 1985.

55. **Schwabe, C., LeRoith, D., Thompson, R.P., Shiloach, J. and Roth, J.** Relaxin extracted from protozoa (*Tetrahymena pyriformis*). *J. Biol. Chem.* 258: 2778-2782, 1983.

56. **Perez-Cano, R., Maruphy, P.K., Girgis, S.I., Arnett, T.R., Blankharn, I., and MacIntyre, I.** Unicellular organisms contain a molecule resembling human calcitonin. *Endocrinology* 110: 673 (Abstract), 1982.

57. **Deftos, L., LeRoith, D., Shiloach, J. and Roth, J.** Salmon calcitonin-like immunoactivity in extracts of *Tetrahymena pyriformis. Horm. Metab. Res.*17: 82-85, 1985.

58. **Macchia, V., Bates, R.W. and Pastan, I.** Purification and properties of thyroid stimulating factor isolated from *Clostridium perfringens*. *J. Biol. Chem.* 242: 3726-3730, 1967.

59. **Weiss, M., Ingbar, S.H., Winblad, S., and Kasper, D.L.** Demonstration of a saturable binding site for thyrotropin in *Yersinia enterocolitica*. *Science* 219: 1331-1333, 1983.

60. **Bhatnagar, Y.M., and Carraway, R.** Bacterial peptides with C-terminal similarities to bovine neurotensin. *Peptides* 2: 51-59, 1981.

61. **LeRoith, D., Liotta, A.S., Roth, J., Shiloach, J., Lewis, M.E., Pert, C.B. and Krieger, D.T.** Corticotropin and β-endorphin-like materials are native to unicellular organisms. *Proc. Natl. Acad. Sci. U.S.A.* 79: 2086-2090, 1982.

62. **LeRoith, D., Shiloach, J., Roth, J., Liotta, A.S., Krieger, D.T., Lewis, M. and Pert, C.B.** Evolutionary origins of vertebrate hormones: material very similar to adrenocorticotropic hormone, β-endorphin, and dynorphin in protozoa. *Trans. Assoc. Am. Physicians.* 94: 52-60, 1981.

63. **Zipser, B., Ruff, M.R., O'Neill, J.B., Smith, C.C., Higgins, W.J. and Pert, C.B.** The opiate receptor: a single 110kDa recognition molecule appears to be conserved in *Tetrahymena*, leech, and rat. *Brain Res.* 463: 296-304, 1988.

64. **Josefsson, J.O. and Johansson, P.** Naloxone-reversible effects of opioids on pinocytosis in *Amoeba proteus*. *Nature* 282: 78-80, 1979.

65. **LeRoith, D., Roberts Jr., C., Lesniak, M.A. and Roth, J.** Receptors for intercellular messenger molecules in microbes: similarities to vertebrate receptors and possible implications for diseases in man. *Experientia* 42: 782-788, 1986.

66. **Csaba, G., Nagy, S.U.** Presence (HPL, prostaglandin) and absence (triiodothyronine, thyroxine) of hormones in *Tetrahymena*: experimental facts and open questions. *Acta Physiol. Hung.* 70: 105-110, 1987.

67. **Feldman, K.D., Tokes, L.G., Stathis, P.A., Miller, S.C., Kurz, W. and Harvey, D.** Identification of 17β-estradiol as the estrogenic substance in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 81: 4722-4726, 1984.

68. **Burshell, A., Stathis, P.A., Do, Y., Miller, S.C. and Feldman, D.** Characterization of an estrogen-binding protein in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* 259: 3450-3456, 1984.

69. **Loose, D.S., Stover, E.P., Restrepo, A., Stevens, D.A. and Feldman, D.** Estradiol binds to a receptor-like cytosol binding protein and initiates a biological response in *Paracoccidioides brasiliensis*. *Proc. Natl. Acad. Sci. U.S.A.* 80:7659-7663, 1983.

70. **Loose, D.S. and Feldman, D.** Characterization of a unique corticosterone-binding protein in *Candida albicans*. *J. Biol. Chem.* 257: 4925-4930, 1981.

71. **Verhaert, P. and de Loof, A.** Immunocytochemical localization of a methionine-enkephalin-resembling neuropeptide in the central nervous system of the American cockroach, *Periplaneta americana*. *J. Comp. Neurol.* 239: 54-61, 1985.

72. **Stefano, G.B. and Scharrer, B.** High affinity binding of an enkephalin analog in the cerbral ganglion of the insect *Leucophaea maderae* (Blattaria). *Brain Res.* 225: 107-114, 1981.

73. **Schoofs, L., Jegou, S., Tonon, M-C., Huybrechts, R., Vaudry, H. and De Loof, A.** Biological, chromatographical, and radioimmunological evidence for a melanotropin-like peptide in the central nervous system of *Locusta migratoria*. *Gen. Comp. Endocrinol.* 71: 36-44, 1988.

74. **Schoofs, L., Jegou, S., Vaudry, H., Verhaert, P. and De Loof, A.** Proopiomelanocortin- and enkephalin-related substances in the central nervous system of *Locusta migratoria* and *Sarcophaga bullata*. *Gen. Comp. Endo.* 66: 46-47. (Abst#132), 1987

75. **Alumets, J., Hakanson, R., Sundler, F. and Thorell, J.** Neuronal localization of immunoreactive enkephalin and β-endorphin in the earthworm. *Nature* 279: 805-806, 1979.

76. **Remy, C.H. and Dubois, M.P.** Alpha-endorphin-like cells in the infra-oesophageal ganglions of the earthworm *Dendrobaena subrubicunda* : Immunocytological localization. *Experientia* 35: 137-138, 1979.

77. **Duvaux-Miret, O., Dissous, C., Gautron, J.P., Pattou, E., Kordon, C. and Capron, A.** The helminth *Schistosoma mansoni* expresses a peptide similar to human β-endorphin and possesses a proopiomelanocortin-related peptide. *New Biol.* 2: 93-99, 1990.

78. **Odell, W.D. and Franchimont, P.** (Eds.) Principles of Competitive Binding Assays. 2nd Edition, Wiley, 1983.

79. **Laver, W. G., Air, G.M., Webster, R.G. and Smith-Gill, S.J.** Epitopes on protein antigens: misconceptions and realities. *Cell* 61: 553-556, 1990.

80. **De Loof, A. and Schoofs, L.** Homologies between the amino acid sequences of some vertebrate peptide hormones and peptides isolated from invertebrate sources. *Comp. Biochem. Physiol.* 25B: 459-468, 1990.

81. **Rafaeli, A., Moshitzky, P. and Applebaum, S.W.** Diuretic action and immunological cross-reactivity of corticotropin and locust diuretic hormone. *Gen. Comp. Endocrinol.* 67: 1-6, 1987.

82. **Proux, J.F., Miller, C.A., Li, J.P., Carney, R.L., Girardie, A., Delaage, M. and Schooley, D.A.** Identification of an arginine vasopressin-like diuretic hormone from *locusta migratoria. Biochem. Biophys. Res. Comm.* 149: 180-186, 1987.

83. **Thorndyke, M.C.** Peptides in Invertebrates, in Holmgren, S. (Ed) The Comparative Physiology of Regulatory Peptides. Chapter 9, pp. 203-228, Chapman and Hall 1989.

84. **Jung, L.J. and Scheller, R.H.** Peptide processing and targeting in the neuronal secretory pathway. *Science* 251: 1330-1335, 1991.

85. **Gubler, U., Seeburg, P.H., Hoffman, B.J., Gage, L.P. and Udenfriend, S.** Molecular cloning establishes proenkephalin as precursor of enkephalin-containing peptides. *Nature* 295: 206-208, 1982.

86. **Herbert, E., Oates, E., Martens, G., Comb, M., Rosen, H. and Uhler, M.** Generation of diversity and evolution of opioid peptides. *Cold Spring Harb. Symp. Quant. Biol.* 47: 375-383, 1983.

87. **Schneider, L.E. and Taghert, P.H.** Isolation and characterization of a Drosophila gene that encodes multiple neuropeptides related to Phe-Met-Arg-Phe-NH_2 (FMRFamide). *Proc. Natl. Acad. Sci. U.S.A.* 85: 1993-1997, 1988.

88. **Leung, M.K. and Stefano, G.B.** Isolation and identification of enkephalins in pedal ganglia of *Mytilus edulis* (Mollusca). *Proc. Natl. Acad. Sci. U.S.A.* 81: 955-958, 1984.

89. **Creighton, T.E.** Proteins. Structures and Molecular Properties. Ch 3. pp 93-131, Freeman, 1983.

90. **Hanson, E.D.** Phylogenetic Conclusions: A Review. In: Animal Diversity. 3rd Edition. Ch 10, pp 141-148. Prentice Hall, 1972.

91. **Miller, M.** (Professor of Zoology, University of Auckland). Personal communication, 1992.

92. **Loomis, W.F.** Similarities in eukaryotic genomes. *Comp. Biochem. Physiol.* 95B: 21-27, 1990.

93. **Herbert, E. and Uhler, M.** Biosynthesis of polyprotein precursors to regulatory peptides. *Cell* 30: 1-2, 1982.

94. **Kurjan, J. and Herskowitz,I.A.** Structure of a yeast pheromone gene (MFα): a putative α-factor precursor contains four tandem copies of mature α-factor. *Cell* 30: 933-943, 1982.

95. **Straus, E., Yalow, R.S. and Gainer, H.** Molluscan gastrin: concentration and molecular forms. *Science* 190: 687-689, 1975.

96. **Barrington, E.J.W.** The phylogeny of the endocrine system. *Experientia* 42: 775-782, 1986.

97. **Sporn, M.B. and Roberts, A.B.** Peptide growth factors are multifunctional. *Nature* 332: 217-219, 1988.

98. **Tsou, K., Khachaturian, H., Akil, H. and Watson, S.J.** Immunocytochemical localization of pro-opiomelanocortin-derived peptides in the adult rat spinal cord. *Brain Res.* 378: 28-35, 1986.

99. **Watson, S.J., Akil, H., Richard, C.W. and Barchas, J.D.** Evidence for two separate opiate peptide neuronal systems. *Nature* 275: 226-228, 1978.

100. **Van Houten, M., Khan, M.N., Walsh, R.J., Baquiran, G.B., Renaud, L.P., Bourque, C,H., Sgro, S., Gauthier, S., Chretien, M. and Posner, B.** NH_2-terminal specificity and axonal localization of adrenocorticotropin binding sites in rat median eminence. *Proc. Natl. Acad. Sci. U.S.A.* 82: 1271-1275, 1985.

101. **Alvarez-Monn, M., Kehrl, J.H., Fauci, A.S.** A potential role for adrenocorticotropin in regulating human B lymphocyte functions. *J. Immunol.* 135: 3823-3826, 1985.

102. **Weigent, D.A. and Blalock, J.E.** Interactions between the neuroendocrine and immune systems: common hormones and receptors. *Immunol. Rev.* 55: 47-63, 1987.

103. **Ottaviani, E., Petraglia, F., Montagnani, G., Cossarizza, A., Monti, D. and Franceschi, C.** Presence of ACTH and β-endorphin immunoreactive molecules in the freshwater snail *Planorbarius corneus* (L.)(Gastropoda, Pulmonata) and their possible role in phagocytosis. *Regul. Peptides* 27: 1-9, 1990.

104. **Gizang-ginsberg, E. and Wolgemuth, D.J.** Expression of the proopiomelanocortin gene is developmentally regulated and affected by germ cells in the male mouse reproductive system. *Proc. Natl. Acad. Sci. U.S.A.* 84: 1600-1604, 1987.

105. **Keshet, E., Polakiewicz, R.D., Itin, A., Ornoy, A. and Rosen, H.** Proenkephalin A is expressed in mesodermal lineages during organogenesis. *EMBO J.* 8: 2917-2923, 1989.

106. **Cossu, G., Cusella-De Angelis, M.G., Senni, M.I., De Angelis, L,Vivarelli, E., Vella, S., Bouche, M., Boitani, C. and Molinaro, M.** Adrenocorticotropin is a specific mitogen for mammalian myogenic cells. *Dev. Biol.* 131: 331-336, 1989.

107. **Estivariz, F.E., Iturizza, F., McLean, L., Hope, J. and Lowry, P.J.** Stimulation of adrenal mitogenesis by N-terminal proopiocortin peptides. *Nature* 297: 419-422, 1982.

108. **Cheitlin, R. and Ramachandran, J.** Regulation of actin in rat adrenocortical cells by corticotropin. *J. Biol. Chem.* 256: 3156-3158, 1981.

109. **Saito, E., Iwasa, S. and Odell, W.D.** Widespread presence of large molecular weight adrenocorticotropin-like substances in normal rat extrapituitary tissues. *Endorinology* 113: 1010-1019, 1983.

110. **DeBold, C.R., Nicholson, W.E. and Orth, D.N.** Immunoreactive proopiomelanocortin peptides and POMC-like messenger ribonucleic acid are present in many rat nonpituitary tissues. *Endocrinology* 66: 2648-2657, 1988.

111. **Botstein D.** Personal communication, 1990.

112. **Julius, D., Brake, A., Blair, L. Kunisawa, R. and Thorner, J.** Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast prepro-α factor. *Cell* 37: 1075-1089, 1984.

113. **Thomas, L., Cooper, A., Bussey, H. and Thomas, G.** Yeast KEX1 protease cleaves a prohormone processing intermediate in mammalian cells. *J. Biol. Chem.* 265: 10821-10824, 1990.

114. **Thomas, G., Thorne, B.A., Thomas, L., Allen, R.G., Hruby, D.E., Fuller, R., and Thorner, J.** Yeast KEX2 endopeptidase correctly cleaves a neuroendocrine prohormone in mammalian cells. *Science* 241: 226-230, 1988.

115. **Peterson, J.B., Nelson, D.L., Ling, E. and Angeletti, R.H.** Chromogranin A-like proteins in the secretory granules of a protozoan, *Paramecium tetraurelia*. *J. Biol. Chem.* 262: 17264-17267, 1987.

116. **Kochert, G.** Sex pheromones in algae and fungi. *Ann. Rev. Plant Physiol.* 29: 461-486, 1978.

117. **LeRoith, D., Roberts Jr., C., Lesniak, M.A. and Roth, J.** Receptors for intercellular messenger molecules in microbes: similarities to vertebrate receptors and possible implications for diseases in man. *Experientia* 42: 782-788, 1986.

118. **Muggeo, M., Ginsberg, B.H., Roth, J., Neville, D.M., De Meyts, P. and Kohn, C.R.** The insulin receptor invertebrates is functionally more conserved during evolution than insulin itself. *Endocrinology* 104: 1393-1402, 1979.

119. **Berridge, M.J.** Inositol trisphosphate and diacylglycerol as second messengers. *Biochem. J.* 220: 345-360, 1984.

120. **Boulton, T.G., Yancopoulos, G.D., Gregory, J.S., Slaughter, C., Moomaw, C., Hsu, J. and Cobb, M.H.** An insulin-stimulated protein kinase similar to yeast kinases involved in cell cycle control. *Science* 249: 64-67, 1990.

121. **Metzger, D., White, J.H. and Chambon, P.** The human oestrogen receptor functions in yeast. *Nature* 334: 31-36, 1988.

122. **Gabrielsen, O.S., Reppe, S., Saether, O., Blingsmo, O.R., Sletten, K., Gordeladze, J.O., Hogset, A., Gautvik, V.T., Alestrom, P., Oyen, T.B. and Gautvik, K.M.** Efficient secretion of human parathyroid hormone by *Saccharomyces cerevisiae*. *Gene* 90: 255-262, 1990.

123. **Gellerfors, P., Axelsson, K., Helander, A., Johansson, S., Kenne, L., Lindqvist, S., Pavlu, B., Skottner, A. and Fryklund, L.** Isolation and characterization of a glycosylated form of human insulin-like growth factor I produced in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 264: 11444-11449, 1989.

124. **Privalsky, M.L., Sharif, M. and Yamamoto, K.R.** The viral *erbA* oncogene protein, a constitutive repressor in animal cells, is a hormone-regulated activator in yeast. *Cell* 63: 1277-1286, 1990.

125. **Sambrook, J., Fritsch, E.F. and Maniatis, T.** A Manual of Molecular Cloning Techniques. Cold Spring Harbor Laboratory, 1990.

126. **Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K.**(Eds) Current Protocols in Molecular Biology. Wiley Interscience 1987.

127. **Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J.** (Eds) PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego, CA, 1990.

128. **Sanford, Y.M. and Orias, E.** Phenylketonuric *Tetrahymena*: phenylalanine hydroxylase mutants and other tyrosine auxotrophs. *Proc. Natl. Acad. Sci. U.S.A.* 78: 7614-7618, 1981.

129. **Lue, N.F. and Kornberg, R.D.** Accurate initiation at RNA polymerase II promoters in extracts from *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 84: 8839-8843, 1987.

130. **Philippsen, P., Stotz, A. and Scherf, C.** DNA of *Saccharomyces cerevisiae*. *Meth. Enzymol.* 194: 169-182, 1991.

131. **Berson, S.A. and Yalow, R.S.** Radioimmunoassay of ACTH in plasma. *J. Clin. Invest.* 47: 2725-2732, 1968.

132. **Salacinski, P.R.P., McLean, C., Sykes, J.E.C., Clement-Jones, V.V. and Lowry, P.J.** Isolation of proteins, glycoproteins and peptides using a solid-phase oxidizing agent 1,3,4,6-tetrachloro-3α,6α-dephenyl glycoluril (iodogen). *Anal. Biochem.* 117: 136-146, 1981.

133. **Rae, P.A. and Schimmer, B.P.** Iodinated derivatives of adrenocorticotropic hormone. Preparation of iodinated α^{1-24}-ACTH with full hormonal activity. *J. Biol. Chem.* 249: 5649-5653, 1974.

134. **Hofmann, K., Romovacek, H., Stehle, C.J. and Finn, F.M.** Radioactive probes for adrenocorticotropic hormone receptors. *Biochemistry* 25: 1339-1346, 1986.

135. **Zahradnik, R., Brennan, G., Hutchison, J.S. and Odell, W.D.** Immunoradiometric assay of corticotropin with use of avidin-biotin separation. *Clin. Chem.* 35: 804-807, 1989.

136. **Ehrlich, P.H. and Moyle, W.R.** Specificity considerations in cooperative immunoassays. *Clin. Chem.* 30: 1523-1532, 1984.

137. **Moyle, W.R., Anderson, D.M. and Ehrlich, P.H.** A circular antibody-antigen complex is responsible for increased affinity shown by mixtures of monoclonal antibodies to human chorionic gonadotropin. *J. Immunol.* 131:1900-1905, 1983.

138. **Hirono, M., Endoh, H., Okada, N., Numata, O. and Watanabe,Y.** *Tetrahymena* actin: Cloning and sequencing of the *Tetrahymena* actin gene and identification of its gene product. *J. Mol. Biol.* 194: 181-192, 1987.

139. **Meinkoth, J. and Wahl. G.** Review: hybridization of nucleic acids immobilized on solid supports. *Anal. Biochem.* 138: 267-284, 1984.

140. **Roberts, J.L., Seeburg, P.H., Shine, J., Herbert, E., Baxter, J.D. and Goodman, H.** Corticotropin and β-endorphin: construction and analysis of recombinant DNA complementary to mRNA for the common precursor. *Proc. Natl. Acad. Sci. U.S.A.* 76: 2153-2157, 1979.

141. **Reed, K.C. and Mann, D.A.** Rapid transfer of DNA from agarose gels to nylon membranes. *Nucl. Acid Res.* 13: 7207-7221, 1985.

142. **Twomey, T.A. and Krawetz, S.A.** Parameters affecting hybridization of nucleic acids blotted onto nylon or nitrocellulose membranes. *Biotechniques* 8: 478-481, 1990.

143. **Nierzwicki-Bauer, S.A., Gebhardt, J.S., Linkkila, L. and Walsh, K.** A comparison of UV cross-linking and vacuum baking for nucleic acid immobilization and retention. *Biotechniques* 9: 472-478, 1990.

144. **Bennetzen, J.L. and Hall, B.D.** Codon selection in yeast. *J. Biol. Chem.* 257: 3026-3031, 1982.

145. **Davis, T.N. and Thorner, J.** Isolation of the yeast calmodulin gene using synthetic oligonucleotide probes. *Meth. Enzymol.* 139: 248-262, 1987.

146. **Martin, F.H., Castro, M.M., Aboul-ela, F. and Tinoco, I. Jr.** Base pairing involving deoxyinosine: implications for probe design. *Nucl. Acid Res.* 13: 8927-8938, 1985.

147. **Caruthers, M.H.** Gene synthesis machines: DNA chemistry and its uses. *Science* 230: 281-285, 1985.

148. **Schowalter, D.B. and Sommer, S.S.** The generation of radiolabelled DNA and RNA probes with polymerase chain reaction. *Analyt. Biochem.* 177: 90 94, 1989.

149. **Ohara, O., Dorit, R.L. and Gilbert, W.** One-sided polymerase chain reaction: the amplification of cDNA. *Proc. Natl. Acad. Sci. U.S.A.* 86: 5673-5677, 1989.

150. **Casey, J. and Davidson, N.** Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nucl. Acid Res.* 4: 1539-1552, 1977.

151. **McConaughy, B.L., Laird, C.D. and McCarthy, B.J.** Nucleic acid reassociation in formamide. *Biochemistry* 8: 3289-3294, 1969.

152. **Cigan, A.M. and Donahue, T.F.** Sequence and structural features associated with translational initiator regions in yeast : a review. *Gene* 59: 1-18, 1987.

153. **Kozak, M.** Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucl. Acid Res.* 9: 5233-5252, 1981.

154. **Kozak, M.** Translation of insulin-related polypeptides from messenger RNAs with tandemly reiterated copies of the ribosome binding site. *Cell* 34: 971-978, 1983.

155. **Guarente, L.** Regulatory proteins in yeast. *Ann. Rev. Genet.* 21: 425-452, 1987.

156. **Struhl, K.** Promoters, activator proteins, and the mechanism of transcriptional activation in yeast. *Cell* 49: 295-297, 1987.

157. **Breathnach, R. and Chambon, P.** Organization and expression of eucaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50: 349-83, 1981.

158. **Struhl, K.** Molecular mechanisms of transcriptional regulation in yeast. *Ann. Rev. Biochem.* 58: 1051-1077, 1989.

159. **Guarente, L.** Yeast promoters: positive and negative elements. *Cell* 36: 799-800, 1984.

160. **Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A.** Splicing of messenger RNA precursors. *Ann. Rev. Biochem.* 55: 1119-50, 1986.

161. **Fink, G.R.** Pseudogenes in yeast. *Cell* 49: 5-6, 1987.

162. **Baltimore, D.** Retroviruses and retrotransposons: the role of reverse transcription in shaping the eukaryotic genome. *Cell* 40: 481-482, 1985.

163. **Ptashne, M.** How eukaryotic transcriptional activators work. *Nature* 335: 683-689, 1988.

164. **Sibley, D.R., Benovic, J.L., Caron, M.G. and Lefkowitz, R.J.** Phosphorylation of cell surface receptors: a mechanism for regulating signal transduction pathways. *Endocrine Rev.* 9: 38-56, 1988.

165. **Vinson, C.R., Sigler, P.B. and McKnight, S.L.** Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* 246: 911-916, 1989.

166. **White, M.J., Hirsch, J.P. and Henry, S.A.** The *OPI1* gene of *Saccharomyces cerevisiae*, a negative regulator of phospholipid biosynthesis, encodes a protein containing polyglutamine tracts and a leucine zipper. *J. Biol. Chem.* 266: 863-872, 1991.

167. **Neuberg, M., Schuermann, M., Hunter, J.B. and Muller, R.** Two functionally different regions in fos are required for the sequence-specific DNA interaction of the fos/jun protein complex. *Nature* 338: 589-90, 1989.

168. **Rogers, S., Wells, R. and Rechsteiner, M.** Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* 234: 364-368, 1986.

169. **Kahana, C. and Nathans, D.** Nucleotide sequence of murine ornithine decarboxylase mRNA. *Proc. Natl. Acad. Sci. U.S.A.* 82: 1673-1677, 1985.

170. **DeMartino, G.N. and Croall, D.E.** Calcium-dependent proteases from liver and heart. *Prog. Clin. Biol. Res.* 180: 117-126, 1985.

171. **Doolittle, R.F.** Similar amino acid sequences: chance or common ancestry? *Science* 214: 149-159, 1981.

172. **Klarskov, K., Hojrup, P., Andersen, S.O. and Roepstorff, P.** Plasma-desorption mass spectrometry as an aid in protein sequence determination. Application of the method on a cuticular protein from the migratory locust (*Locusta migratoria*). *Biochem. J.* 262: 923-930, 1989.

173. **Larocca, D., Chao, L.A., Seto, M.H. and Brunck, T.K.** Human T-cell leukemia virus minus strand transcription in infected T-cells. *Biochem. Biophys. Res. Comm.* 163:1006-1013, 1989.

174. **Bodner, M., Castrilo, J.L., Theill, L.E., Deerinck, T., Ellisman, M. and Karin, M.** The pituitary-specific transcription factor GHF-1 is a homeobox-containing protein. *Cell* 55: 505-518, 1988.

175. **Ingraham, H.A., Chen, R., Mangalam, H.J., Elsholtz, H.P., Flynn, S.E., Lin, C.R., Simmons, D.M., Swanson, L. and Rosenfeld, M.G.** A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell* 55: 519-529, 1988.

176. **Bancroft, I. and Wolk, C.P.** Characterization of an insertion sequence (IS891) of novel structure from the cyanobacterium *Anabaena* sp. strain M-131. *J. Bacteriol.* 171: 5949-5954, 1989.

177. **Michaelis, U., Schlapp, T. and Roedel, G.** Yeast nuclear gene CBS2 required for translational activation of cytochrome b, encodes a basic protein of 45kDa. *Mol. Gen. Genet.* 214: 263-270, 1988.

178. **McPhaul, M.J., Noble, J.F., Simpson, E.R., Mendelson, C.R. and Wilson, J.D.** The expression of a functional cDNA encoding the chicken cytochrome P-450-arom (aromatase) that catalyzes the formation of estrogen from androgen. *J. Biol. Chem.* 263:16358-16363, 1988.

179. **Zinoni, F., Birkmann, A., Stadtman, T.C. and Bock, A.** Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.* 83: 4650-4654, 1986.

180. **Kohchi, T., Shirai, H., Fukuzawa, H., Sano, T., Komano, T., Umesono, K., Inokuchi, H., Ozeki, H. and Ohyama, K.** Structure and organization of *Marchantia polymorpha* chloroplast genome. IV. Inverted repeat and small single copy regions. *J. Mol. Biol.* 203: 353-372, 1988.

181. **Schiestl, R.H., Reynolds, P., Prakash, S. and Prakash, L.** Cloning and sequence analysis of the *Saccharomyces cerevisiae* RAD9 gene and further evidence that its procuct is required for cell cycle arrest induced by DNA damage. *Mol. Cell. Biol.* 9: 1882-1896, 1989.

182. **Smoller, D., Friedel, C., Schmid, A., Bettler, D., Lam, L. and Yedvobnick, B.** The Drosophila neurogenic locus mastermind encodes a nuclear protein unusually rich in amino acid homopolymers. *Genes Dev.* 4:1686-700, 1990.

183. **Semerjian, A.V., Malloy, D.C. and Poteete, A.R.** Genetic structure of the phage P22 PL operon. *J. Mol. Biol.* 207: 1-13, 1989.

184. **De Robertis, E.M., Oliver, G. and Wright, C.V.E.** Homeobox genes and the vertebrate body plan. *Scientific American* 263: 26-32, 1990.

185. **Buratowski, S., Hahn, S., Sharp, P.A. and Guarente, L.** Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* 334: 37-42, 1988.

186. **Ogden, J.E., Starway, C., Kim, S., Mellor, J., Kingsman, A.J. and Kingsman, S.M.** Efficient expression of the *Saccharomyces cerevisiae* PGK gene depends on an upstream activation sequence but does not require TATA sequences. *Mol. Cell. Biol.* 6: 4335-4343, 1986.

187. **Wierenga, R.K. and Hol, W.G.J.** Predicted nucleotide-binding properties of p21 protein and its cancer-associated variant. *Nature* 302: 842-848, 1983.

188. **Roush, C.L., Kennelly, P.J., Glaccum, M.B., Helfman, D.M., Scott, J.D. and Krebs, E.G.** Isolation of the cDNA encoding rat skeletal muscle myosin light chain kinase. *J. Biol. Chem.* 263: 10510-10516, 1988.

189. **Pellman, D., Garber, E.A., Cross, F.R., Hanafusa, H.** An N-terminal peptide from $p60^{src}$ can direct myristylation and plasma membrane localization when fused to heterologous proteins. *Nature* 314: 374-377, 1985.

190. **Graff, J.M., Gordon, J.I. and Blackshear, P.J.** Myristoylated and non-myristoylated forms of a protein are phosphorylated by protein kinase C. *Science* 246: 503-506, 1989.

191. **Hanks, S.K., Quinn, A.M. and Hunter, T.** The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. *Science* 241: 42-52, 1988.

192. **Hussain, K., Elliott, E.J. and Salmond, G.P.C.** The ParD(-) mutant of *Escherichia coli* carries a gyrAam mutation. The complete sequence of gyrA. *Mol. Microbiol.* 1: 259-273, 1987.

193. **Sharp, P.M., Cowe, E., Higgins, D.G., Shields, D.C., Wolfe, K.H. and Wright, F.** Codon usage patterns in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucl. Acid. Res.* 16: 8207-8211, 1988.

194. **Guoy, M. and Gautier, C.** Codon usage in bacteria: correlation with gene expressivity. *Nucl. Acid. Res.* 10: 7055-7073, 1982.

195. **Struhl, K.** Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 82: 8419-8423, 1985.

196. **Sachs, A.B. and Kornberg, R.D.** Nuclear polyadenylate-binding protein. *Mol. Cell. Biol.* 5: 1993-1996, 1985.

197. **Guerrini, A.M., Ascenzioni, F., Tribioli, C. and Donini, P.** Transformation of *Saccharomyces cerevisiae* and*Schizosaccharomyces pombe* with linear plasmids containing 2μ sequences. *EMBO J.* 4: 1569-1573, 1985.

198. **Rine, J.** Gene overexpression in studies of *Saccharomyces cerevisiae*. *Meth. Enzymol.* 194: 329-251, 1991.

199. **Schneider, J.C. and Guarente, L.** Vectors for expression of cloned genes in yeast: regulation, overproduction, and underproduction. *Meth. Enzymol.* 194: 373-388, 1991.

200. **Fowler, A. and Zabin, I.** The amino acid sequence of β-galactosidase in *E. coli*. *Proc. Natl. Acad. Sci. U.S.A.* 74: 1507-1510, 1977.

201. **Palzkill, T.G. and Newlon, C.S.** A yeast replication origin consists of multiple copies of a small conserved sequence. *Cell* 53: 441-450, 1988.

202. **Som, T., Armstrong, K.A., Volkert, F.C. and Broach, J.R.** Autoregulation of 2μm circle gene expression provides a model for maintenance of stable plasmid copy levels. *Cell* 52: 27-37, 1988.

203. **Furautani, Y., Morimoto, Y., Shibahara, S., Noda, M., Takahashi, H., Hirose, T., Asai, M., Inayama, S., Hayashida, H., Miyata, T. and Numa, S.** Cloning and sequence analysis of cDNA for ovine corticotropin-releasing factor precursor. *Nature* 301: 537-540, 1983.

204. **Scheller, R.H.** Personal communication.

205. **Schoofs, L.** Personal communication.

206. **Olivera, B.M., Rivier, J., Clark, C., Ramilo, C.A., Corpuz, G.P., Abogadie, F.C., Mena, E.E., Woodward, S.R., Hillyard, D.R. and Cruz, L.J.** Diversity of *Conus* neuropeptides. *Science* 249: 257-263, 1990.

207. **Misra, B.R. and Misra, H.P.** Vasoactive intestinal peptide, a singlet oxygen quencher. *J. Biol. Chem.* 265: 15371-15374, 1990.

208. **Boniface, J.J. and Reichert, L.E.** Evidence for a novel thioredoxin-like catalytic roperty of gonadotropic hormones. *Science* 245: 61- 64, 1990.

209. **Baron, M., Norman, D.G. and Campbell, I.D.** Protein modules. *TIBS* 16:41 47, 1991.

210. **Salvato, M., Shimomaye, E.M. and Oldstone, M.B.** The primary structure of the lymphocytic choriomeningitis virus L gene encodes a putative RNA polymerase. *Virology* 169: 377-384, 1989.

211. **Pasero, L., Massei-Pierron, Y., Abadie, B., Chicheportiche, Y. and Marchis-Mouren, G.** Complete amino acid sequence of the five disulfide bridges in porcine pancreatic alpha-amylase. *Biochem. Biophys. Acta* 869: 147-157, 1986.

212. **Komalis, C.P., Barlow, J.H., and Butterworth, P.H.** Cloned cDNA for rabbit erythrocyte carbonic anhydrase I: A novel erythrocyte-specific probe to study development in erythroid tissues. *Proc. Natl. Acad. Sci. U.S.A.* 82: 663-667, 1985

213. **Julliard, J.H., Shibasaki, T., Ling, N. and Guillemin, R.** High-molecular-weight immunoreactive β-endorphin in extracts of human placenta is a fragment of immunoglobulin G. *Science* 208: 183-185, 1980.

214. **LeRoith, D., Shiloach, J., Berelowitz, M., Frohman, L.A., Liotta, A.S., Krieger, D.T. and Roth, J.** Are messenger molecules in microbes the ancestors of the vertebrate hormones and tissue factors? *Federation Proc.* 42: 2602-2607, 1983.