Structure function relationships in Chinese hamster adenine phosphoribosyl transferase

by

Barry Noel Ford
B.Sc. University of Alberta, 1988

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biology

Abstract

Adenine phosphoribosyl transferase is a ubiquitous enzyme which salvages endogenous adenine, via the nucleotide AMP, for use by the cell. This activity, in conjunction with other interconnected purine salvage mechanisms is an energy-efficient way for the cell to satisfy its purine requirements. APRT is a target molecule in certain human diseases, for chemotherapeutics, and *in vivo* mutagenesis studies. There is little known about structure-function relationships in APRT. In the absence of solved three-dimensional crystal structures, we have explored structure-function relationships in APRT by sequence comparison, *in vitro* mutagenesis and kinetic analysis, protein crosslinking, and *in vivo* selection of mutant enzymes with altered substrate affinities. Chinese hamster APRT shares identifiable sequence similarities to all other phosphoribosyl transferases, and many other nucleotide binding proteins, in regions which probably serve closely similar functions across diverse protein families. Predicted secondary structures of CHO APRT are very similar to other APRT molecules, and to a lesser degree to other phosphoribosyl transferases. Residues of part of the generalized nucleotide binding motif of APRT were found to have specific roles in binding substrate, which can be extrapolated to the same functional elements in other nucleotide binding proteins. In addition, mutants identified by selection for altered substrate affinities are widely dispersed in the primary sequence. Although APRT is thought to exist as a dimer in its native context, certain mutants of APRT which have impaired ability to form dimers appear to have near-wildtype activity.

Table of Contents

LIST OF ABBREVIATIONS

APRT	adenine phosphoribosyl transferase (protein)
aprt	adenine phosphoribosyl transferase (mammalian gene)
apt †	adenine phosphoribosyl transferase (bacterial gene)
ATP	adenosine triphosphate
AMP	adenosine monophosphate
AdR	adenosine
%C	percent (w/w) of total acrylamide present as bisacrylamide
CHO	Chinese hamster ovary (cell line)
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
DNA	deoxyribonucleic acid
dTTP	deoxythymidine triphosphate
E. coli	*Escherichia coli*
g	gram(s)
GMP	guanosine monophosphate
HPRT	hypoxanthine phosphoribosyl transferase (protein)
hgprt	hypoxanthine/guanine phosphoribosyl transferase (gene)
hpt †	hypoxanthine phosphoribosyl transferase (bacterial gene)
IMP	inosine monophosphate
K_M	Michealis constant
kDa	kilodaltons
μ	micro
m	milli (*e.g.* millimole)
M	molar
ml	millilitre
MBP	maltose binding protein (Appendix II)
MCS	multiple cloning site
mRNA	messenger ribonucleic acid
n	nano (*e.g.* nanomole)
NT	nucleotide
ntbc	nucleotide binding cassette
OPRT	orotate phosphoribosyl transferase (protein)
p	pico (*e.g.* picomole)
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pup	purine nucleoside phosphorylase

PRPP	1'-phosphoribosyl 5'pyrophosphate
%T	total concentration of acrylamide monomer (grams per 100 ml)
V_{max}	maximal velocity of an enzyme
XPRT	xanthine phosphoribosyl transferase (protein)

List of Tables

List of Figures

Acknowledgements

I gratefully acknowledge the generous scientific and financial support of Dr. Barry Glickman, who has also been a mentor and friend. At the Centre for Environmental Health, Dr. Johan de Boer and James Holcroft have contributed directly and indirectly to the project. Pauline Tymchuk has been a calming force, keeping in order years of documents, meetings, and of course, at least one confused student. I would like to thank John Curry, Mike Parlee, and Andrew McArthur for assistance with various computational issues. My family have been a persistent source of strength and good food. Finally, I would like to thank Dr. Moyra Brackley, with whom I have shared an office while writing this dissertation, for teaching me the value of getting something finished.

Dedication

for Catharine

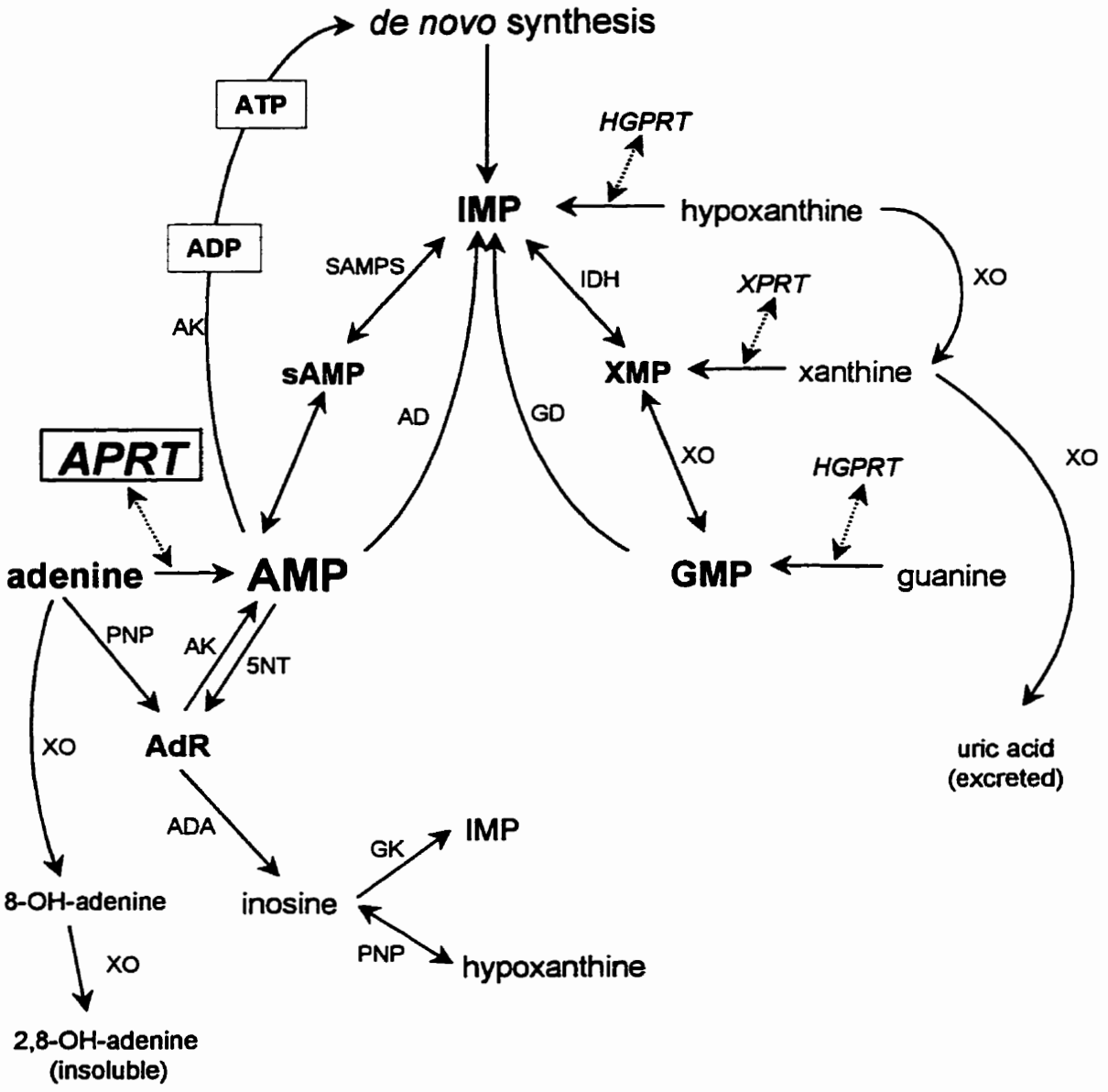
1.1 Perspectives

In the normal cell, adenine is the most abundant purine base. It is a component of the main energy storage molecule of cells, adenosine triphosphate (ATP). Adenine may be synthesised *de novo* from precursors, including phosphoribosylpyrophosphate (PRPP), or recovered for use in the cell by a specific enzyme, adenine phosphoribosyl transferase (APRT). APRT is a member of a family of enzymes which transfer the phosphoribosyl (Prib) moiety of phosphoribosyl pyrophosphate, to acceptor substrates which range from nitrogenous bases and amino acids, to nucleotides. Phosphoribosyl transferases are widespread in nature, with functions in *de novo* nucleotide and amino acid synthesis, and nucleotide salvage (Dean *et al.*, 1968; Bell and Koshland, 1970; Queen *et al.*, 1989).

APRT catalyzes the addition of phosphoribosyl to free adenine in the presence of divalent metal cations, usually magnesium, resulting in the formation of adenosine monophosphate (AMP). This so-called salvage reaction is the primary means by which free adenine is recovered by mammalian cells. Separate phosphoribosyl transferases exist for the parallel recovery of uracil, guanine and hypoxanthine, and xanthine. In the case of purine base salvage, the recovery of each base is interconnected with IMP, the end product of *de novo* synthesis, as shown in Figure 1.1, such that any single purine nucleoside (such as AMP or GMP) can satisfy the needs of the cell

As a result of scavenging, and *de novo* adenine synthesis pathways, APRT is a "non-essential" enzyme. This means that in the cell, deficiency of APRT is generally not lethal (Henderson *et al.*, 1968). As a result, cells which contain mutations in the *aprt* gene that destroy or reduce function of the enzyme product can survive. This characteristic

Figure 1.1 : The relationship of adenine phosphoribosyl transferase (APRT) to other components of purine metabolism, the central role of IMP, and the interconverison and alternate salvage of purine bases. In some species (*e.g. E. coli*), hypoxanthine-guanine phospho-ribosyl transferase (HGPRT) is replaced by two discrete enzymes, HPRT and GPRT. Cofactors and cosubstrates are not indicated. Alternate salvage of guanine, hypoxanthine, and xanthine, to their respective nucleosides, by purine nucleoside phosphorylase and 5'nucleo-tidase, is not shown. AD; adenylate deaminase. AdR; adenosine. AK; adenosine kinase. AMP; adenosine monophosphate. GD; guanylate deaminase. GK; guanosine kinase. GMP; guanosine monophosphate. IDH; IMP dehydrogenase. IMP; inosine monophosphate. PNP; purine nucleoside phosphor-ylase. SAMPS; adenylosuccinate synthetase. sAMP; succinyl AMP. XO; xanthine oxidase. XMP; xanthine monophosphate. XPRT; xanthine phosphribosyl transferase



makes it attractive to use *aprt* as a target in studies of the genetic effects of mutagens through the selection of viable cells that have lost (due to mutation) APRT activity.

APRT deficiency in humans has a range of clinical effects, from gouty arthritis to urolithiasis (Emmerson *et al.*, 1975; Fox *et al.*, 1973; Fox, 1976). At the cellular level, increased levels of adenine and PRPP, which may occur in *aprt*$^-$ (APRT-deficient) cells, have been shown to cause growth inhibition of lymphocytes (Snyder *et al.*, 1976). The related enzyme hypoxanthine-guanine phosphoribosyl transferase, when lacking in humans, has similar effects at the cellular level, but much more damaging effect for the organism, leading in some cases to a crippling neurodegenerative condition called Lesch-Nyhan disease (Lesch and Nyhan, 1964).

In addition to adenine, APRT can catalyze phosphoribosyl transfer to certain analogs which are structurally related to adenine. Analogues of adenine, which are not particularly toxic *per se*, become very poisonous to the cell after processing by APRT into nucleotide analogs (Smith and Matthews, 1957; Parks *et al.*, 1973). These base analogues have been used *in vitro* for selecting *aprt*$^-$ mutants and include 8-azaadenine, 2,6-diaminopurine (Le Page and Greenlees, 1955), 2-fluoroadenine (Montgomery, 1982), 6-N-hydroxylamino purine, 4-amino-pyrazolo-pyrimidine (Avila and Casanova, 1982) , 6-amino-8-methylpurine and 4-carbamoylimidazolium-5-olate (mizoribine aglycone; Koyama and Kodama, 1982). Figure 1.2 illustrates a comparison of adenine to the structures of some of these compounds. Alterations at the purine ring in these analogues can be atom substitutions, open rings, addition of an additional or atypical sidegroup, or changes in bond structures.

Figure 1.2

The normal adenine phosphoribosyl transferase is able to bind many of these analogues and catalyse their reaction with PRPP, thus yielding cytotoxic nucleotide analogues. Indeed, the presence of a functional phosphoribosyl transferase is required for these analogues to kill the cell. Several of the analogues of adenine have been previously evaluated for their ability to become cytotoxic in cells which posses functional APRT (Smith and Matthews, 1957; Hitchings and Elion, 1961; Bennet *et al.*, 1966; Montgomery, 1982) . It has been reported that the analogues generally bind with lower affinity than the normal substrate (Krenitsky *et al.*, 1969a). As a result, it would be expected that normal adenine phosphoribosyl transferase enzyme would produce mainly AMP, and little of the toxic nucleotide analogue, based on the relative affinity of the enzyme for the different substrates. However, estimates by HPLC of the intracellular concentration of nucleotide analogues suggests that only a few molecules of the analogue as nucleotide are sufficient to cause cell death (Parks *et al.*, 1973; Van Diggelen *et al.*, 1979; Smolenski *et al.*, 1991). Prokaryotic cells seem to be able to tolerate much higher media concentrations of purine analogs than mammalian cells, but this may be due in part to the poorer transit of analogs across the bacterial cell membrane.

The precise mechanism of this toxicity is unknown, but this has not impeded investigations of the chemotherapeutic potential of analogues of various bases, primarily processed by phosphoribosyl transferases, in the treatment of neoplasia (Hitchings, 1950; Hitchings and Elion, 1961; Bhalla *et al.*, 1984; Choi *et al.*, 1992), parasitic diseases (Nakamura and James, 1951; Queen *et al.*, 1989), hematological disorders (Parks *et al.*, 1973; Hashimoto *et al.*, 1990), gout, and others (Roblin *et al.*, 1945; Natsumeda *et al.*, 1984). For example, the malarial parasite *Plasmodium falciparum* expresses its own

APRT, which has approximately ten-fold higher affinities for adenine and PRPP than its human counterpart (Queen *et al.*, 1989). Using low doses of appropriate adenine analogs, it may be possible to selectively kill the parasite *in situ*. Conversely, the salvage of nucleosides and bases has been investigated as a potential source of resistance to chemotherapeutic nucleotide analogues (Pillwein *et al.*, 1990; Fox *et al.*, 1991).

Adenine analogues have also been exploited in many investigations into mutagenic mechanisms using the *aprt* gene as a mutational target (Kocahryan *et al.*, 1975; Adair, 1980; Taylor, 1985; Grosovsky *et al.*, 1986; Phear *et al.*, 1987; de Boer and Glickman., 1989), . Several groups have developed a substantial collection of mutants in the coding sequence of the gene, most coding for amino acid alterations (Drobetsky *et al.*, 1987; de Boer and Glickman, 1991). These collections of induced mutations have been compared to spontaneous mutations occurring both *in vivo* and *in vitro* (Nalbantoglou *et al.*, 1987; Hidaka *et al.*, 1987; Hidaka *et al.*, 1988). The mutations are distributed unevenly over the length of the primary structure. Many of the phenotypically identifiable mutants in these collections have no detectable residual enzyme activity, nor are the mutant cell lines able to survive in conditions which require adenine for growth (Thompson *et al.*, 1980; Carver *et al.*, 1980; Taylor *et al.*, 1985, Turker and Martin, 1985). Conversely, some mutants selected on the basis of resistance to adenine analogues, still have detectable enzyme activity, implying that the protein has altered affinity for adenine analogues, such that insufficient toxic analogue is produced by the enzyme to cause a lethal response in the cell. A similar degree and scope of exploitation of the *hgprt* gene has also been undertaken (Lieberman and Ove, 1960; Stout, 1985; King *et al.*, 1994). In the case of *hgprt*, mutational studies are simplified by the actual or functional hemizygosity of

the gene human cells.

The ability of an enzyme to discriminate and bind substrates in a highly specific manner is based upon interactions between amino acid side chains in the enzyme and various atoms in the substrate. While biochemical work has been reported on APRT (Groth and Young, 1971; Holden *et al.*, 1979) and HGPRT (Hughes *et al.*, 1975; Ali *et al.*, 1982), as well as on some of the related enzymes (Bell and Koshland, 1970; Kleeman and Parsons, 1976; Victor *et al.*, 1979), to date little has been done from the molecular point of view. X-ray crystallographic structures are available only for a few phosphoribosyl transferases, and none for APRT. A model for part of the phosphoribosyl transferase molecule has been proposed by Busetta (1988). These enzymes, which are relatively small (Chinese hamster APRT monomer is approximately 20 kDa), have structural features that are predicted to be similar to other nucleotide binding proteins such as dehydrogenases (Argos *et al.* 1983) and the *ras* oncogene protein (Fry *et al.*, 1986; Valencia *et al.*, 1991).

1.2 Physiological role of adenine phosphoribosyl transferase

1.2.1 APRT in context

APRT occupies a peripheral role in the overall scheme of nucleotide metabolism. as illustrated in Figure 1.1, showing the interconnected pathways of purine base synthesis, salvage, and catabolism (Murray, 1971; Nygaard, 1976; Musick, 1981; Chappel and Slaytor, 1992). In the cell, free adenine is either salvaged as AMP, or is catabolized into 2,8-dihydroxyadenine by xanthine oxidase. Dihydroxyadenine has limited solubility, so the cellular or plasma concentrations of this molecule must be maintained at low levels, to prevent precipitation of the acid. In people who are APRT deficient, dihydroxyadenine

kidney stones, leading to renal insufficiency or failure are the commonest cause of symptoms (Fujimori *et al.*, 1985; Zoellner and Gresser, 1990; Chen *et al.*, 1991; Hakoda *et al.*, 1991, Soga *et al.*, 1995). Alternative routes of purine metabolism, such as the removal of excess guanine and inosine via uric acid, also catalyzed by xanthine oxidase, have higher capacity due to the greater solubility of uric acid (Wilson *et al.*, 1983; Wilson and Kelley, 1984; Yakota *et al.*, 1991) . This is somewhat paradoxical, since adenine in nucleosides and nucleotides is 45

more abundant than other purine bases, its degradation products should also therefore be more abundant. But the poor solubility of 2,8-dihydroxyadenine, and its absence of accumulation in normal cells, suggests that most adenine which is ultimately degraded, must first be converted to IMP after salvage by APRT. Thus, it is advantageous for the cell to be able to process some, or most, adenine excess through this alternative pathway of catabolism. APRT provides the route by which free adenine may be either salvaged for use as other bases, or routed through a more efficient catabolic mechanism.

The physiological utility of APRT other than in routing excess adenine to catabolism, is probably energy conservation. *De novo* synthesis of a molecule of adenine as AMP (using preformed precursors), requires an energy expenditure approximately equivalent to six ATP molecules, whereas salvage of adenine as AMP by APRT has a net cost of just one ATP molecule. If the energy cost of precursors is also considered, salvage of purine bases generally should be advantageous for the cell. Obviously, it should be more efficient for the cell to recycle free adenine than to produce it *de novo*. If this is the case, then cells which are deficient in APRT might be expected to exhibit reduced growth

rates. There is some evidence that this may be the case in human lymphocytes (Snyder *et al.*, 1976; Snyder *et al.*, 1993), but this does not seem to apply to bacteria, or immortalized cells in culture (Levine *et al.*, 1982; Nguyen *et al.*, 1984; Dare *et al.*, 1992).

An additional role for APRT is found in muscle tissue in the so-called ATP cycle. During vigorous exercise, muscular ATP is depleted, and energy costs can be reduced by exploiting nucleotide salvage and salvage intermediates (Murphy and Tulley, 1984; Tullson and Terjung, 1991).

1.2.2 Clinical manifestations of APRT deficiency

The clinical relevance of alterations in APRT activity is ambiguous. There is some evidence that APRT deficiency may be a significant causal factor in the formation of kidney accretions leading to dysuria, renolithiasis and renal failure due to the accumulation of insoluble 2,8-dihydroxyadenine (Emmerson *et al.*, 1975; Soga *et al.*, 1995). Complete absence of APRT activity leads to dihydroxyadenine renolithiasis in adolescence or early adulthood in greater than 85% of affected individuals (Fujimori *et al.*, 1985; Sahota *et al.*, 1991; Yokota *et al.*, 1991; Kambayashi *et al.*, 1994). A retrospective analysis of pathological specimens of kidney stones is underway in Britain (A. Simmonds, pers. comm.), in order to establish the frequency of occurrence of 2,8-dihydroxyadenine stones in renal patients. It has been reported that the usual treatment for uric acid (gouty) kidney stones, caused by a partial defect in HGPRT, is also effective in the treatment of dihydroxyadenine stones caused by APRT deficiency (Simmonds *et al.*, 1992). This may lead to an underassessment of the frequency of renolithiasis linked to APRT.

In the population of the United Kingdom (approximately 5.8×10^7), only a few

homozygous APRT-deficient individuals have been found. But the frequency of heterozygosity of *aprt* is estimated from population sampling to be about 0.01 (Johnson *et al.*, 1977; Kamatani *et al.*, 1987). That is to say, in the United Kingdom, there ought to be about 500,000 heterozygous individuals, assuming there is no strong prenatal selective pressure against homozygous deficiency, and about 1400 homozygous-deficient individuals (many of whom would be symptomatic with kidney disease attributable to 2,8-dihydroxyadenine). Less than 20 symptomatic individuals have been identified (H.A. Simmonds, pers. comm.). Thus there appears to be a considerable deficit in the frequency of homozygous-deficient persons. It has been suggested that this deficiency may be due to an obligate requirement for the embryo, or a specific embryonic tissue, to exploit salvage of purines during some crucial stage of development. In most cases, it is proposed, complete deficiency of APRT activity causes abortion at that stage. There is however, only sparse experimental evidence for this hypothesis, including the observation that an X-linked locus, responsive to suppression by methylation imprinting during embryogenesis, may exert control over *aprt* expression during the same period (Moore and Whittingham, 1992; Singer *et al.*, 1992). It has been reported that the purine salvage phosphoribosyl transferases are differentially expressed during early embryogenesis (Schopf *et al.*, 1984; Moore and Whittingham, 1992; Alexiou and Leese, 1994). Conversely, homozygous APRT-deficient mouse embryos live to full term with normal development, but later develop lethal dihydroxyadenine renolithiasis (Engles *et al.*, 1996). Alternatively, it is possible that the assertion of symptomatic deficiency is overstated, and that many persons, completely lacking APRT activity, are otherwise asymptomatic.

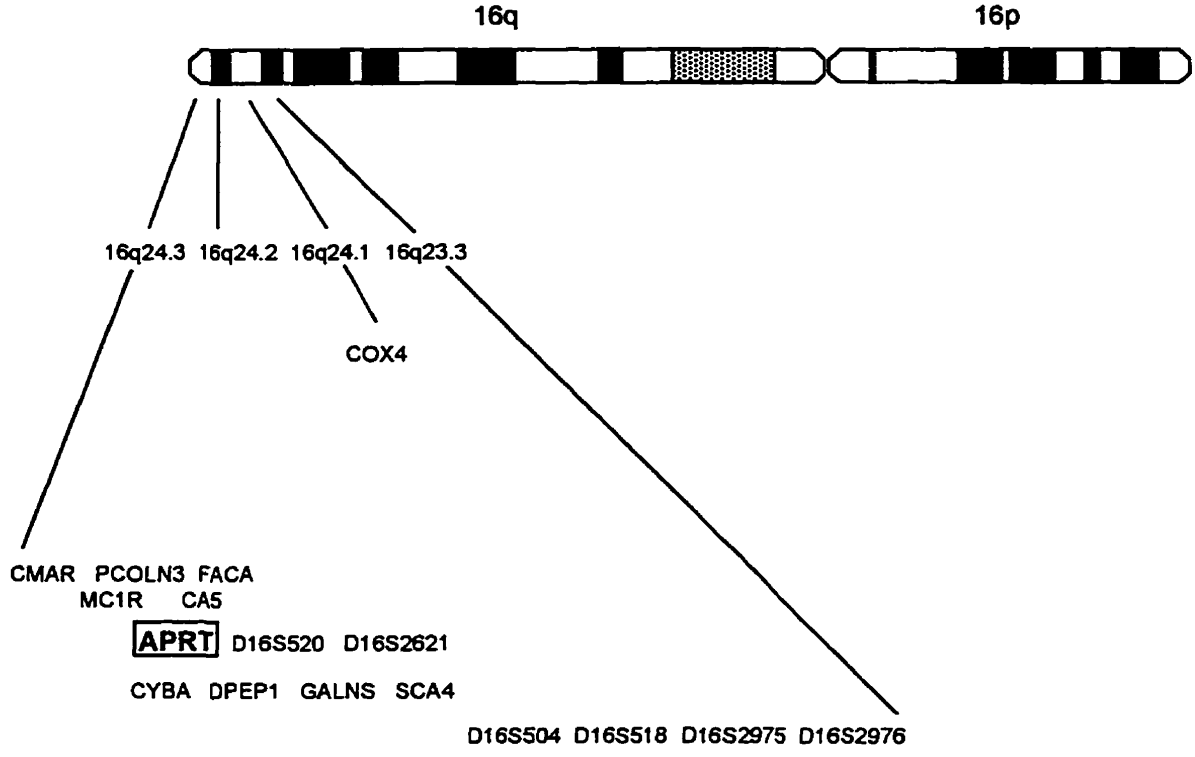
An interesting new clinical role for the *aprt* gene has recently been described (Stambrook *et al.*, 1996). *Aprt* is on human chromosome 16q24.2 (Barg *et al.*, 1982). In this region are found several other important genes, including cytochrome C oxidase, Fanconi's anemia, and others (Fratini *et al.*, 1986; Cleton Jansen *et al.*, 1995). An idiogram of this chromosomal region in Figure 1.3 indicates the position of *aprt* relative to other markers, including RFLPs. In certain neoplasias, genomic instability has been identified as an indicator of disease progression, related to altered or diminished DNA repair capacity (Boyer *et al.*, 1995; Peltomaki and de la Chappelle, 1997) . The region which includes *aprt* appears to be susceptible to deletions and other forms of genomic instability (Cooper *et al.*, 1991; Cooper *et al.*,1992; Smith and Grosovsky, 1993), such that *aprt* is a useful marker in genetic analysis of clonality and progression (Harwood *et al.*, 1991; Phear *et al.*, 1996; Shao *et al.*, 1996; Gupta *et al.*, 1997). It is possible that the same region contains important DNA-repair genes, which have not been characterized, for which *aprt* may be a linkage marker. It is interesting to speculate that disturbances of methylation patterns during neoplastic progression (Branch *et al.*, 1995), may have an impact on the susceptibility of the *aprt* gene and adjacent 16q markers to loss of heterozygosity and DNA misrepair (Sanfilippo *et al.*, 1994).

1.3 Biochemical Background to APRT

1.3.1 The APRT reaction

APRT activity was first detected in bacterial cells as a sensitivity or resistance to purine analogues (Roblin *et al.*, 1945; Smith and Matthews, 1957). This observation led to the testing of purine analogs against parasitic diseases (Nakamura and James, 1951

Figure 1.3 Chromosomal location of the human *aprt* gene, and surrounding markers including genetic and RFLP markers. The locus map and adjacent markers are taken from the human chromosome genome database (GDB5.6) at the National Centres for Biological Information (Callen *et al.*, 1991). The CHO locus which is not as well mapped, resides on chromosome 11 of the Chinese hamster genome. Genetic markers: APRT; adenine phosphoribosyltransferase. COX4; cytochrome C oxidase. CA5; carboni anhydrase. CYBA; cytochrome b-245, alpha subunit. DPEP1; dipeptidase 1. FACA; Fanconi's anemia. GALNS; galactosamine synthetase. MC1R; melanocortin 1 receptor. PCOLN3; procollagen Type III. SCA4; spinocerebellar ataxia. The markers D16S520 *etc.* indicate RFLP probe markers which have been mapped in the region.



; Hitchings and Elion, 1961). The biochemistry of APRT was later explored (Hori and Henderson, 1966a, 1966b), followed by the first report of a clinical case of APRT deficiency (Kelley *et al.*, 1968). To date biochemical data on APRT from many species across three kingdoms has been collected (Table 1.1). Notable is the range of substrate affinities, especially for PRPP. This variation is suggestive of different requirements for adenine salvage. For example, in *Plasmodium falciparum*, like most parasitic protozoans, which possesses no *de novo* adenine synthetic pathways, adenine salvage is a crucial survival mechanism (Queen *et al.*, 1989; Asahi *et al.*, 1996). Similarly the Australian termite, *Nasutitermes walkeri*, has a highly active APRT enzyme, but no salvage of hypoxanthine or guanine (Chappell and Slaytor, 1992).

Analysis of isolated APRT has revealed a burst of AMP synthesis when adenine is added to the purified enzyme (Kenimer *et al.*, 1975; Nagy and Ribet, 1977; Arnold and Kelley, 1978). A schematic of the APRT reaction sequence is shown in Figure 1.4. Note that only the first step, interaction with PRPP, is magnesium dependent (Berlin 1969; Gadd and Henderson, 1970). This is a consistent observation among phosphoribosyl transferases (Dodin *et al.*, 1982; Yuan *et al.*, 1992; Bhatia and Grubmeyer, 1993; Tao *et al.*, 1996). Indeed, there is evidence from solution NMR studies that it is the interaction of magnesium ion with PRPP that "locks" the PRPP into a conformation suitable to binding with the enzyme (Smithers and O'Sullivan, 1982). In isolated preparations, burst synthesis of AMP after adenine addition, can be observed even in the presence of excess amounts of strong chelating agents (Montero and Llorente, 1991). This implies that the cosubstrate, or at least part of it, must be resident in a stable or metastable form, on the

Table 1.1 (1 of 3): Characteristics of phopshoribosyl transferases from various sources. Values of K_M for base substrate and PRPP are in μM. Values not reported are blank. Where K_M has been measured in the presence of alternate substrates, these are denoted by abbreviations: ade; adenine gua; guanine. hypo; hypoxanthine. xan; xanthine. glut; glutamine. amm; ammonia.

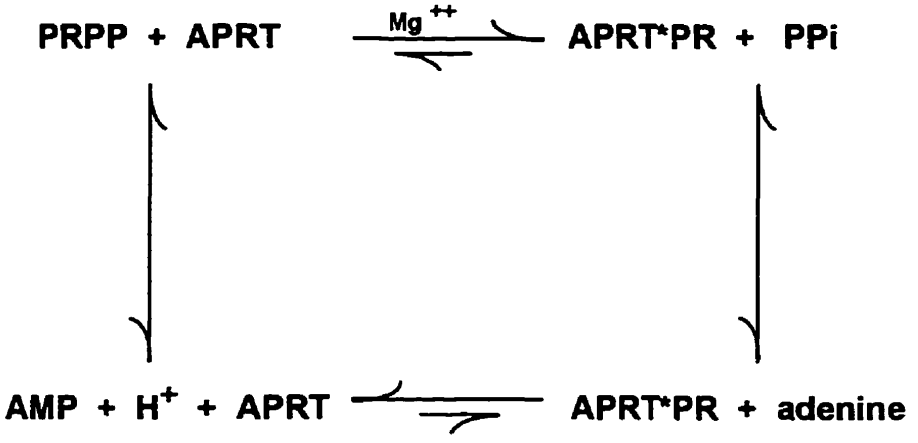
Table 1.1 (2 of 3):

Source	K_M base	K_M PRPP	Mol Wt.(Dal)	Reference
APRT				
Cricetulus griseus	1.0	3.0	18,000	Hershey and Taylor, 1978
Homo sapiens	1.4-2.6	6.2-6.5	18,000	Arnold and Kelley, 1978; Srivastava *et al.*, 1971; Dean *et al.*, 1968; Hori and Henderson, 1966
Mus musculus	6.6	1.2	18,000	Okada *et al.*, 1986
Rattus norvegicus	0.9	2.0-5.0	22,000	Natsumeda *et al.*, 1984; Goth *et al.*, 1978
Arabadopsis thaliana	4.5	290	27,000	Lee and Moffat, 1993
Helianthus tuberosa	5.5	6.4		le Floc'h *et al.*, 1978
Plasmodium falciparum	0.8	0.7	18,000	Queen *et al.*, 1989
Artemia cystis	2.0	30	15,000	Montero and Llorente, 1991
Schizosaccaromyces pombe	69	32	30,000	Nagy and Ribet, 1977
E.coli	10	150	20,000	Hochstadt-Ozer, 1971
B. subtilis	22	9	23,000	Berlin, 1969
S. tymphimurium	10	3.4		Kalle and Gots, 1963
HGPRT				
Homo sapiens	3.5-17 hypo 5 gua	6.6-16 hypo 2-5.5 gua	34,000	Giacomello and Salerno, 1977; Stot and Caskey, 1986
Artemia cystis	1.0 hypo 1.0 gua	15	19,000	Montero and Llorente, 1991
Cricetulus griseus	0.52 hypo 1.1 gua	2 hypo 4 gua		Natsumeda *et al.*, 1984
Rattus norvegicus	70 gua	50	33,000	Kim *et al.*, 1992

Table 1.1 (3 of 3)

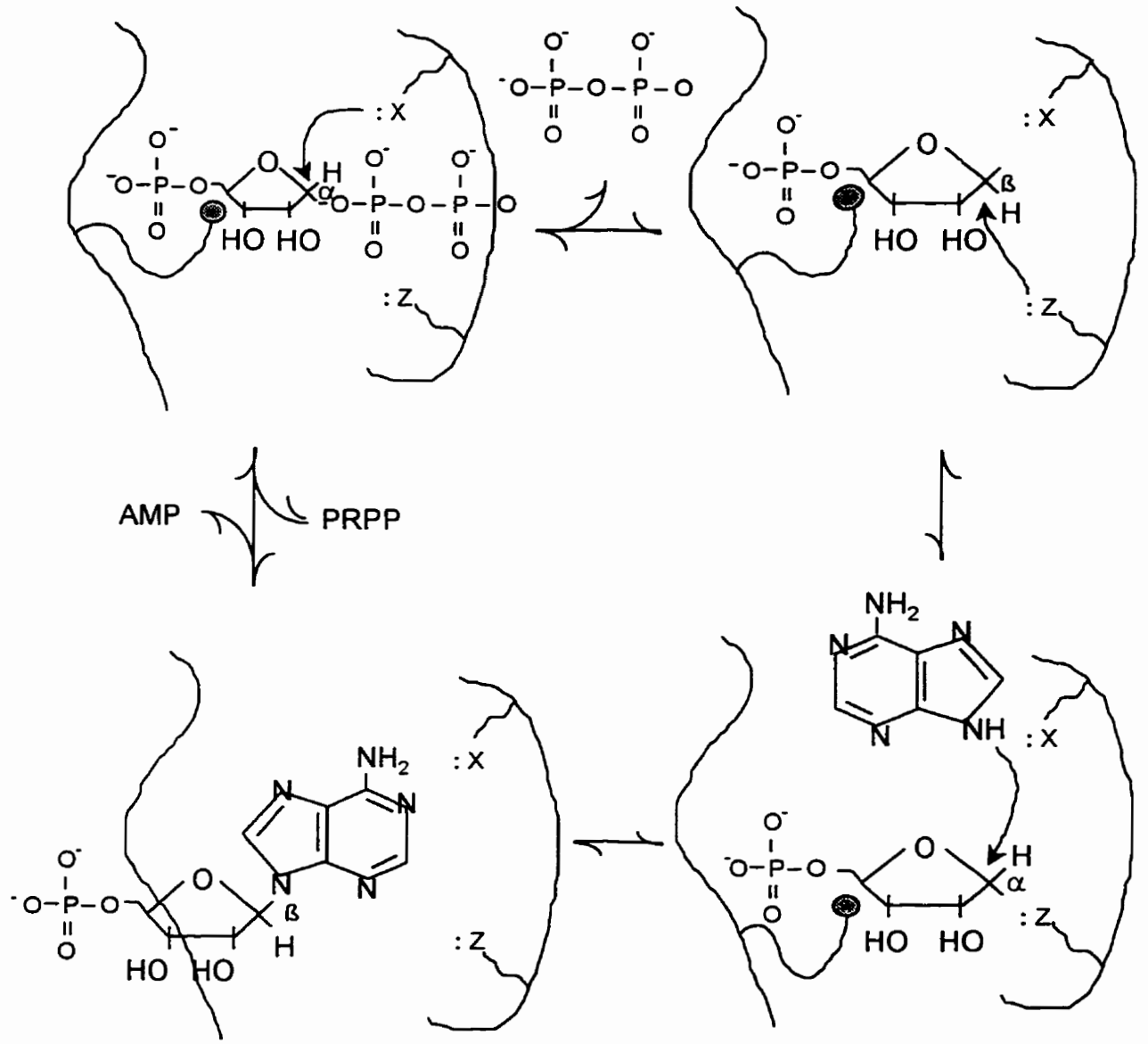
Source	Km base	Km PRPP	Mol Wt.(Dal)	Reference
HGPRT (cont.)				
Giardia lamblia	74 gua		26,300	Sommer and Wang, 1996
Tritrichomonas foetus	4.1 hypo 3.8 gua 52.4 xan	74.2 hypo 31.4 gua 38 xan	24,000	Beck and Wang, 1993
S.cerevisiae	26 hypo 41 gua	47	30,000	Ali and Sloan, 1982
Schizosaccarmyces pombe	28	100	24,000	Nagy and Ribet, 1977
Schistosoma mansoni	3.0 gua 5.4 hypo	18.2 gua 9.3 hypo		Yuan *et al.*, 1992
E.coli	2.6 gua 169 hypo 39 xan	95	18,600	Liu and Milman, 1983
OPRT				
S.cerevisiae	15-35	38	20,000	Ashton *et al.*, 1989; Syed and Sloan, 1990; Shostak *et al.*, 1990
Quinolinate PRT				
S.typhimurium	20	32	32,400	Hughes *et al.*, 1993
Pseudomonas sp.	100	340		
ATP-PRT				
S.typhimurium	90-200	17-67		Kleeman and Parsons, 1976 Martin, 1963
Glutamine Amido-PRT				
Homo sapiens	460 glu 710 amm	400		Boss *et al.*, 1983

Figure 1.4 : Reaction schematic for APRT. PR; phospribose moiety of PRPP. PPI; pyrophosphate. APRT*PR; activated enzyme-phosphoribose intermediate. Note that only the interaction between PRPP and the enzyme is magnesium-dependent. Subsequent steps are independent of metal ion. The reaction pattern of interaction with first one substrate, then the next, with a stable (or metastable) activated enzyme intermediate is the so-called Bi-Bi-Ping-Pong reaction.



isolated enzyme (Groth and Young, 1971). In addition, the final product of the reaction exhibits inverted stereochemistry around the C-1 of the ribose, where the catalytic events take place (Popjak, 1970; Chelsky and Parsons, 1975). This suggests that there is a multistep interaction of the PRPP with the catalytic site, in a so-called surface walk, where reactive intermediate is moved from one reactive side-group to another in the active site (Spector, 1982). A model of this scheme has been proposed by Spector, shown in Figure 1.5. This reaction scheme is called bi-bi Ping-Pong[®], restating the observation that there are two reactants, and two products (AMP and pyrophosphate), which are used and released in an ordered nonsimultaneous reaction. This differs from HGPRT, OPRT, and XPRT, which exhibit Ping Pong[*] bi-bi reaction schemes, in which a stable enzyme-PRPP intermediate cannot be isolated, and no burst of product formation is observed when substrate base is added to purified enzyme (Bell and Koshland, 1970; Syed *et al.*, 1987). In these enzymes, the PRPP-enzyme complex can undergo substitution of PRPP with radiolabelled PRPP in solution, in the absence of cosubstrate. In contrast to APRT, the initial burst of product synthesis upon addition of base to the enzyme PRPP mixture can be obliterated by preincubation with EDTA or other chelators (Ali and Sloan, 1982; Ashton *et al.*, 1989). Thus some critical component of the APRT reaction mechanism differs from other phosphoribosyl transferases, and this difference may manifest as a unique three-dimensional structural difference, which in turn should be identifiable as a conserved sequence among APRTs, representing a difference in primary structure from other phosphoribosyl transferases.

Figure 1.5 : Adaptation from Spector (1988) of a model of the catalytic events leading to inversion of stereochemistry around C-1 of the ribose moiety. In addition, the phosphate-ribose ester bond at C-5 that is retained in the nucleotide is the region that is invariant in nucleotide structures. This region may be important for non-catalytic binding and the specificity of binding of these molecules to ntbc proteins.



Direct physical analysis of purified APRT from several systems, reveals that APRT exists as a homodimeric molecule (Dean *et al.*, 1968; Srivastava and Beutler, 1971; Thomas *et al.*, 1973). In comparison, HGPRT is homotrimeric (Krenitsky *et al.*, 1969b; Arnold *et al.*, 1974; Olsen and Milman, 1974; Holden and Kelley, 1978) or homotetrameric (Johnson *et al.*, 1979), and OPRT is either monomeric or a homodimer, and indeed, may also be a component of a larger bifunctional enzyme (Scapin *et al.*, 1993; Agharjari *et al.*, 1994). This heterogeneity of quaternary structure would suggest that these molecules may have different structural elements for determining or stabilizing the final assembled molecule (Russell and Barton, 1994).

1.3.2 Implications of the reaction scheme

From the reaction schematic, and the model of the surface-walk proposed by Spector (1982), it is apparent that the catalytic site of APRT must contain multiple groups to which the PRPP moiety must bind. Although Spector's original model shows an actual transit of the reactive phosphoribosyl intermediate between groups within the catalytic site, it is also possible that the PRPP is initially simultaneously bound to two side-chains, one of which becomes detached after the release of the pyrophosphate, such that the substrate is never free during any stage of the reaction. Indeed, it may also be possible that the catalytic groups are actually distinct from the substrate binding groups, implying that the C-1 end of the ribose is free to react with multiple side-chains. In such a case, catalytically nonfunctional mutants of the enzyme might be able to bind PRPP with similar affinity to the wildtype, but be unable to catalyze phosphoribosyl transfer. Conversely, it may be possible to identify mutants which bind substrates with altered affinity, but still be

functional. Such a catalytically competent enzyme, mutated in a region with strong homology to a region in APRT, with altered substrate binding properties, has previously been described in the β-subunit of proton-ATPase from *E. coli* (Al Shawi *et al.*, 1988). Unfortunately, binding of PRPP to catalytically incompetent enzymes would be predicted to be very difficult to measure using conventional kinetic methods. All of these models of the interactions at the catalytic site are consistent with available data.

1.3.3 Predictions on structure-function of APRT

From biochemical data on APRT, it is apparent that there must be regions of primary structure quite different from other phosphoribosyl transferases in order to account for the different reaction pathway. The presence of a stable enzyme-phosphoribosyl intermediate, which cannot be identified for other phosphoribosyl transferases, suggests that there may be discrete residues of APRT which may be identifiable from comparative sequence analysis, which fulfill the unique needs of the surface walk model proposed by Spector.

Any of these predictions may be valid, but more central to the problem is that the catalytic sites of APRT, may actually not be conserved in structure, across long evolutionary periods, since the APRT reaction is somewhat different from the other phosphoribosyl transferase reaction schemes. Thus one would expect to identify regions of primary structure which are conserved in APRT among phosphoribosyl transferases, and sequences which are unique to APRT molecules.

By the same reasoning, the dimerization domains of APRT might be identifiable from sequence comparisons, by virtue of their conserved structural nature, relative to

other phosphoribosyl transferases. One might predict that the dimerization functions of APRT have been preserved through evolution, while the corresponding regions of other phosphoribosyl transferases may be highly divergent. Characteristics of interchain interaction domains from multimeric proteins, have been described by several authors (Argos, 1988; Perry *et al.*, 1989; Shindalyov *et al.*, 1994; Henriksen *et al.*, 1996), and may provide testable models against which sequence domains of APRT may be assessed. Using simple elimination (*e.g.* sequences which are catalytic are probably not also responsible for dimerization), it should be possible to reduce the likely regions of dimer interactions to one or two candidates. Under some conditions, it may also be possible to determine whether APRT must form a dimer in order to become functional, such that molecules which can be shown to have poor or no dimerization, still exhibit considerable APRT activity. If residues of each monomer are required to form the active site, such as has been recently described in OPRT from *S. typhimurium* (Henriksen *et al.*, 1996), one would predict that monomeric molecules are not catalytically competent.

1.4 Sequence comparison of APRT from various species.

A useful starting point in structure/function analysis is to identify regions of protein sequence which are evolutionarily conserved among functionally equivalent forms of the molecule of interest. Figure 1.6 is an alignment of adenine phosphoribosyl transferases, organized with hamster sequences first (see Chapter 2). Regions of strong sequence similarity are indicated. The data are adapted from various sources as noted in the figure legend, using alignment methodology and programs as described in Chapter 2. These conserved regions presumably represent catalytically or structurally indispensable regions

under strong selective pressure. Note that the human enzyme differs from the CHO enzyme by 16 amino acids, most of these being highly conservative substitutions, thus there is no reason to assume *a priori* that these two enzyme differ in any significant way in tertiary structure. The *E. coli* enzyme however shares only about 40% percent sequence identity or similarity with the human / CHO sequences. Thus it is possible that the bacterial enzymes, or others which are less similar, despite their shared reaction mechanism, are structurally much different from the vertebrate types. Note also that the sequence of the *Arabadopsis thaliana* enzyme is not more similar to the vertebrate type (41-46% identity / similarity) or to fungal sequences (20-22%) than to the bacterial types (50-58% identity / similarity). A similar comparison can be made relative to the *Drosophila sp.* APRT sequences. Thus sequence divergence within diverse eukaryotes is as great as that between eukaryotes and prokaryotes.

1.5 Mutagenesis of *aprt*

1.5.1 *In vivo* mutagenesis

Studies of mutational phenomena at the molecular level require a target gene which can be phenotypically monitored after mutagenic treatments. Adenine phosphoribosyl transferase presents a reasonable model system in mutagenesis for just this reason. Forward mutations can be identified by the phenotype of resistance to cytotoxic analogs of adenine, while revertants can be identified by their ability to sequester adenine as AMP, or to utilize adenine as the sole source of purine base in cultured cells.

Based on these properties, *aprt* has been used extensively in studies of mutagenesis, especially in cultured mammalian cells, where *aprt* has the advantage of

Figure 1.6 (1 of 2): Alignment of APRT primary protein sequences using Clustal V (Higgins and Sharp, 1989). Identical residues (>10/11) are indicated by **x** above the sequence. Highly conserved tracts (>7/11 identical) are indicated with * above the sequence, moderately-conserved residues (identical >4/11, conservative changes >5/11) with a !. Numbering of residues conforms to the CHO APRT sequence numbering from the initiating methionine. Details of the alignment method are in Chapter 2.

	Species
1	*Cricetulus griseus*
2	*Rattus norvegicus*
3	*Mus musculus*
4	*Homo sapiens*
5	*Drosophila melanogaster*
6	*Saccaromyces cerevisiae I*
7	*Saccaromyces cerevisiae II*
8	*Pseudomonas aeruginosa*
9	*Pseudomonas stutzeri*
10	*Escherichia coli*
11	*Arabadopsis thaliana*

```
          1         10         20         30         40
          |         |          |          |          |
               !   !   ** !**** **!!*!** *!* !*   !    ! !
1     -----MAESELQLVAQRIRSFPDFPIPGVLFRDISPLLKDPASFRASIRL
2     -----MSESELQLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRASIRL
3     -----MSEPELKLVARRIRVFPDFPIPGVLFRDISPLLKDPDSFRASIRL
4     -----MADSELQLVEQRIRSFPDFPTPGVVFRDISPVLKDPASFRAAIGL
5     MSPSISAEDKLDYVKSKIGEYPNFPKEGILFRDIFGALTDPKACVYLRDL
6     ---MSISESYAKEIKTAFRQFTDFPIEGEQFEDFLPIIGNPTLFQKLVHT
7     ---MSIA-SYAQELKLALHQYPNFPSEGILFEDFLPIFRNPGLFQKLIDA
8     ------MIFDEFTLKSQIRAVPDFPKPGVVFRDITPLFQSPRALRMTVDS
9     ------MIFDEFSIKTLIRPVQDFPRPGVVFRDITPLFQSPKALRMVADS
10    ---MTATAQQLEYLKNSIKSIQDYPKPGILFRDVTSLLEDPKAYALSIDL
11    MATEDVQDPRIAKIASSIRVIPDFPKPGIMFQDITTLLLDTEAFKDTIAL

          50         60         70         80         90
          |          |          |          |          |
         *!  *!         !  *!*** XXXX*** *X  X*!* X*!*X *XX**
1     LASHLKSTHGG-KIDYIAGLDSRGFLFGPSLAQELGLGCVLIRKRGKLPG
2     LAGHLKSTHGG-KIDYIAGLDSRGFLFGPSLAQELGVGCVLIRKRGKLPG
3     LASHLKSTHSG-KIDYIAGLDSRGFLFGPSLAQELGVGCVLIRKQGKLPG
4     LARHLKATHGG-RIDYIAGLDSRGFLFGPSLAQELGLGCVLIRKRGKLPG
5     LVDHIRESAP--EAEIIVGLDSRGFLFNLLIATELGLGCAPIRKKKKLAG
6     FKTHLEEKFAKEKIDFIAGIEARGLLFGPSLALALGVGFVPIRRVGKLPG
7     FKLHLEEAFPEVKIDYIVGLESRGFLFGPTLALALGVGFVPVRKAGKLPG
8     FVQRYIEA----DFSHIGAMDARGFLIGSAVAYALNKPLVLFRKQGKLPA
9     LIQRYVEA----DFTHIGALDARGFLVGSILAYELNKPLVLFRKQGKLPA
10    LVERYKNA----GITKVVGTEARGFLFGAPVALGLGVGFVPVRKPGKLPR
11    FVDRYKDK----GISVVAGVEARGFIFGPPIALAIGAKFVPMRKPKKLPG
```

Figure 1.6 (cont. 2 of 2) :

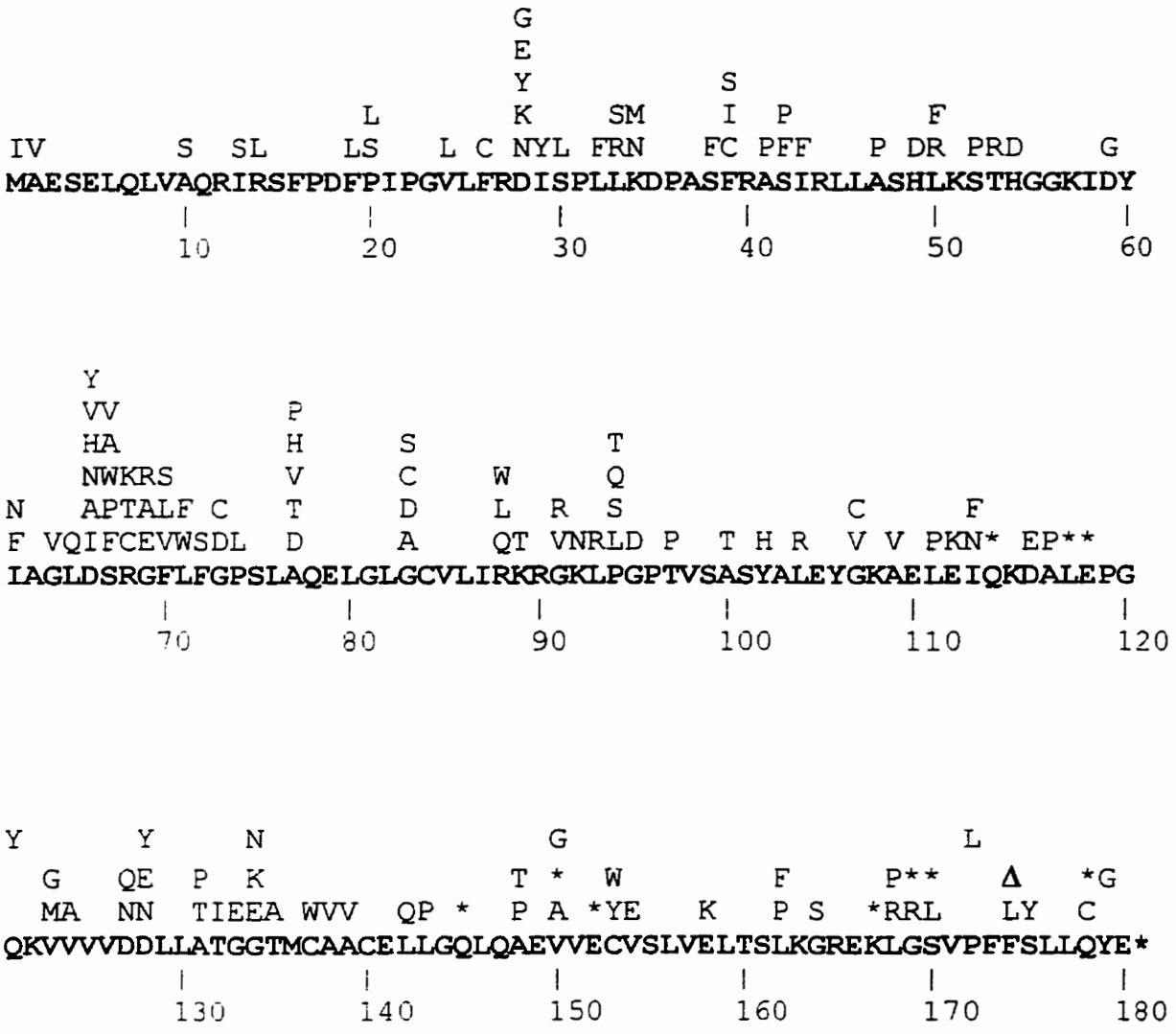
```
          100         110         120         130         140
          |           |           |           |           |
           *  !!!    !X!! !!*  **!!X***XXXXXXX*! XX  X!
1     PTVSASYALEYGKAELEIQKDALE-PGQKVVVVDDLLATGGTMCAACELL
2     PTVSASYSLEYGKAELEIQKDALE-PGQKVVIVDDLLATGGTMCAACELL
3     PTVSASYSLEYGKAELEIQKDALE-PGQRVVIVDDLLATGGTMFAACDLL
4     PTLWASYSLEYGKAELEIQKDALE-PGQRVVVVDDLLATGGTMNAACELL
5     EVVSVEYKLEYGSDTFELQKSAIKQPGQKVVVVDDLLATGGSLVAATELI
6     ECASITFTKLDHEEIFEMQVEAIP-FDSNVVVVDDVLATGGTAYAAGDLI
7     ECFKATYEKEYGSDLFEIQKNAIP-AGSNVIIVDDIIATGGSAAAAGELV
8     DVLAEGYQTEYGEAFLEVHADSLC-EGDSVLIFDDLIATGGTLLAAASLV
9     DVLSQAYSTEYGEAHLEIHADSLC-EGDSVLLFDDLIATGGTLLAAAQLV
10    ETISETYDLEYGTDQLEIHVDAIK-PGDKVLVVDDLLATGGTIEATVKLI
11    KVISEEYSLEYGTDTIEMHVGAVE-PGERAIIIDDLIATGGTLAAAIRLL


          150         160         170         180
          |           |           |           |
       !* *  *!X    !!!X  X X! XX     ! **
1     GQLQAEVVECVSLVELTSLKGREKLGSVPFFSLLQYE------
2     SQLRAEVVECVSLVELTSLKGREKLGPVPFFSLLQYE------
3     HQLRAEVVECVSLVELTSLKGRERLGPIPFFSLLQYD------
4     GRLQAEVLECVSLVELTSLKGREKLAPVPFFSLLQYE------
5     RKVGGVVVESLVVMELVGLEGRKRLDGCKVHSLIKY-------
6     RQVGAHILEYDFVLVLDSLHGEEKLSA-PIFSILHS-------
7     EQLEANLLEYNFVMELDFLKGRSKLNA-PVFTLLNAQKEALKK
8     RRLGARVFEAAAIIDLPELGGSTRLQDAGISTFSLTAFALDER
9     RRMRANIHEAAAIIDLPELGGSQKLQDIGIPTFTLTAFALSDR
10    RRLGGEVADAAFIINLFDLGGEQRLEKQGITSYSLVPFPGH--
11    ERVGVKIVECACVIELPELKGKEKLGETSLFVLVKSAA-----
```

being an endogenous gene, constitutively transcribed (Thompson *et al.*, 1980; Carver *et al.*, 1980; Taylor *et al.*, 1985; Drobetsky *et al.*, 1987, 1995; Sage *et al.*, 1996). Figure 1.7 illustrates the extent of the mutations known at the protein level recovered in *aprt*. Few of these studies however, have been developed with a view to examining the effect of these mutations on the enzyme's properties, with the exception of a recent retrospective analysis by de Boer and Glickman (1991). It is generally assumed that resistance to adenine analogs must be accompanied by partial or total deficiency of APRT activity. This position is however disputable, as several publications on APRT as a model system, describe mutants resistant to the selective agent, which apparently still exhibit the ability to utilize adenine (Kalle and Gots, 1963; Turker and Martin, 1985; de Boer and Glickman, 1991; Khattar *et al.*, 1997). It may be possible to resolve this question by deliberately selecting mutants under conditions which require both the use of adenine, and resistance to one or more adenine analogs. In addition, it may be illuminating to test whether such molecules exhibit resistance to multiple analogues (*e.g.* does resistance to 8-azaadenine imply resistance to 2,6-diaminopurine), or can be localized to a functional domain or tract in the primary structure of the enzyme, which may mediate substrate specificity, such that some substrates would be excluded from the active site in the final structure of these mutant enzymes. A final consideration about *in vivo* mutagenesis in structure-function analysis, is that no prior assumptions about the role or significance of residues is required.

1.5.2 *In vitro* mutagenesis

In vitro mutagenesis (IVM) offers a convenient way to construct mutants at specific residues in order to test hypotheses about the role of individual residues in the

Figure 1.7 : Chinese hamster APRT sequence with mutants recovered in *in vivo* mutagenesis studies from our laboratory and others. Above the wild type sequence are indicated the amino acid alterations that were found in mutant collections. Asterisks denote termination mutants. Sources of data are described in text.



polypeptide. IVM using synthetic oligo-nucleotides was first developed by Zoller and Smith (1982). Current methods utilize oligonucleotides to produce specific or semi-randomized changes in DNA sequence, on templates prepared as either single-stranded DNA (ssDNA) plasmids or bacteriophage, double stranded DNA (dsDNA) plasmids, or PCR-amplified DNA (see Botstein and Shortle, 1985, for a review). Site-specific *in vitro* mutagenesis has numerous advantages, including absence of second site (adventitious) mutations, predictable amino acid changes, and the availability of kit-based commercial products. Site-specific IVM is limited by the need to clone or PCR amplify sequences of interest. IVM is however, not generally as useful in saturation mutagenesis, since there are as many as 3N (three times the number of nucleotides in the coding sequence) possible single mutations in the sequence of interest. For CHO *aprt*, this would require in excess of 1500 individual oligonucleotides, and is clearly cost prohibitive. Using partially randomized oligonucleotides, where multiple bases are incorporated at some positions during synthesis, such that more than one mutation may be generated for each reaction mixture, is one solution to this problem,

More recently, a novel approach to *in vitro* mutagenesis has been developed, wherein a template is mutagenized during PCR in conditions which reduce polymerase fidelity and promote base misincorporations (Leung *et al.*, 1989; Lin-Goerke *et al.*, 1997). Using this method, multiple mutations may be produced in a PCR-amplified fragment, which may be subsequently cloned into a suitable vector. This method produces mutations only in the PCR-amplified region, and can be very effective in terms of effort and cost.

1.6 Kinetic characterization of mutants

Using crude extracts from proteins expressed from suitable vectors in *E. coli*, it should be possible to rapidly determine a basal level of APRT activity of most mutants. If the protein of interest is expressed in a fully folded form (disregarding any requirements for post-translational processing), it is likely to be active, and its activity should be readily recoverable from cell extracts. This approach can be complemented by a simple screening of mutant clones on selective media, if the clones express a detectable phenotype.

From a fusion vector system, large quantities of even non-functional mutant proteins can be purified in a few steps to essential homogeneity. The advantages of this system are rapid purification, no intermediate assays of enzyme activity until the final purification step is completed, and mutant enzymes need not be functional in order to be tracked through purification. On the other hand, a significant time commitment to each mutant must be made beyond crude extract stages, and a rapid or *in vivo* assay of enzyme activity is not always possible.

If an expressed mutant enzyme has detectable activity in crude extracts, it may be possible to further purify the protein using the same approach as used for the normal enzyme. If the mutant protein has very poor stability relative to the normal protein, From such purified extracts, aliquots can be evaluated with respect to several biochemical parameters, including kinetic constants for substrates, and stability to heat, pH and ionic concentration. In practice, because of the need to perform assays on extracts on multiple days, extracts are often stored either refrigerated, or frozen. This leads to another crude measure of mutant protein function, stability to freeze-thaw or long term storage.

One liability of using protein extracts is that for undefined reasons, mutant enzymes may have significantly poorer stability under any specific extraction or storage condition. It is even possible that assay conditions, such as incubation at 37 °C, can quickly destroy protein activity, making comparative kinetic assays impossible. A possible solution to this problem is *in vivo* assessment of the effects of mutations in the protein of interest. Variations in growth conditions of the expressing cells, such as elevated or reduced incubation temperature, variation in ionic concentration of the medium, or addition of alternate substrates (*e.g.* analogs of adenine which produce cytotoxic nucleotides after processing by APRT) may be used (Parks *et al.*, 1973; Montgomery, 1982) . This approach requires expression of a functional protein product, which preferably can be continuously produced under the desired conditions. Where these conditions can be satisfied, *in vivo*, such an approach may be informative as to the role of specific changes in the sequence under investigation. *In vivo* activity can also be used to search for mutants which have a specific alteration in function, such as an ability to utilize a novel or variant substrate, or conversely, to become resistant to toxic substrates.

1.7 Relationship of structural to functional changes

The observation that many enzymes are not highly reactive to "classes" of substrates, but only to specific molecules within a class, delineates an important theoretical boundary in structure-function analysis (Popjak, 1970). Clearly, enzymes do not generally or randomly enhance reaction rates, but have a specific and stable catalytic site, which preferentially excludes or ignores some substrates, and accepts other ones. The observation that a chemical which is structurally similar to the substrate can interfere with

the rate of reaction in a competitive fashion while not being susceptible to the reaction *per se*, implies that access to the active site is controlled by structural components separate from the parts which actually perform the chemical work. Whatever the structural composition of the catalytic site, it has to be three-dimensional, and contain catalytic and noncatalytic groups. Such deductions arise from substrate affinity and kinetic studies even without any *a priori* knowledge of the nature of enzyme catalysts themselves.

Amino acid residues, and especially the side chains of such residues, provide the primary functional and structural determinants of proteins. The composite shape and chemical nature is not however a simple aggregate function, as if enzymes were some sort of complex solution. These composite molecules have definite structural and catalytic properties, different from and greater than a simple mixture of the components. It is these unique properties that are the subject of analysis of structure-function relationships.

The peptide bond is the uniform link connecting amino acids in a polypeptide. Two amino acids are linked at their respective amino and carboxyl ends, in a condensation reaction, with the removal of a molecule of water (Fersht, 1985). This bond is important in understanding why proteins are not a simple mixture of their components. The peptide linkage removes the oppositely charged amino and carboxyl ends of the amino acids from consideration as primary determinants of protein function. The resultant chain of amino acids has a polarity of structure, defined by the residuals of the amino and carboxylic acid groups, which is maintained irrespective of the number of peptide units in the entire chain. This linear polarity of structure is overlaid by higher orders of structure, including electric charge distribution (Desideri *et al.*, 1992; Antosiewicz *et al.*, 1996), and secondary

structures, such as α-helix or β-sheet (and Corey, 1951; Pauling *et al.*, 1951; Crick, 1953; Barlow and Thornton, 1988).

These higher orders of structure are essentially all determined by the side-chains of the amino acids in the polypeptide. The peptide bonds of the polypeptide are, with some exceptions, repetitive elements of essentially similar nature, and can't explain the diversity of structure and function observed among enzymes. Could a simple linear sequence of amino acids have sufficient uniqueness to account for the catalytic specificity of proteins? A series of ten amino acids residues will be much larger than the region of events around a given chemical reaction, say breaking a single bond. Including determining substrate specificity, such a reaction could only involve one or a few amino acid residues of such a linear sequence. With just 20 amino acids in most cells, this suggests a rather limited spectrum of catalytic functions, which is not at all the observation. But the polypeptide can flex or bend, driven by interactions between side chains of amino acids and the solvent, and bring more than one residue to bear upon the site of the reaction. This is really the essence of protein structure, and by extension, of the catalytic site.

It would be extremely convenient if we could simply obtain the sequence of our most interesting protein, apply computational analysis to the sequence, and obtain a good picture of the three-dimensional relationships of amino acids in the structure (Bowie and Sauer, 1989). Unfortunately, there appear to be few limitations on the shape, size, and residue content of the secondary structure or active site of proteins (Efimov, 1993). The best of current computerised methods have a predictive accuracy for secondary structure ranging from 50 to 65% for single proteins for which three-dimensional structures are

available (von Heijne, 1987; Cohen *et al.*, 1990). If we have multiple sequences, varying somewhat, but still performing similar structural or functional roles, the power of such approaches can be improved somewhat (Schulz *et al.*, 1986; Attwood *et al.*, 1991; Valencia *et al.*, 1991; Yokuma *et al.*, 1992; Pascarella and Argos, 1995).

Evaluating, based on empirical data, which residues are most often involved in helices versus sheets, has however, led to some useful tools for predicting secondary structure in protein molecules (Chothia *et al.*, 1977; Vishwanadhan *et al.*, 1991; Rost *et al.*, 1993; for a review see von Heijne, 1987). While the power of such methods is variable, general trends in the secondary structure of the protein in question may be seen. Unfortunately, the functional roles of alpha helices and beta sheets are unfixed. Alpha helices or beta sheets may be adventitious structural components, the scaffold for an active site, or participate in the formation of higher orders of structure by interaction with other proteins. A third, less well understood element of secondary structure is the "loop". Loops may act as structural elements which connectng important helix or sheet motifs, as in the nucleotide-binding fold predicted in phosphoribosyl transferases (Rossmann *et al.*, 1974; Argos *et al.*, 1983; Wierenga *et al.*, 1985; Bussetta , 1988). It may also form part of the catalytic site, as in the case of the phosphate binding loop of the family of nucleotide binding proteins which includes kinases, and certain oncogenes (Saraste *et al.*, 1990). Algorithms which predict the existence of helix or sheet, or turn or coil, all important structural elements, are limited in their direct utility for identifying functionally important regions or residues.

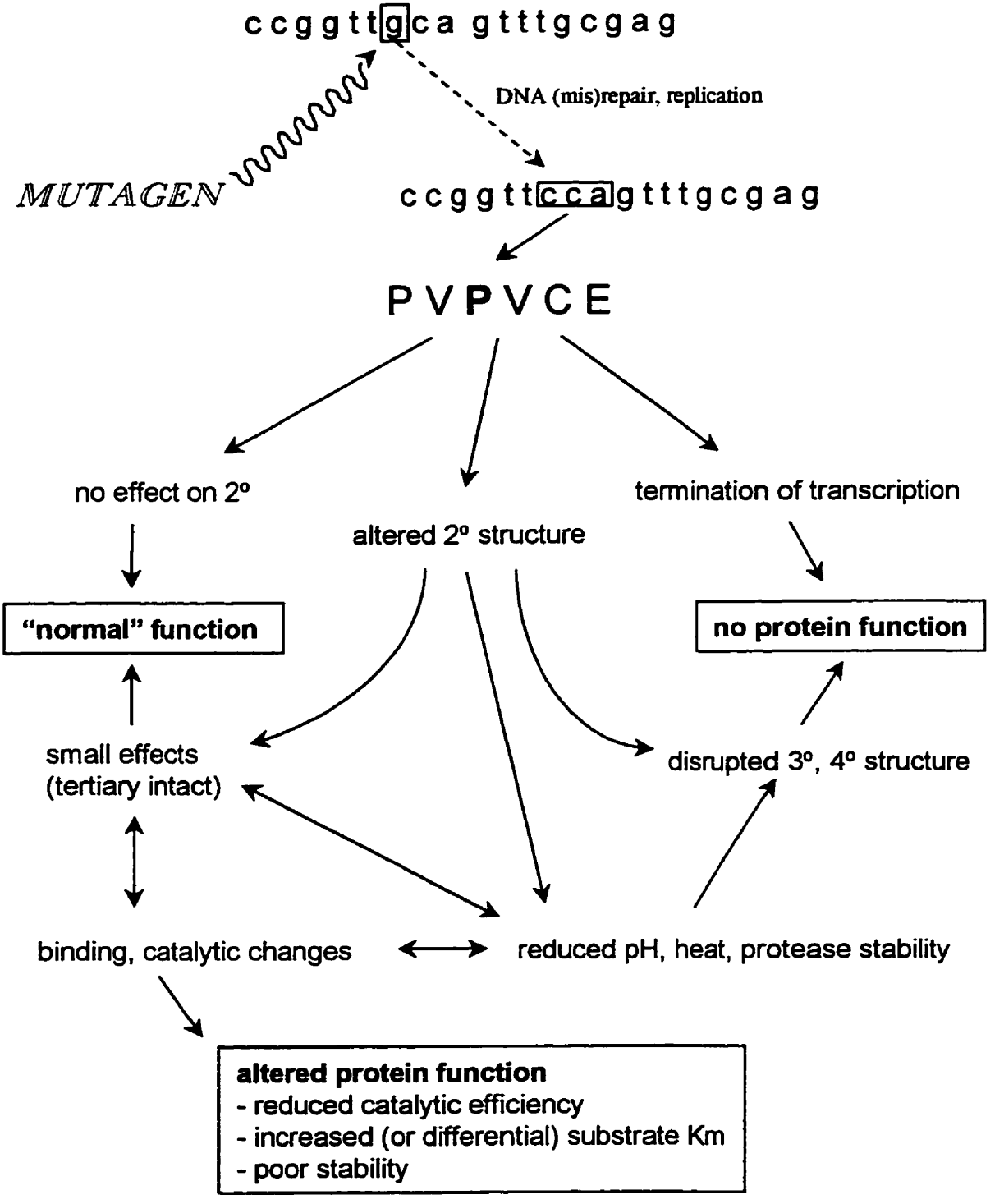
It is sometimes possible to reduce the work of finding structural elements

important in protein function by examining the three-dimensional model of the protein. However, to be certain of the effect that a given residue change will have on the molecule, remembering the limited predictive power of computerised methods, it is necessary to analyse a novel molecule with that change. To date, a few of the phosphoribosyl transferases have been crystallised and X-ray structures reported (Musick, 1977; Edwards *et al.*, 1988; Scapin *et al.*, 1995), but APRT has only recently been crystallised (Phillips *et al.*, 1996). No structure for this latter crystal from *Leishmania donovani* APRT has yet been reported. In many cases however, even solved crystal structures are not sufficient to predict the effects of amino acid substitutions on catalytic or structural parameters (Pai *et al.*, 1977; Berger and Evans, 1990).

An alternative means of probing the role of specific protein residues that is complementary to sequence analysis or structural prediction, involves recombinant DNA technology. A cloned DNA sequence, expressed in a suitable host, may be deliberately mutagenized, site-specifically or randomly, to generate novel sequences (Zoller and Smith, 1982). These new proteins may then be characterised with repsect to changes in binding affinity, catalytic competence, or stability. In turn, inferences about the likely role of the altered residue may be made (Carter and Wells, 1987; Sung *et al.*, 1988). For example, if a nonconservative change (such as substituting a charged for an uncharged amino acid side group) at a specific site causes no measurable change in binding affinity for a substrate, it is reasonable to argue that he residue is directly involved in the substrate binding site. The range of possible outcomes of mutagenic changes is summarised in Figure 1.8.

Obviously then, a few appropriate changes in amino acid sequence could be very

Figure 1.8 : Possible outcomes of mutational event, illustrating the interaction between structural and functional outcomes. The original theoretical peptide primary sequence across the region shown was PVAVCE. Only direct effects on gene product structure or function are considered. Alterations in cellular distribution, post-translational modifications etc. are not illustrated.



informative provided assays to measure potentially altered enzyme properties are available. Generally, any residue change which causes gross structural changes in the enzyme is unlikely to leave the enzyme with significant residual function. This loss of activity may be due to global or localized structural disruption, discrete loss of binding or catalytic function, or a combination of these. Subtle overall changes in structure, which preserve enzyme activity, may manifest as differential sensitivity to heat, pH, or ionic strength. Changes in catalytic domains may present as specific kinetic changes, such as shifts in apparent K_M for substrates, modulation of substrate specificity, or alterations in inhibition characteristics.

Changes in quantity or quality of enzyme activity, detectable *in vitro*, may not correspond to the apparent effects of the same mutation on the enzyme *in vivo*. Some mutants which are identified as nonfunctional by phenotype or enzyme assay from native systems (*e.g.* APRT expressed in CHO cells in culture), may have considerable function in the heterologous expression system. The basis for these differences, often ascribed to proteolytic susceptibility, or codon bias during translation, is unclear. To some degree this is a criticism of *in vitro* manipulation (*e.g.* heterologous expression) as a tool for probing structure- function in proteins. Residue substitutions which yield loss of function in the normal environment for a given protein, may be meaningful than the same changes in an *in vitro* system where some activity can still be detected.

1.8 Objectives

Despite an extensive history of biochemical analysis, much of the nature of structure and function in nucleotide binding proteins remains obscure. Recent NMR

work on adenylate kinase (AK) for example, which had previously been co-crystallized with its substrates by several invesitgators, has shown that the bound substrate has a different relationship to the enzyme in solution from that in the crystal (Pai *et al.*, 1977; Tian *et al.*, 1990; Vetter *et al.*, 1990). Further, investigation on mutants of adenylate kinase with altered biochemical properties have been informative regarding elements of the catalytic event, for example the position of the nucleotide in the active site, not apparent from crystal structure models (Reinstein *et al.*, 1988; Reinstein *et al.*, 1990; Dahnke *et al.*, 1991) . Thus it seems that for nucleotide-binding enzymes which have already been characterized extensively, studies of the characteristics of mutants, comparative sequence analysis, and crystallography may be complementary.

In the absence of an entire suite of perfect analytical tools, or good three-dimensional crystal or NMR structural data, inferences about structure and function must come from alternative methods. This is very much the case with adenine phosphoribosyl transferase. The crystal structures of some related proteins are available, but to date, few of the APRT lineage have been successfully crystallized, and none have been analyzed by X-ray crystallography. Thus if we are to expand the knowledge base about structure/ function in APRT, we must infer the role of sequence or predicted structural motifs by comparison to related molecules, and even to the structure of the substrates themselves. Analysing the effects on function or structure of specific mutants, or selecting mutants with unique properties, may be informative about the roles of altered residues in the final protein. The clinical, cellular, and pharmacological relevance of APRT, along with its potential as a model for nucleotide-binding proteins generally, make it an important

candidate for analysis from the point of view of structure/function relationships. Little data on structure-function relationships studied by other approaches in this family of enzymes, has been published. The objective of the work described herein is, in summary, to build a more coherent picture of structure-function realtionships in APRT, and to attempt to synthesize available biochemical, sequence analysis, and mutational data on this protein.

Chapter 2 Sequence analysis of phosphoribosyl transferases

2.1 Introduction

As a first step in understanding structure function relationships in a protein of unsolved three dimensional structure, a useful approach is to examine related proteins. In the case of the phosphoribosyl transferases, there are multiple sequences available, representing an assortment of substrate specificities, reaction schemes, and of course, source with regard to organism. Phosphoribosyl transferases share certain commonalties of function, and these relationships should be reflected in similarities or identities of sequence (Doolittle, 1981; Pascarella and Argos, 1995) . Indeed, it is an underlying assumption of structure function studies that sequence similarities or identity in proteins of related function, and probable common, but ancient evolutionary lineage, likely represent homologous sequences (Dayhoff, 1978; Rossman *et al.*, 1974). Such homologies would be expected to persist if selective pressure works to maintain them. Since such sequences are susceptible to selective pressures, changes in those sequences ought to be reflected at a biochemical level. In contrast, changes at sites which are biochemically less relevant ought not to be under strong selective pressure, and thus should accumulate over time as differences between evolutionary lineages (Ohta, 1992). By extension, sequence tracts in related proteins which undergo extensive change would not be expected to lie within regions of essential structural or functional elements. The foregoing assumes that mutations in homologous or identical sequences have equivalent biochemical effects between cellular systems, which while reasonable, remains unproven.

One opportunity presented by protein sequence analysis involves probing the

assumptions of comparative sequence analysis. Reconstruction of phylogenetic relationships between species and Kingdoms has been carried out for many genes and proteins including rodent APRT sequences (Fieldhouse *et al.*, 1997; Hashimoto *et al.*, 1997; Kuma *et al.*, 1997). It is likely that previously reported evolutionary relationships of species will be confirmed when their phylogeny is reconstructed using adenine phosphoribosyl transferase sequences. But if we compare sequences of functionally related families of proteins such as phosphoribosyl transferases, it may be possible to discern additional relationships between the various sequences. For example, if one hypothesizes that APRT, HGPRT, and OPRT arose from a common ancestral sequence, in an ancient cell, then we might expect that any substrate-specific set of sequences would reproduce a conventional phylogeny of species relationships, consistent with the origin and history of the molecules. But a combined phylogeny of different substrate-specific sequences would yield a sorting of sequences by substrate, since the sequences will not be expected to have been interacting across species boundaries, and thus should not substantially influence each other's evolutionary history. On the other hand, if a gene arose after a given ancestral division in response to different pressures, that sequence may sort within its species relationships rather than with its substrate cousins. It is entirely likely that some combination of such patterns will emerge from phylogenetic analysis of phosphoribosyl transferases.

In addition to primary structure analysis, folded or secondary structural information may be useful. Although such analyses are imperfect (Taylor, 1992; Frishman and Argos, 1997;) they can be informative. In addition, when multiple functionally related proteins

with similar sequences and presumptively similar structures are examined in parallel, it may be possible to enhance the power of such analyses (Rao and Rossman, 1973; Geourjon and Deleage, 1997; Persson and Argos, 1997) . Many powerful algorithms exist that are useful for these studies, and several have been implemented in commercially packaged computer programs. As an adjunct to these packages, remote software servers can provide similar analyses for comparison (Smith *et al.*, 1997).

2.2 Materials and Methods

2.2.1 Sequence data for phosphoribosyl transferases

Sequences of phosphoribosyl transferases and other proteins used in the analyses were obtained from Internet search and retrieval servers at the National Center for Biotechnology Information http://www. ncbi.nlm.nih.gov, or the Human Genome Center at Baylor College of Medicine (http://kiwi.imgen.bcm.tmc.edu: 8088/search-launcher/launcher.html). In Table 2.1, the names, accession numbers, and species codes (see subsequent figures) of the phosphoribosyl transferase sequences used are shown.

2.2.2 Global alignments and motifs

Prior to alignment, sequences were converted to minimal text listings to permit manipulation by multiple programs, adaptations of the Needleman-Wunsch algoritm for sequence alignment (1969). These simplified files were then used in sequence alignment programs for Microsoft Windows including MEGALIGN (DNASTAR), Clustal V and Clustal X (Higgins, 1997), and map3 for UNIX (Huang, 1997). Alignments were assessed for global alignment quality by examination, and refined in some cases by hand.

Table 2.1 (1 of 3) : Sequence data by species and accession number. Multiple accessions for a complete sequence are also listed. Key to source databases : pir; Protein Information Resource, gb; GenBank, embl; European Moelcular Biology Laboratory, sp; Swiss Prot, ent Entrez . All of these databases are indexed nonredundantly at the National Center for Biotechnology Information (http://http://www.ncbi.nlm.nih.gov/Entrez/protein.html). Also indicated are the codes used to indentify species in subsequent figures (Fig 2,7, 2.8).

Species	Source Accession		Species Code
APRT			
Cricetulus griseus	pir S36334 gb M25901		cho ut1
Mus musculus	sp P08030 gb M11310		mus
Mus spicilegus	gb U28720		spi
Mus pahari	gb U28721		pah
Rattus norvegicus	sp P36972 gb 543829	gb L04970	rat
Gerbillus campestris	gb U28961		gerb
Stochomys longicaudatus	gb U28723 sp P47958		stoch
Mastomys hildebrantii	sp Q64427 gb U28722		hil
Homo sapiens	sp P07741		hum
Drosophila melanogaster	sp P12426 gb L06280		drosm
Drosophila pseduobscura	sp P54363		drosb
Saccharomyces cerevisiae I	sp P36973 gb 542830		sacci
Saccharomyces cerevisiae II	sp P49435		saccii
Arabadopsis thaliana	sp P31166		arab
Triticum aestivum	gb U22442		trit
Synechocystis sp.	ent 1653085		synec
Escherichia coli	gb M14040 sp P07672		eco
Pseudomonas aeruginosa	sp Q04633		pseaer
Pseudomonas stutzeri	pir S37398 gb 480812		psst
Haemophilus influenzae	sp P43856		haem
Streptomyces coelicolor	gb 1076066		scoe
Mycobacterium tuberculosis	gb Q50637 gb 1478211		myct
Mycobacterium lepriae	pir S72724		mycl
Caenorhabditis elegans	gb U80438		eleg
Leishmania donovanii	gb 159398		leish
Mycoplasma pneumoniae	ent 1674130 sp P75388		mycop
Mycoplasma genitalium (SGC3)	sp E64230 sp P47518		mycog
Methanococcus jannaschii	gb Q59049 gb 1592235		meth
HGPRT / HPRT			
Vibrio harveyi	sp P18134		vibha
Escherichia coli	sp P36766		eco
Rhodobacter capsulatus	sp P37171		rhodo
Lactococcus lactis	sp Q02522		lacto
Bacillus subtilis	sp P37472		bacsu
Trypanosoma cruzi	gb 386724		trypc
Trypanosoma brucei brucei	sp Q07010		trypb
Leishmania donovani	gb 409144		leish
Tritrichomonas foetus	gb 406187		tritf

Table 2.1 (2 of 3) :

Species	Source Accession	Species Code
HGPRT / HPRT		
Plasmodium falciparum (iso G)	sp P20035	pfal1
Plasmodium falciparum (iso K)	sp P07833	pfal2
Plasmodium falciparum	gb160323	pfal
Toxoplasma gondii I	gb 499108 spQ26997 gb 501059	toxoi
Toxoplasma gondii II	gb 501060	toxoii
Schistosoma mansoni I	pir S09614	sman
Schistosoma mansoni II	sp P09383	sman2
Rattus norvegicus	ent 2304983 ent 111792 sp P27605	rat
Mycobacterium tuberculosis	embl Z95436	mycot
Salmonella typhimurium	gb AF008931	salty
Cricetulus longicaudatus	ent 817936	cho
Mus spretus	sp Q64531	mus2
Mus musculus	sp P00493	mus
Meriones unguiculatus	sp P47959	mer
Haemophilus influenzae	sp P45078	hae
Bacillus subtilis II	sp P37472	bacsu
Vibrio harveyi	sp P18134	vibha
Homo sapiens	sp P00492	hum
Cricetulus griseus	sp P00494	cho
XPRT		
Escherichia coli	K01784	eco
UPRT		
Saccharomyces cerevisiae	sp P18562	sacc
Toxoplasma gondii	gb 499106	toxo
OPRT		
Saccharomyces cerevisiae URA10	sp P30402	sacc
Saccharomyces cerevisiae URA5	sp P13298	sac5
Podospora anserina	sp A29459 sp P08309	podan
Colletotrichum graminicola	sp S30118 sp P35788	coll
Sordaria macrospora	sp P18904	sord
Dictyostelium discoides	sp S03826 sp P09556	ddis
Homo sapiens	sp A30148	hum
Lactobacillus plantarum	sp A60993 sp P77889	lacto
Rhizobium leguminosarum	sp P42719	rhiz
Escherichia coli	sp P00495	eco
Bacillus subtilis	sp P25972	bacsu
UMP synthase		
Drosophila melanogaster	pir JU0141	dros
Bos taurus	pir JN0558	bos

Table 2.1 (3 of 3) :

Species	Source Accession	Species Code
PRPP Synthetase (PRPPSYN)		
Rattus norvegicus I	sp P09330	rati
Rattus norvegicus II	gb 206428	ratii
Homo sapiens	sp P11908	eleg
Caenorhabditis elegans	gb 459000	hum
Saccharomyces cerevisiae I	sp P38620	sacci
Saccharomyces cerevisiae II	sp P38063	saccii
Escherichia coli	sp P08330	eco
Salmonella typhimurium	sp P15849	salt
Bacillus subtilis	sp P14193	bacsu
Synechocystis sp.	pir JN0886	synec
Leishmania donovanii	gb 159398	leish
Glutamine amido PRT (GLNPRT)		
Homo sapiens	gb 548638	hum
chicken	sp P12873	chik
Anthranilate PRT (ANTHPRT)		
Azospirillum brasilense	sp P26924	asp
Haloferax volcanii	gb 149037	halo
Quinolinate PRT (QNPRT)		
Salmonella typhimurium	sp P30012 gb 266599	salt

2.2.3 Phylogeny reconstruction

Aligned phosphoribosyl transferase sequences were used mainly in two phylogenetic algorithms, Clustal V (or X) (Higgins, 1997; Thompson *et al.*,1997) and MEGALIGN (DNASTAR). Subset regions of high sequence similarity, and their inverses, were analyzed as well as globally aligned sequences. Finally, selected phosphoribosyl transferases from different Kingdoms were analyzed to determine whether the phylogeny of phosphoribosyl transferases would reconstruct the gross phylogenetic pattern of these disparate groups.

2.2.4 Secondary structure prediction

Secondary structure motifs were evaluated by PROTEAN (DNASTAR), using the algorithms of Garnier and Robson (1978), and Chou and Fasman (1978a,b). Summary results were saved as graphical files.

2.3 Results

2.3.1 Alignment of phosphoribosyl transferases

The final alignments of APRT sequences are shown in Figure 2.1. Amino acid residues of strong similarity are indicated by the symbols along the top of the figure. In addition, alignments of each substrate-specific group of phosphoribosyl transferases to representative APRT sequences is shown in Figure 2.2. Represented are several enzymes related in function to APRT, such as HPRT, OPRT, XPRT, UPRT, and some the phosphoribosyl transferases of *de novo* purine or amino acid synthesis. Again, the most conserved tracts are indicated as noted in the figure legend.

Figure 2.1 (1 of 5): Aligned primary sequences of adenine phosphoribosyl transferases from various organisms. Organisms are arranged top to bottom: mammal (9), insect (1), fungus (2), plant (2), cyanobacteria (1), eubacteria (8), nematode (2), mycoplasma (2), archeabacteria(1). The numbering scheme is that of Chinese hamster (*C.longicaudus*) APRT. Sources of the data (Genbank, PIR,SwissProt) in the figure are described in the text. From sequence 21, *Mycobacterium tuberculosis*, 39 residues are unalignable near the N-terminus, and these are listed below as indicated. Sequence 23, *Mycobacterium lepriae* is incomplete. Identical residues (<24/28) are indicated by **x** over the sequence, highly conserved tracts (>15/28 identical) by a *. Moderately conserved regions (<8/28 identical, <10/28 conservative) are indicated by !.

Sequence	Species
1	*Cricetulus longicaudatus*
2	*Mus musculus*
3	*Mus pahari*
4	*Mus spicelegus*
5	*Rattus norvegicus*
6	*Gerbillus campestris*
7	*Stochomys longicaudatus*
8	*Mastomys hildebrantii*
9	*Homo sapiens*
10	*Drosophila melanogaster*
11	*Drosophila pseudoobscura*
12	*Saccharomyces cerevisiae I*
13	*Saccharomyces cerevisiae II*
14	*Arabadopsis thaliana*
15	*Triticum aestivum*
16	*Synechocystis sp.*
17	*Escherichia coli*
18	*Pseudomonas aerogenes*
19	*Pseudomonas stuzeri*
20	*Haemophilus influenzae*
21	*Streptomyces coelicor*
22	*Mycobacterium tuberculosis*
23	*Mycobacterium lepriae*
24	*Caenorhabditis elegans*
25	*Leishmania donovani*
26	*Mycoplasma pneumoniae*
27	*Mycoplasma genitalium*
28	*Methanococcus jannaschii*

Figure 2.1 (cont. 2 of 5)

```
        1                 10            20            30           40
        |                !|    *!  !!**x  *x!!**x*!!!!   x     !             !!
1       MA-----ESELQLVAQRIRSFPDFPIPGVLFRDISPLLKDPASFRAS-IRL---LASH
2       MS-----EPELKLVARRIRVFPDFPIPGVLFRDISPLLKDPDSFRAS-IRL---LASH
3       MS-----ESELKLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRAS-IRL---LASH
4       MS-----EPELKLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRAS-IRL---LASH
5       MS-----ESELQLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRA-SIRL---LAGH
6       MA-----EPELQLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRAS-IRL---LANH
7       MS-----EPELQLVARRIRSFPDFPVQGVLFRDISPLLKDPVSFRAS-IHL---LASH
8       MS-----EPELQLVARRIRSFPDFPIPGVLFRDISPLLKDPDSFRAS-IRL---LASH
9       MA-----ESDLQLVEQRIRSFPDFPTPGVVFRDISPVLKDPASFRAA-IGL---LARH
10      MSPSISAEDKLDYVKSKIGEYPNFPKEGILFRDIFGALTDPKACVYL-RDL---LVDH
11      MS-QVSAEDKLDYVKSKIGEYPNFPKEGILFRDIFGALTDPKACVYL-RDL---LVQY
12      M--SIS-ESYAKEIKTAFRQFTDFPIEGEQFEDFLPIIGNPTLFQKL-VHT---FKTH
13      M--SI--ASYAQELKLALHQYPNFPSEGILFEDFLPIFRNPGLFQKL-IDA---FKLH
14      MATEDVQDPRIAKIASSIRVIPDFPKPGIMFQDITTLLLDTEAFKDT-IAL---FVDR
15      MAS----DGRVERIASSIRAIPNFPKPGILFQDITTLLLDPQAFRDT-TDL---FVER
16      MD-----------LKALIRDIPDFPKPGIMFRDITTLLNSPEGLRYT-IDS---LVEQ
17      MT---ATAQQLEYLKNSIKSIQDYPKPGILFRDVTSLLEDPKAYALS-IDL---LVER
18      M------IFDEFTLKSQIRAVPDFPKPGVVFRDITPLFQSPRALRMTVDSF----VQR
19      M------IFDEFSIKTLIRPVQDFPRPGVVFRDITPLFQSPKALRMVADSL----IQR
20      MT------TQLDLIKSSIKSIPNYPKEGIIFRDITTLLEVPAAFKAT-IDL----IVE
21      MT---EPTGITELLLSRIRDVADYPEPGVVFKDITPLLADPGAF-AALTDA----LAE
22      MCHGG*---------SRRSSALTPLFA---DRRGLAAVTEALADRASGADLVAGVAIS
23      M---------RPISRRQPSSSRTLLRCSLTDRRW------PAVIDAL-ADV----AAG
24      MT-LPRFDQIRPKIEQHIREVKDFPKKGINFRDIMPLFTNPQLVNELCVVI----ADH
25      MPFKEVSPNSFL-LDDS-HALSQLLKK--SYRWYSPVFS-PRNVPRFADVSSITESPE
26      MAQ---------------FIQEV-KADGIVC---------------------------
27      MDQ------NFKLLDQAIKRFENFPNQGTLFYDITPVFSNPQLFNFVLTQM---AQ--
28      MLLEETLKSCP-IVKRGE--YHYFIHPISDGV---PVV-EPKLLREVATRI---IK--
```

• 22 Unalignable *TWAGDYVLNVIATGLSLKARGKRRRQRWVDDGRVLALGE*

Figure 2.1 (cont. 3 of 5)

```
     50         60                        70         80         90
                                  !!* *!xxxxx*!x*x  !!!* *!!xx * xxx*
1   -LKSTHGGKIDY-------------IAGLDSRGFLFGPSLAQELGLGCVLIRKRG-KLPGP
2   -LKSTHSGKIDY-------------IAGLDSRGFLFGPSLAQELGVGCVLIRKQG-KLPGP
3   -LKSTHSGKIDY-------------IAGLDSRGFLFGPSLAQELGVGCVLIRKQG-KLPGP
4   -LKSTHSGKIDY-------------IAGLDSRGFLFGPSLAQELGVGCVLIRKQG-KLPGP
5   -LKSTHGGKIDY-------------IAGLDSRGFLFGPSLAQELGVGCVLIRKRG-KLPGP
6   -LKSKHGGKIDY-------------IAGLDSRGFLFGPSLAQELGLGCVLIRKRG-KLPGP
7   -LKSTHSGKIDY-------------IAGLDSRGFLFGPSLAQELGVGCVLIRKRG-KLPGP
8   -LKSLHGGKIDY-------------IAGLDSRGFLFGPSLAQELGVGCVLIRKRG-KLPGP
9   -LKATHGGRIDY-------------IAGLDSRGFLFGPSLAQELGLGCVLIRKRG-KLPGP
10  -IRES-APEAEI-------------IVGLDSRGFLFNLLIATELGLGCAPIRKKG-KLAGE
11  -IRQSQP-EVEV-------------IVGLDSRGFLFNLLLATELGVGYTPIRKKG-KLAGE
12  LEEKFAKEKIDF-------------IAGIEARGLLFGPSLALALGVGFVPIRRVG-KLPGE
13  LEEAFPEVKIDY-------------IVGLESRGFLFGPTLALALGVGFVPVRKAG-KLPGE
14  -YKDK---GISV-------------VAGVEARGFIFGPPIALAIGAKFVPMRKPK-KLPGK
15  -YKDK---DITV-------------VAGVEARGFIFGPPIALAIGAKFVPIRKPK-KLPGE
16  CESQEL--VPDH-------------VVGMESRGFLFGMPLAYQMNAGFIPVRKPG-KLPAP
17  -YKN---AGITK-------------VVGTEARGFLFGAPVALGLGVGFVPVRKPG-KLPRE
18  -YIE---ADFSH-------------IGAMDARGFLIGSAVAYALNKPLVLFRKQG-KLPAD
19  -YVEA---DFTH-------------IGALDARGFLVGSILAYELNKPLVLFRKQG-KLPAD
20  QYRD---KGITK-------------VLGTESRGFIFGAPVALALGLPFELVRKPK-KLPRE
21  AAGR---TGATK-------------VVGLEARGFILGAPVALRAGLGFIPVRKAG-KLPGA
22  AISVA---SLTRDVADFPVPGVV--EFKDDARGFLVAAAVATRLEVGVLAVRKGG-KLPRP
23  A-DL---------------------VTGLEFRGTMVVAATAARLDIGVLAIRKSG-KLPLS
24  -VRH---TVGHVDS-----------VAGLEARGFLFGPQVAIQLGVPFVPIRKKG-KLPGA
25  TLK---AIRDFLVQRYRAMSPAPTHILGFDARGFLFGPMIAVELEIPFVLMRKAD-KNAGL
26  ----------------------------PEARGFIFGGALASKTKLPLVLVRKPH-KLSGE
27  FIK---AINEAEA------------IVCPEARGFIFGGALASKTQLPLVLVRKAN-KLPGQ
28  IGD---FEGATK-------------LVTAEAMGIPLVTTLSLYTDIPYVIMRKREYKLPGE
```

Figure 2.1 (cont. 4 of 5)

```
        100          110          120          130          140          150
      ! * !*  !!!    !!!! !!!   !*! ****xxxxxxxxx! !!  x*  !   !!
1   TV-SASY-ALEYGKAE-LEIQKDALE--PGQ-KVVVVDDLLATGGTMCAACELLGQLQAEVV
2   TV-SASY-SLEYGKAE-LEIQKDALE--PGQ-RVVIVDDLLATGGTMFAACDLLHQLRAEVV
3   TI-SASY-ALEYGKAE-LEIQKDALE--PGQ-RVVIVDDLLATGGTMFAACDLLHQLRAEVV
4   TV-SASY-SLEYGKAE-LEIQKDALE--PGQ-RVVIVDDLLATGGTMFAACDLLHQLRAEVV
5   TV-SASY-SLEYGKAE-LEIQKDALE--PGQ-KVVIVDDLLATGGTMCAACELLSQLRAEVV
6   TV-SASY-ALEYGKAE-LEIQKDALE--PGQ-KVVIVDDLLATGGTMCAACQLLGQLRAEVV
7   TL-SASY-ALEYGKAE-LEIQKDALE--PGQ-RVVIVDDLLATGGTMCAACELLNQLRAEVV
8   TL-SASY-ALEYGKAE-LEIQKDALE--PGQ-RVVIVDDLLATGGTMCAACELLNQLRAEVV
9   TL-WASY-SLEYGKAE-LEIQKDALE--PGQ-RVVVVDDLLATGGTMNAACELLGRLQAEVL
10  VV-SVEY-KLEYGSDT-FELQKSAIK--PGQ-KVVVVDDLLATGGSLVAATELIRKVGGVVV
11  VV-SVEY-QLEYGSDT-FELQRTAIQ--PGQ-KVVIVDDLLATGGSLLAASELVRKVGGVVL
12  CA-SITFTKLDHEEI--FEMQVEAIP--FDS-NVVVVDDVLATGGTAYAAGDLIRQESLVVV
13  CF-KATY-EKEYGSDL-FEIQKNAIP--AGS-NVIIVDDIIATGGSAAAAGELVEQVGAHIL
14  VI-SEEY-SLEYGTDT-IEMHVGAVE--PGE-RAIIIDDLIATGGTLAAAIRLLERVGVKIV
15  VI-SEEY-SLEYGTDK-IEMHVGAVQ--PND-RVLIVDDLIATGGTLCAAAKLIERVGAKVV
16  VH-RVEY-DLEYGKDS-LEIHQDAVA--PHH-RVLIVDDLIATGGTAKATAELLTKLGCEVL
17  TI-SETY-DLEYGTDQ-LEIHVDAIK--PGD-KVLVVDDLLATGGTIEATVKLIRRLGG-EV
18  VL-AEGY-QTEYGEAF-LEVHADSLC--EGD-SVLIFDDLIATGGTLLAAASLVRRLGARVF
19  VL-SQAY-STEYGEAH-LEIHADSLC--EGD-SVLLFDDLIATGGTLLAAAQLVRRMRANIH
20  TI-SQSY-QLEYGQDT-LEMHVDAIS--EGD-NVLIIDDLLATGGTVEATVKLVQRLGGA-V
21  TL-SQAY-DLEVSS---AEIEVHAEDLTAGD-RVLVVDDVLATGGTAKASLELIRRAG-AEV
22  VL-SEEY-YRAYGAAT-LEILAEGIE-VAGR-RVVIIDDVLATGGTIGATRRLLERGG-ANV
23  VL-SEDY-IME. . . . .
24  TI-EASY-VKEYGEDR-VEIQEGAIK--NGD-IVFLIDDLLATGGTLRAATDLVVKAGGKVG
25  LIRSEPYEKEY-KEAAPEVMTIRYGSIGKGS-RVVLIDDVLATGGTALSGLQLVEASDAVVV
26  LA-RETY-DLEYRQNSILEMRVDALENCK---RCVIVDDLLATAGTVAAIDKLIARLG SQT
27  LI-SASY-DLEYRKHAVLEMSTTSLIQANNAKRCVIVDDVLATAGTVAAIDQLLKQLN-GET
28  IP--VFQSTGYSKG-QLYLNGIE-----KGD-KVVIIDDVISTGGTMIAIIDALKRAG-AEI
```

Figure 2.1 (cont. 5 of 5)

```
                160             170         180
     *!!  !*         xx  **
1    ECVSLV-ELTSLKGREKLG----SVPFFSLLQYE
2    ECVSLV-ELTSLKGRERLG----PIPFFSLLQYD
3    ECVSLV-ELTSLKGRERLG----PIPFFSLLQYD
4    ECVSLV-ELTSLKGRERLG----PIPFFSLLQYD
5    ECVSLV-ELTSLKGREKLG----PVPFFSLLQYE
6    ECVSLV-ELTSLKGREKLG----PVPFFSLLQYE
7    ECVSLV-ELTSLKGRERLG----PIPYFSLLQYE
8    ECVSLV-ELTSLKGRERLG----PIPFFSLLQYD
9    ECVSLV-ELTSLKGREKLA----PVPFFSLLQYE
10   MELVGLE------GRKRLDGKVHSLIKY
11   ESLVVMELVGLDGRKKLDCKVHSLIKY
12   EYD-FVLVLDSLHGEEKLSAPIFSILHS
13   LEANLLEYN-FVMELDFLKGRSKLNAPVFTLLNAQKEALKK
14   ECA-CVIELPELKGKEKLGETSLFVLVKSAA
15   ECA-CVIELPELKGRDKLGDMPVFVLVQADESV
16   GFA-FIIELAALNGRQCLPDLPIISLVEY
17   ADAAFIINLFDLGGEQRLEKQGITSYSLVPFPGH
18   EAA-AIIDLPELGGSTRLQDAGISTFSLTAFALDER
19   EAA-AIIDLPELGGSQKLQDIGIPTFTLTAFALSDR
20   KHAAFVINLPELGGEKRLNNLGVDCYTLVNFEGH
21   AGLAVLMELGFLGGRARLEPALAGAPLEALLTV
22   AGAAVVVELAGLSGRAALAPLPVHSLSRL
23
24   EAFVLIELAPLN-GRSKLPDVNLTTLISYDSA
25   EMVSILSIPFLKAAEKIHSTANSRYKDIKFISLLSDDAL*
26   VGYCFLIELQKLHGKAKLQPNVATKILLHY
27   VGYCFLIELKKLNGKAKLQPNVVSKILLHY
28   KDIICVIERGE--GKKIVEEKTGYKIKTLVKIDVVDGKVVIL

*25  TEENCGDSKNYTGPRVLSCGDVLAEHPH
```

Figure 2.2 (1 of 7) : Exhaustive alignment of phosphoribosyl transferase sequences (obtained from Genbank, SwissProt and PIR as described in text). Pairwise alignments by Clustal V Multiple Sequence Alignment (Gap penalty 20, Floating gap penalty 20, PAM 250; Higgins and Sharp, 1989), followed by multiple sequence assembly and manual adjustment. Identical residues (>50/58) are indicated by **x** over the sequence, highly conserved tracts (>25/58 identical) by a *. Moderately conserved regions (<20/58 identical, <20/58 conservative) are indicated by !. Gaps are not counted. Note that some sequences are abbreviated in unalignable regions, denoted by * and footnotes. Numbering scheme is that of Chinese hamster APRT.

Figure 2.2 (cont. 2 of 7) : Exhaustive alignment of phosphoribosyl transferase sequences.

```
                                  1                        10          20
                                                             !!!        *
1  Cricet. longicaudus  APRT   MA-----ESE--------------LQLVAQRIRS-FPDF-----P
2  Mus musculus                MS-----EPE--------------LKLVARRIRV-FPDF-----P
3  Mus pahari                  MS-----ESE--------------LKLVARRIRS-FPDF-----P
4  Mus spicelegus              MS-----EPE--------------LKLVARRIRS-FPDF-----P
5  Stochomys longicaudatus     MS-----EPE--------------LQLVARRIRS-FPDF-----P
6  Rattus norvegicus           MS-----ESE--------------LQLVARRIRS-FPDF-----P
7  Gerbillus campestris        MA-----EPE--------------LQLVARRIRS-FPDF-----P
8  Mastomys hildebrantii       MS-----EPE--------------LQLVARRIRS-FPDF-----P
9  Homo sapiens                M-----AESD--------------LQLVEQRIRS-FPDF-----P
10 Streptomyces coelicor       MT-----EPTGI------------TELLLSRIRD-VADY-----P
11 Sacc. cerevisiae I          M---SISESY--------------AKEIKTAFRQ-FTDF-----P
12 Sacc. cerevisiae II         M---SISESY--------------AKEIKTAFRQ-FTDF-----P
13 Triticum aestivum           MAS----DGR--------------VERIASSIRA-IPNF-----P
14 Arabadopsis thaliana        MATEDVQDPR--------------IAKIASSIRV-IPDF-----P
15 Drosoph. pseudoobscura      MSQ-VSAEDK--------------LDYVKSKIGE-YPNF-----P
16 Drosoph. melanogaster       MSPSISAEDK--------------LDYVKSKIGE-YPNF-----P
17 Caenorhabditis elegans      MT-LPRFDQI--------------RPKIEQHIRE-VKDF-----P
18 Leishmania donovani         MPFKEV-SPN--------------SFL-LDDSHA-LSQL-----L
19 Mycoplasma pneumoniae       MAQ-------------------------------FIQE-----V
20 Mycoplasma genitalium       MDQ---------------------NFKLLDQAIKRFENF-----P
21 Synechocystis sp.           MD-------------------------LKALIRD-IPDF-----P
22 Escherichia coli            MTATAQQ-----------------LEYLKNSIKS-IQDY-----P
23 Pseudomonas aeruginosa      M------IFD--------------EFTLKSQIRA-VPDF-----P
24 Pseudomonas stuzeri         M------IFD--------------EFSIKTLIRP-VQDF-----P
25 Heamophilus.influenzae      MT------TQ--------------LDLIKSSIKS-IPNY-----P
26 Mycobact. tuberculosis      MCHGG*------------------GESRRSSAL--TPLFAD----
27 Methano. jannaschii         MLLEET------------------LKSCPIVKRG-EYHYF----I
28 Mastom. unguiculatus HPRT   MAT-RSPSI------------------VIGDDEPGYDLDLFC-IP
29 Mus musculus                MPT-RSPSV------------------VISDDEPGYDLDLFC-IP
30 Homo sapiens                MAT-RSPGV------------------VISDDEPGYDLDLFC-IP
31 Cricet. longicaudus         MAT-RSPSV------------------VISDDEPGYDLDLFC-IP
32 Schistosoma mansonii        MSSNMIKADCV----------------VIEDSFRGFPEYFCT-SP
33 Plasmodium falciparum       MPIPNNPGA------GENAFDP-----VFVKDDDGYDLDSFM-IP
34 Trypanosoma cruzi           MPREYEFA--EKILFTEEEIRTRIME-VAKRIADDYKGKGLR--P
35 Trypanosoma brucei          M---ACKYDFATSVLFTEAELHTRMRGVAQRIADDYSNCNLK--P
36 Tritrichomonas foetus       MTETPMMDDLE------RVLYNQDD--IQKRIRELAAELTEFYED
37 Leishmania donovani II      M---RRNYPMSAHTLVTQEQVWAATAKCAKKIAEDYRS-FKLT--
38 Toxoplasma gondi            MASKPIEDYGK----GKGRIEP---MYIPDNTFYNADDFL---VP
39 Escherichia coli            MVR--------DMKHTVEVMIPEAE--IKARIAELGRQITERYKD
40 Vibrio harveyi              MKHTV------------EVMISEQE--VQERIRELGKQITERYQ-
41 Rhodobacter capsulatus      MSQ---------SGYVIDQMISAKA--IAARVEALGAEITEAFKD
42 Lactococcus lactis          MLE------KNLDKAIEKVLVSEEE--IIEKSKELGEILTKEYEG
43 Bacillus subtillus II       MMK----------HDIEKVLISEEE--IQKKVKELGAELTSEYQD
44 Homo sapiens          OPRT  MAV--------ARAALGP--LVTGLYDVQAFKFGDFVLKSGLSSP
45 Trichoderma reesei          MAT------TSQLPAYKQDFLKSAIDG-GVLKFGSFELKSKRISP
46 Dictyostelium discoides     MN-------------IKELVLKLMKL--MAIKLGEFKLKSGIISP
47 Sacc. cerevisiae 10         MS-----ASTTSLEEYQKTFLELGLEC-KALRFGSFKLNSGRQSP
48 Sacc. cerevisiae 5          MP--------IMLEDYQKNFLELAIEC-QALRFGSFKLKSGRESP
49 Podospora anserina          MS---------ELPQYKKDFLKSAIDG-NILKFGSFLLKSGRTSP
50 Sordaria macrospora         MA----------ALRYKADFLKASIDG-GVLKFGSFELKSKRISP
51 Colletotrich. graminicola   MASQ--------LPPYKQDFLKAAIA--GVLKFGSFELKSKRISP
52 Rhizobium trifolii          MIQTTFPDRAVMAELLAKML-WEIKAV-HFNAAQPYKLASGMASP
53 Coxiella burnetii           MNQAT-----EIAKLL-----LNIKAV-TLNLHEPYRYTSGILSP
54 Bcillus subtillus           MGGN------QILKQIIAKHLLDIQAV-FLRPNEPFTWASGILSP
55 Escherichia coli            MKP------------YQRQFIEFALSK-QVLKFGEFTLKSGRKSP
56 Lactobacill. plantarum      MSE---------IAATIATDLLTIKAV-TLRPNDPFTWASGIKSP
57 Bos taurus            UPRT  MAA--------ADALLGS--LVTGLYDVQAFKFGNFVLKSGLSSP
58 Drosoph. melanogaster       MVA-------QNSDKMRALALKLFEIN--AFKFGDFKMKVGINSP

26 Unalignable                 *TWAGDYVLNVIATGLSLKARGKRRRQRWVDDGRVLALGE*
```

Figure 2.2 (cont. 3 of 7) : Exhaustive alignment of phosphoribosyl transferase sequences.

```
                30        40              50        60
             !!!!         !                        !!!!
1       IPGVLFRDISPLLKDPASFRAS-IRL---LASH-LKSTHGGKIDY----------IAGLD-
2       IPGVLFRDISPLLKDPDSFRAS-IRL---LASH-LKSTHSGKIDY----------IAGLD-
3       IPGVLFRDISPLLKDPDSFRAS-IRL---LASH-LKSTHSGKIDY----------IAGLD-
4       IPGVLFRDISPLLKDPDSFRAS-IRL---LASH-LKSTHSGKIDY----------IAGLD-
5       VQGVLFRDISPLLKDPVSFRAS-IHL---LASH-LKSTHSGKIDY----------IAGLD-
6       IPGVLFRDISPLLKDPDSFRAS-IRL---LAGH-LKSTHGGKIDY----------IAGLD-
7       IPGVLFRDISPLLKDPDSFRAS-IRL---LANH-LKSKHGGKIDY----------IAGLD-
8       IPGVLFRDISPLLKDPDSFRAS-IRL---LASH-LKSLHGGKIDY----------IAGLD-
9       TPGVVFRDISPVLKDPASF-RAAIGL---LARHLKA-THGGRIDY----------IAGLD-
10      EPGVVFKDITPLLADPGAF-AALTDA---LAEAAGR----TGATK----------VVGLE-
11      IEGEQFEDFLPIIGNPTLF-QKLVHT---FKTHLEEKFAKEKIDF----------IAGIE-
12      IEGEQFEDFLPIIGNPTLFQKL-VHT---FKTHLEEKFAKEKIDF----------IAGIE-
13      KPGILFQDITTLLLDPQAFRDT-TDL---FVERYKDKD----ITV----------VAGVE-
14      KPGIMFQDITTLLLDTEAF-KDTIAL---FVDRYKDK----GISV----------VAGVE-
15      KEGILFRDIFGALTDPKAC-V-LRDL---LVQYIRQ--SQPEVEV----------IVGLD-
16      KEGILFRDIFGALTDPKAC-VYLRDL---LVDHIRE--SAPEAEI----------IVGLD-
17      KKGINFRDIMPLFTNPQLVNELCVV----IADHVRH--TVGHVDS----------VAGLE-
18      KKSYRWYS--PVFS-PRNVPRF-ADVSSITESPETLKAIRDFLVQRYRAMSPAPTHILGFD-
19      KADG-------------------------------------------------IVCPE-
20      NQGTLFYDITPVFSNPQLFNFVLTQM---AQFIKAIN--EAEA----------IVCPE-
21      KPGIMFRDITTLLNSPEGLRYT-IDS---LVEQCESQ--ELVPDH----------VVGME-
22      KPGILFRDVTSLLEDPKAY-ALSIDL---LVERYKN----AGITK----------VVGTE-
23      KPGVVFRDITPLFQSPRALRMT-VDS---FVQRYIE----ADFSH----------IGAMD-
24      RPGVVFRDITPLFQSPKALRMV-ADS---LIQRYVE----ADFTH----------IGALD-
25      KEGIIFRDITTLLEVPAAFKAT-IDL---IVEQYRD----KGITK----------VLGTE-
26      RRGLAAVTEALAD------RASGADLVAGVAISVADVV--ASLTRDVADFPVPGVVEFKDD-
27      HPISDGV---PVV-EPKLLREVATRI-----IKIGDF---EGATK----------LVTAE-
28      KHYAEDLEKVFIPHGLIMD-RTERLA---RDVMKEM---GGHHI----------VALCVL
29      NHYAEDLEKVFIPHGLIMD-RTERLA---RDVMKEM---GGHHI----------VALCVL
30      NHYAEDLERVFIPHGLIMD-RTERLA---RDVMKEM---GGHHI----------VALCVL
31      NHYVEDLEKVFIPHGVIMD-RTERLA---RDVMKEM---GGHHI----------VALCVL
32      RYDECLDY-VLIPNGMIKD-RLEKMS---MDIVDYYEACNATSI----------TLMCVL
33      AHYKKYLTKVLVPNGVIKN-RIEKLA---YDIKKVY---NNEEF----------HILCLL
34      ---------------------------------------YVNPL----------VLISVL
35      ---------------------------------------LENPL----------VIVSVL
36      ---------------------------------------YKNP----------VMICVL
37      ---------------------------------------TDNPL----------YLLCVL
38      PHCKPYIDKILLPGGLVKD-RVEKLA---YDIHRTY---FGEEL----------HIICIL
39      -------------------------------------SGSDM----------VLVGLL
40      -------------------------------------GSEDL----------VMVGLL
41      -------------------------------------TDRL----------VVVGLL
42      ------------------------------------KNP----------LVLGIL
43      ------------------------------------TFP----------LAIGVL
44      IY---IDLRGIVSRPRLLS-QVADIL---FQTAQNAG--ISFDT----------VCGVPY
45      YFFNA-GEFHTARLARRIA-SAFAKT---IIEAQEKAG-LEFDI----------VFGPAY
46      IY---IDLRVTVSSPPL-L-AAIAEM---MYQKVY----KSGNA----------QETPAL
47      YFFNL-SLFNSGKLLANLA-TAYATA---IIQSE-----LKFDV----------IFGPAY
48      YFFNL-GLFNTGKLLSNLA-TAYAIA---IIQSD-----LKFDV----------IFGPAY
49      YFFNA-GDFYRADLLQAIS-TAYAKC---IIEAH-KSG-LDFDI----------VFGPAY
50      YFFNA-GDFYRADLLNSIS-TAYALT---IDSLP-----IQYDI----------IFGPAY
51      FFNA-GDFYRADLLRAISL--AYAHT---VIEAREATG-LDFDV----------VFGPAY
52      VYI---DCRKLLSFPRIRS-TVMDFA---ASTL------LRDAG----------FEQFDC
53      IY---CDNRLIISYPEKRK-MIIEAF---LQ--------LIEKN----------HLSFDI
54      IY---CDNRLTLSFPEVRN-DVASGI---SK--------LVKEH----------FPEAEM
55      YFFTA-GLFNTGRDLALLG-RFYAEA---LVDSG-----IEFDL----------LFGPAY
56      IY---TDNRLTISYPAVRQ-HIAHGI---AAIIKREY--PATEV----------IAGVAT
57      VY---IDLRGIISRPSILN-QVAEML---FQTAENAE--INFDT----------VCGVPY
58      VY---FDLRVIVSYPDVMQ-TVSDLL---VEHIKDKQ--LSAKH----------VCGVPY
```

Figure 2.2 (cont. 4 of 7) : Exhaustive alignment of phosphoribosyl transferase sequences.

```
                    70         80         90              100
                 !!!  !!! ! !! !        ! !    !          **
1           ---SRG--FLFGPSLAQELGLGCVLIRKRG-KLPGPTV-SA---SYAL
2           ---SRG--FLFGPSLAQELGVGCVLIRKQG-KLPGPTV-SA---SYSL
3           ---SRG--FLFGPSLAQELGVGCVLIRKQG-KLPGPTI-SA---SYAL
4           ---SRG--FLFGPSLAQELGVGCVLIRKQG-KLPGPTV-SA---SYSL
5           ---SRG--FLFGPSLAQELGVGCVLIRKRG-KLPGPTL-SA---SYAL
6           ---SRG--FLFGPSLAQELGVGCVLIRKRG-KLPGPTV-SA---SYSL
7           ---SRG--FLFGPSLAQELGLGCVLIRKRG-KLPGPTV-SA---SYAL
8           ---SRG--FLFGPSLAQELGVGCVLIRKRG-KLPGPTL-SA---SYAL
9           ---SRG--FLFGPSLAQELGLGCVLIRKRG-KLPGPTL-WA---SYSL
10          ---ARG--FILGAPVALRAGLGFIPVRKAG-KLPGATL-SQ---AYDL
11          ---ARG--LLFGPSLALALGVGFVPIRRVG-KLPGECA-SI---TFTK
12          ---ARG--LLFGPSLALALGVGFVPIRRVG-KLPGECA-SI---TFTK
13          ---ARG--FIFGPPIALAIGAKFVPIRKPK-KLPGEVI-SE---EYSL
14          ---ARG--FIFGPPIALAIGAKFVPMRKPK-KLPGKVI-SE---EYSL
15          ---SRG--FLFNLLLATELGVGYTPIRKKG-KLAGEVV-SV---EYQL
16          ---SRG--FLFNLLIATELGLGCAPIRKKG-KLAGEVV-SV---EYKL
17          ---ARG--FLFGPQVAIQLGVPFVPIRKKG-KLPGATI-EA---SYVK
18          ---ARG--FLFGPMIAVELEIPFVLMRKAD-KNAGLLIRSE---PYEK
19          ---ARG--FIFGGALASKTKLPLVLVRKPH-KLSGELA-RE---TYDL
20          ---ARG--FIFGGALASKTQLPLVLVRKAN-KLPGQLI-SA---SYDL
21          ---SRG--FLFGMPLAYQMNAGFIPVRKPG-KLPAPVH-RV---EYDL
22          ---ARG--FLFGAPVALGLGVGFVPVRKPG-KLPRETI-SE---TYDL
23          ---ARG--FLIGSAVAYALNKPLVLFRKQG-KLPADVL-AE---GYQT
24          ---ARG--FLVGSILAYELNKPLVLFRKQG-KLPADVL-SQ---AYST
25          ---SRG--FIFGAPVALALGLPFELVRKPK-KLPRETI-SQ---SYQL
26          ---ARG--FLVAAAVATRLEVGVLAVRKGG-KLPRPVL-SE---EYYR
27          ---AMG--IPLVTTLSLYTDIPYVIMRKREYKLPGEIP-VF---QSTG
28          -KGG----YKFFADLLDYIKSLN-RNTDRSI--PMTVD-FIRLKSYCN
29          -KGG----YKFFADLLDYIKALN-RNSDRSI--PMTVD-FIRLKSYCN
30          -KGG----YKFFADLLDYIKALN-RNSDRSI--PMTVD-FIRLKSYCN
31          -KGG----YKFFADLLDYIKALN-RNSDRSI--PMTVD-FIRLKSYCN
32          -KGG----FKFLADLVDGLERTV---RARGIVLPMSVE-FVRVKSYVN
33          -KGSRG--FFTALLKHLSRIHNYSAVETSK---PLFGEHYVRVKSYCN
34          -KGS----FMFTADLC-RALSDFNV---------PVRME-FICVSSYGE
35          -KGS----FVFTADMVR-ILGDFGV---------PTRVE-FLRASSYGH
36          -TGA----VFFYTDLL-KHLDFQLE---------PD----YIICSSYSG
37          -KGS----FIFTADLAR-FLADEGV---------PVKVE-FICASSYGT
38          -KGSRG--FFNLLIDYLATIQKYSGRESSV---PPFFEHYVRLKSYQN
39          -RGS----FMFMADLCR-EVQV-----------SHEVD-FMTASSYGS
40          -RGS----FVFMADLA-RAIEL-----------THQ-D-FMTASSYGN
41          -RGS----FVFIADLIR-EIGV-----------PCEVD-FLEASSYGN
42          -RGS----VPFLAELIK-HIDC-----------HLETD-FMTVSSYHG
43          -KGA----LPFMADLIK-HIDT-----------YLEMD-FMDVSSYGN
44          -TA-----LPLATVICS-TN-QI----------PM-------------
45          -KG-----IPLCSAITIKLGDVA----------PQNLD----RVSYSF
46          VCG-----VPYTALPIA-TGMSIANNI------PM-------------
47          -KG-----IPLATAATIKLGQIR----------PR----AKYAVGRVL
48          -KG-----IPLAAIVCVKLAEIGGT--------------KFQGIQYAF
49          -KG-----IPLAAIVCVKLAEIGGS--------------KFQNIQYAF
50          -KG-----IPLATAATDKLAQLD----------PE----TYGKICYSF
51          -KG-----IPLATSTTDKLAELD----------PA----RYGTTCYSF
52          IAGGETAGIPFAALLADRLGL------------PM-------------
53          VAGTATAGIPHAAWIADRLDL------------PM-------------
54          IAGTATAGIPHAALAADHLNL------------PM-------------
55          -KG-----IPIATTTAVALAEHHDLDL------PYCF-----------
56          ------AGIPHAAWVAE-LLNL-----------PL-------------
57          -TA-----LPLATIVCS-TH-EI----------PM-------------
58          -TA-----LPRATIVSV-QQ---GT--------PM-------------
```

Figure 2.2 (cont. 5 of 7) : Exhaustive alignment of phosphoribosyl transferase sequences.

```
                     110         120         130         140
          !!    !              * !****xxx*!*x!*!
1         --EY--GKAELEIQKDALEP--GQKVVVVDDLLATGGTMCAACEL
2         --EY--GKAELEIQKDALEP--GQRVVIVDDLLATGGTMFAACDL
3         --EY--GKAELEIQKDALEP--GQRVVIVDDLLATGGTMFAACDL
4         --EY--GKAELEIQKDALEP--GQRVVIVDDLLATGGTMFAACDL
5         --EY--GKAELEIQKDALEP--GQRVVIVDDLLATGGTMCAACEL
6         --EY--GKAELEIQKDALEP--GQKVVIVDDLLATGGTMCAACEL
7         --EY--GKAELEIQKDALEP--GQKVVIVDDLLATGGTMCAACQL
8         --EY--GKAELEIQKDALEP--GQRVVIVDDLLATGGTMCAACEL
9         --EY--GKAELEIQKDALEP--GQRVVVVDDLLATGGTMNAACEL
10        --EV--SSAEIEVHAEDLTA--GDRVLVVDDVLATGGTAKASLEL
11        --LD--HEEIFEMQ-VEAIPF-DSNVVVVDDVLATGGTAYAAGDL
12        --LD--HEEIFEMQ-VEAIPF-DSNVVVVDDVLATGGTAYAAGDL
13        --EY--GTDKIEMHVGAVQP--NDRVLIVDDLIATGGTLCAAAKL
14        --EY--GTDTIEMHVGAVEP--GERAIIIDDLIATGGTLAAAIRL
15        --EY--GSDTFELQRTAIQP--GQKVVIVDDLLATGGSLLAASEL
16        --EY--GSDTFELQKSAIKP--GQKVVVVDDLLATGGSLVAATEL
17        --EY--GEDRVEIQEGAIK--NGDIVFLIDDLLATGGTLRAATDL
18        --EYKEAAPEVMTIRYGS-IGKGSRVVLIDDVLATGGTALSGLQL
19        --EY--RQNSILEMRVDALE-NCKRCVIVDDLLATAGTVAAIDKL
20        --EYRKHAVLEMSTTSLIQANNAKRCVIVDDVLATAGTVAAIDQL
21        --EY--GKDSLEIHQDAVAP--HHRVLIVDDLIATGGTAKATAEL
22        --EY--GTDQLEIHVDAIKP--GDKVLVVDDLLATGGTIEATVKL
23        --EY--GEAFLEVHADSLCE--GDSVLIFDDLIATGGTLLAAASL
24        --EY--GEAHLEIHADSLCE--GDSVLLFDDLIATGGTLLAAAQL
25        --EY--GQDTLEMHVDAISE--GDNVLIIDDLLATGGTVEATVKL
26        --AY--GAATLEILAEGIEV-AGRRVVIIDDVLATGGTIGATRRL
27        --YS--KGQLYLNGIE-----KGDKVVIIDDVISTGGTMIAIIDA
28        --DQSTGDIKVIGGDDLSALT-GKNVLIVEDIIDTGKTMQTLLSL
29        --DQSTGDIKVIGGDDLSTLT-GKNVLIVEDIIDTGKTMQTLLSL
30        --DQSTGDIKVIGGDDLSTLT-GKNVLIVEDIIDTGKTMQTLLSL
31        --DQSTGDIKVIGGDDLSTLT-GKNVLIVEDIIDTGKTMQTLLSL
32        --DVSIHEPILTGLGDPSEYK-DKNVLVVEDIIDTGKTITKLISH
33        --DQSTGTLEIVS-EDLSCLK-GKHVLIVEDIIDTGKTLVKFCEY
34        --GVTSSGQVRMLLDTRHSIE-GHHVLIVEDIVDTALTLNYLYHM
35        --DTKSCGRVDVKADGLCDIR-GKHVLVLEDILDTALTLREVVDS
36        --TKSTG-NLTISKDLKTNIE-GRHVLVVEDIIDTGLTMYQLLNN
37        --GVETSGQVRMLLDVRDSVE-NRHILIVEDIVDSAITLQYLMRF
38        --DNSTGQLTVLS-DDLSIFR-DKHVLIVEDIVDTGFTLTEFGER
39        --GMSTTRDVKILKDLDEDIR-GKDVLIVEDIIDSGNTLSKVREI
40        --TMESSRDVRILKDLDDDIK-GKDVLIVEDIIDTGNTLNKIREI
41        --ETTSTREVRVLKDLRGIIG-GRDVLVVEDIIDTGHTISKVMEM
42        --GTKSSGEVKLILDVDTAVK-GRDILIVEDIIDTGRTLKYLKEL
43        --STVSSGEVKIIKDLDTSVE-GRDILIIEDIIDSGLTLSYLVEL
44        LIRRKETKDYGTKRLVEGTINPGETCLIIEDVVTSG-SSVLETVE
45        --DRKEAKDHGEGGNIVGASLKGKRVLIVDDVITAG-TAKRDAIE
46        VVRRKEAKAYGTKQLIEGRFKEGDNVLVVEDLVTSGASDLETVRD
47        VRDRKEAKDHGEGGNIVGAPLKGKRVLIVDDVISRC-TAKREAIA
48        --NRKKVKDHGEGGIIVGASLEDKRVLIIDDVMTAG-TRINEAFE
49        --NRKEAKDHGEGGIIVGSALENKRILIIDDVMTAG-TAINEAFE
50        --DRKEAKDHGEGGNIVGAPLKGKRILIVDDVITAG-TAKREAIA
51        --DRKEAKDHGEGGNIVGAPLKGQKVLIVDDVITAG-TAKREAIA
52        IYVRKQPKGHGRNAQIEGNMPEGSRVLVIEDLTTAGGSMFKFIDA
53        IYVRAKAKTHGKQNQIEGRIRKGQRALIVEDLISTGKSALAAGLA
54        CYVRSKPKAHGKGNQIEGAVQEGQKTVVIEDLISTGGSVLEACAA
55        --NRKEAKDHGEGGNLVGSALQGR-VMLVDDVITAG-TARESMEI
56        IYVRAKPKDHGAGRQIEGVLKPGQKVVMLDDLISTGGS-VLQAAK
57        LIRRKEKKDYGTKRLIEGAVNPGDTCLIIEDVVSSG-SSVWETAE
58        LVRRKEAKAYGTKKLVEGIFNAGDTCLIVEDVVTSG-SSILDTVR
```

Figure 2.2 (cont. 6 of 7) : Exhaustive alignment of phosphoribosyl transferase sequences.

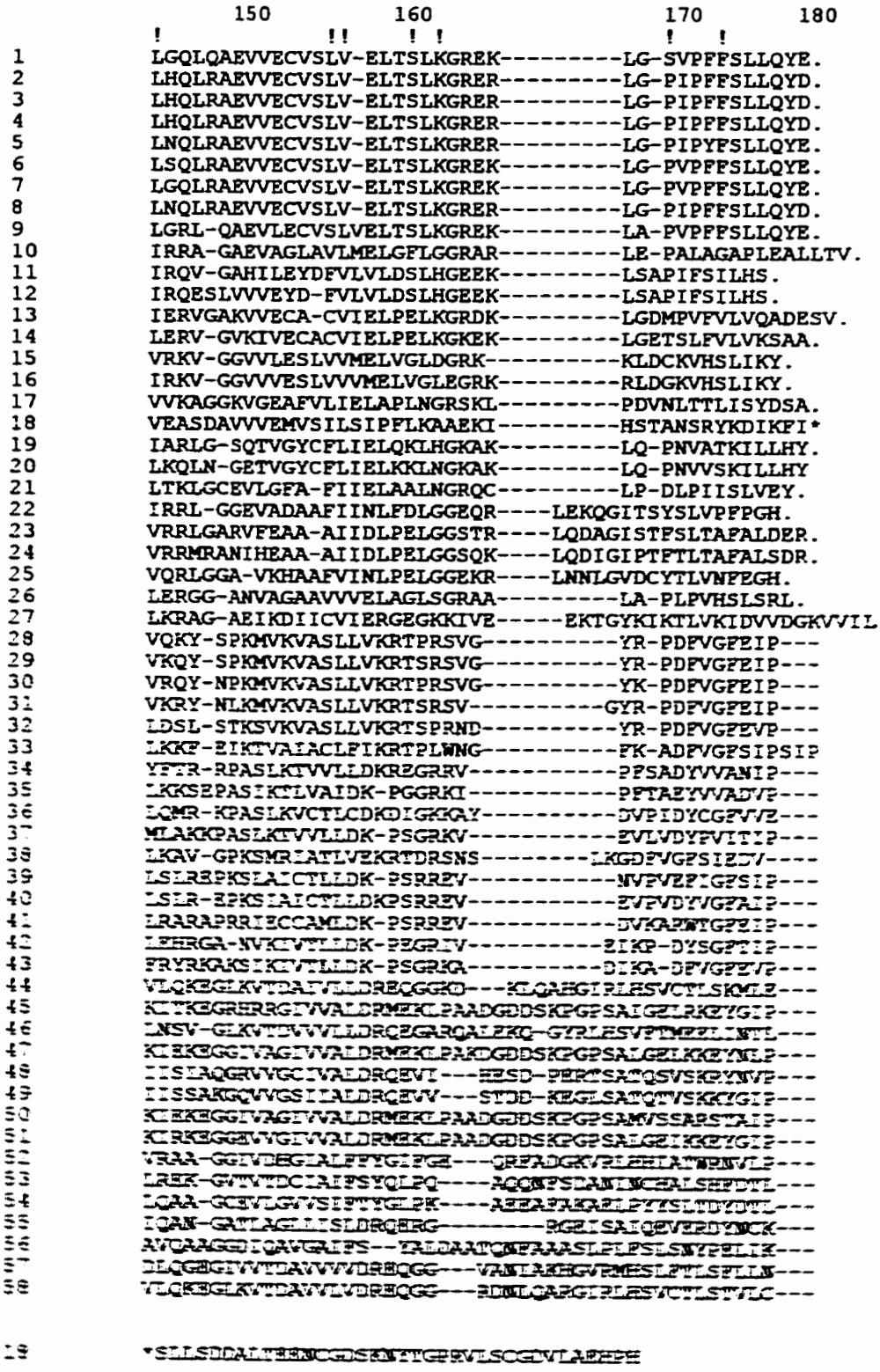


Figure 2.2 (cont. 7 of 7) : Exhaustive alignment of phosphoribosyl transferase sequences

```
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28          DKFVVGYALDYNEYFRDLNHVCVISESGKAKYKA.
29          DKFVVGYALDYNEYFRNLNHVCVISETGKAKYKA.
30          DKFVVGYALDYNEYFRDLNHVCVISETGKAKYKA.
31          DKFVVGYALDYNEYFRDLNHICVISETGKAKYKA.
32          NRFVVGYALDYNDNFRDLHHICVINEVGQKKFSVPCTSKPV.
33          DHFVVGYSLDYNEIFRDLDHCCLVNDEGKKKYKATSL.
34          NAFVIGYGLDYDDTYRELRDIVVLRPEVYAERE*
35          NVFVVGYGLDYDQSYREVRDVVILKPSVYETWGKELERRKAAGEAKR.
36          NRYIIGYGFDFHNKYRNLPVIGILKESVYT.
37          HAFVIGYGMDYAESYRELRDICVLKKEYYEKPESKV.
38          --WIVGCCYDFNEMFRDFDHVAVLSDAARKKFEK.
39          DEFVVGYGIDYAQRYRHLPYIGKVILLDE.
40          DEFVVGVGIDYAQKYRDLPFIGKVVPQE.
41          DEFVVGYGLDYAQNHRNLPFIGTVRFTDPQK.
42          NEFVVGFGLDYEENYRNLPYVGVLKPEVYNK.
43          DAFVVGYGLDYAERYRNLPYIGVLKPAVYES.
44          ILEQQKKVDAETVGR.
45          IFAILTLDDIIDGMKGFATPEDIKNTEDYRAKYKATD.
46          IEAGKLTGRTLELVQSFLDANRNV.
47          IYAILTLDDIIEGIKGLVGEEDIKRTEEYREKYKATD.
48          VLSIVSLTQVVQFMGLRLSPEQKSAIENYRKAYGI.
49          VISIVSLIHIITYLEGRITAEEKSKIEQYLQTYGASA.
50          IFAILTLDDIIEGMRGLASPEDVKKTEEYRAKYKATD.
51          IFSILTLDDIIEGAK.
52          SPGSRSSSTTRRCRKSSPSSMRRWLGRERMVA.
53          IEMAVDEGYIDEIEKQKALAWNKDPEHWQP.
54          TEVALENGNIHSDDLKKLQTWKRNPESKDWFKK.
55          VISIITLKDLIAYLEERLEMAEHLAAVKAYREEFGV.
56          VARQADYIDDEELAS.
57          ILEQQKKINAETVER.
58          TLHEAGRIEKSTVEA.


34          * AARQKKQRAIGSADTDRDAKREFHSKY.
```

Figure 2.3 (1 of 2) : Detail alignment of the region of the A motif. In this area the consensus sequence for phosphoribosyl transferases is poor. A general pattern for the A motif (L-I-V-M)$_{2-4}$GXXXXGK(T/S)(L-I-V-M) is approximated in APRT and HPRT, but in OPRT or UPRT sequences is essentially undetectable.

```
1  Cricet. longicaudus  APRT
2  Mus musculus
3  Mus pahari
4  Mus spicelegus
5  Stochomys longicaudatus
6  Rattus norvegicus
7  Gerbillus campestris
8  Mastomys hildebrantii
9  Homo sapiens
10 Streptomyces coelicor
11 Sacc. cerevisiae I
12 Sacc. cerevisiae II
13 Triticum aestivum
14 Arabadopsis thaliana
15 Drosoph. pseudoobscura
16 Drosoph. melanogaster
17 Caenorhabditis elegans
18 Leishmania donovani
19 Mycoplasma pneumoniae
20 Mycoplasma genitalium
21 Synechocystis sp.
22 Escherichia coli
23 Pseudomonas aeruginosa
24 Pseudomonas stuzeri
25 Heamophilus.influenzae
26 Mycobact. tuberculosis
27 Methano. jannaschii
28 Mastom. unguiculatus HPRT
29 Mus musculus
30 Homo sapiens
31 Cricet. longicaudus
32 Schistosoma mansonii
33 Plasmodium falciparum
34 Trypanosoma cruzi
35 Trypanosoma brucei
36 Tritrichomonas foetus
37 Leishmania donovani II
38 Toxoplasma gondi
39 Escherichia coli
40 Vibrio harveyi
41 Rhodobacter capsulatus
42 Lactococcus lactis
43 Bacillus subtillus II
44 Homo sapiens          OPRT
45 Trichoderma reesei
46 Dictyostelium discoides
47 Sacc. cerevisiae 10
48 Sacc. cerevisiae 5
49 Podospora anserina
50 Sordaria macrospora
51 Colletotrich. graminicola
52 Rhizobium trifolii
53 Coxiella burnetii
54 Bcillus subtillus
55 Escherichia coli
56 Lactobacill. plantarum
57 Bos taurus            UPRT
58 Drosoph. melanogaster
```

Figure 2.3 (2 of 2) :

```
A motif                              LIVMGXXXXG--KSLIVM
1          IAGLDSRG---FLFGPSLAQELGLGCVLIRKRG--KLPGPTV-SA--SYAL
2          IAGLDSRG---FLFGPSLAQELGVGCVLIRKQG--KLPGPTV-SA--SYSL
3          IAGLDSRG---FLFGPSLAQELGVGCVLIRKQG--KLPGPTI-SA--SYAL
4          IAGLDSRG---FLFGPSLAQELGVGCVLIRKQG--KLPGPTV-SA--SYSL
5          IAGLDSRG---FLFGPSLAQELGVGCVLIRKRG--KLPGPTL-SA--SYAL
6          IAGLDSRG---FLFGPSLAQELGVGCVLIRKRG--KLPGPTV-SA--SYSL
7          IAGLDSRG---FLFGPSLAQELGLGCVLIRKRG--KLPGPTV-SA--SYAL
8          IAGLDSRG---FLFGPSLAQELGVGCVLIRKRG--KLPGPTL-SA--SYAL
9          IAGLDSRG---FLFGPSLAQELGLGCVLIRKRG--KLPGPTL-WA--SYSL
10         IAGLDARG---FILGAPVALRAGLGFIPVRKAG--KLPGATL-SQ--AYDL
11         IAGIEARG---LLFGPSLALALGVGFVPIRRVG--KLPGECA-SI--TFTK
12         IAGIEARG---LLFGPSLALALGVGFVPIRRVG--KLPGECA-SI--TFTK
13         VAGVEARG---FIFGPPIALAIGAKFVPIRKPK--KLPGEVI-SE--EYSL
14         VAGVEARG---FIFGPPIALAIGAKFVPMRKPK--KLPGKVI-SE--EYSL
15         IVGLDSRG---FLFNLLLATELGVGYTPIRKKG--KLAGEVV-SV--EYQL
16         IVGLDSRG---FLFNLLIATELGLGCAPIRKKG--KLAGEVV-SV--EYKL
17         VAGLEARG---FLFGPQVAIQLGVPFVPIRKKG--KLPGATI-EA--SYVK
18         ILGFDARG---FLFGPMIAVELEIPFVLMRKAD--KNAGLLIRSE--PYEK
19         IVCPEARG---FIFGGALASKTKLPLVLVRKPH--KLSGELA-RE--TYDL
20         IVCPEARG---FIFGGALASKTQLPLVLVRKAN--KLPGQLI-SA--SYDL
21         VVGMESRG---FLFGMPLAYQMNAGFIPVRKPG--KLPAPVH-RV--EYDL
22         VVGTEARG---FLFGAPVALGLGVGFVPVRKPG--KLPRETI-SE--TYDL
23         IGAMDARG---FLIGSAVAYALNKPLVLFRKQG--KLPADVL-AE--GYQT
24         IGALDARG---FLVGSILAYELNKPLVLFRKQG--KLPADVL-SQ--AYST
25         VLGTESRG---FIFGAPVALALGLPFELVRKPK--KLPRETI-SQ--SYQL
26         EFKDDARG---FLVAAAVATRLEVGVLAVRKGG--KLPRPVL-SE--EYYR
27         LVTAEAM----GIPLVTTLSLYTDIPYVIMRKREYKLPGEIP-VF--QSTG
28         VALCVLKGG--YKFFADLLDYIKSLN---RNTD-RSIPMTVDFIRLKSYCN
29         VALCVLKGG--YKFFADLLDYIKALN---RNSD-RSIPMTVDFIRLKSYCN
30         VALCVLKGG--YKFFADLLDYIKALN---RNSD-RSIPMTVDFIRLKSYCN
31         VALCVLKGG--YKFFADLLDYIKALN---RNSD-RSIPMTVDFIRLKSYCN
32         TLMCVLKGG--FKFLADLVDGLERTV---RARG-IVLPMSVEFVRVKSYVN
33         HILCLLKGSR-GFFTALLKHLSRIHNYSAVETS--KPLFGEHYVRVKSYCN
34         VLISVLKGS--FMFTADLCR-ALSDFNV---------PVRME-FICVSSYGE
35         VIVSVLKGS--FVFTADMVR-ILGDFGV---------PTRVE-FLRASSYGH
36         VMICVLTGA---VFFYTDLL-KHLDFQLE--------PD----YIICSSYSG
37         YLLCVLKGS--FIFTADLAR-FLADEGV---------PVKVE-FICASSYGT
38         HIICILKGSR-GFFNLLIDYLATIQKYSGRESSV--PPFFEHYVRLKSYQN
39         VLVGLLRGS--FMFMADLCR-EVQV-----------SHEVD-FMTASSYGS
40         VMVGLLRGS--FVFMADLAR-AIEL-----------THQ-D-FMTASSYGN
41         VVVGLLRGS--FVFIADLIR-EIGV-----------PCEVD-FLEASSYGN
42         LVLGILRGS--VPFLAELIK-HIDC----------HLETD--FMTVSSYHG
43         LAIGVLKG---ALPFMADLIK-HIDT---------YLEMD--FMDVSSYGN
44         VCGVPYT----ALPLATVICS-TN-QI---------PM-------------
45         VFGPAYK----GIPLCSAITIKLGDVA---------PQNLD----RVSYSF
46         QETPALVCG-VPYTALPIA-TGMSIANNI-------PM-------------
47         IFGPAYK----GIPLATAATIKLGQIR---------PR----AKYAVGRVL
48         IFGPAYK----GIPLAAIVCVKLAEIGGT-------------KFQGIQYAF
49         VFGPAYK----GIPLAAIVCVKLAEIGGS-------------KFQNIQYAF
50         IFGPAYK----GIPLATAATDKLAQLD---------PE----TYGKICYSF
51         VFGPAYK----GIPLATSTTDKLAELD---------PA----RYGTTCYSF
52         QFDCIAGGETAGIPFAALLADRLGL-----------PM-------------
53         SFDIVAGTATAGIPHAAWIADRLDL-----------PM-------------
54         EAEMIAGTATAGIPHAALAADHLNL-----------PM-------------
55         LFGPAYK----GIPIATTTAVALAEHHDLDL-----PYCF-----------
56         IAGVAT----AGIPHAAWVAE-LLNL----------PL-------------
57         VCGVPYT----ALPLATIVCS-TH-EI---------PM-------------
58         VCGVPYT----ALPRATIVSV-QQ---GT-------PM-------------
```

2.3.2 Motifs of phosphoribosyl transferases

Following exmaination of aligned APRT protein sequences (Figure 2.1) , several features are apparent. In particualr it can be seen that there are several identifiable regions of sequence similarity. The first region is the so-called Walker A domain, also identified by similarity as the phosphoryl binding loop
in related ATP binding proteins, and indicated in Figure 2.3. This domain has a poor consensus prediction.

In Figure 2.4, the highest scoring of the highlighted regions of the aligned APRT molecules from Figure 2.1 is shown in detail. This corresponds to the so-called Walker B motif of the nucleotide-binding cassette (ntbc), as characterized for nucleotide and polynucleotide kinases (Walker *et al.*, 1982). A consensus sequence for this domain is also shown in Figure 2.4.

In addition, the B domain is of interest because this sequence can be identified in nucleotide-binding proteins other than phosphoribosyl transferases, as shown in Figure 2.5. In the region between residues 120 to 134 (using the numbering scheme for CHO APRT), strong sequence conservation can be identified, even though the flanking sequences appear to be completely unrelated. The same pattern of divergence is seen in HPRT molecules as well, as shown in Figure 2.3.

2.3.3 Phylogeny of phosphoribosyl transferases

The alignments of each group of phosphoribosyl transferases is shown in Figure 2.2. From these alignments a reconstruction of the evolution of adenine phosphoribosyl

Figure 2.4 (1 of 2): Primary sequences of adenine phosphoribosyl transferases from various organisms, aligned around the region about D127 and D128 of CHO APRT. The APRT consensus for the B motif is indicated; capital letters indicate high conservation (>20/27), X indicates no apparent conservation. The numbering scheme is that of Chinese hamster (*C. longicaudatus*) APRT.

1 *Cricetulus longicaudatus*
2 *Mus musculus*
3 *Mus pahari*
4 *Mus spicelegus*
5 *Rattus norvegicus*
6 *Gerbillus cempestris*
7 *Stochomys longicaudatus*
8 *Mastomys hildebrantii*
9 *Homo sapiens*
10 *Drosophila melanogaster*
11 *Drosophila pseudoobscura*
12 *Saccaromyces cerevisiae I*
13 *Saccaromyces cerevisiae II*
14 *Arabadopsis thaliana*
15 *Triticum aestivum*
16 *Synechocystis sp.*
17 *Escherichia coli*
18 *Pseudomonas aerogenes*
19 *Pseudomonoas stuzeri*
20 *Haemophilus influenzae*
21 *Streptomyces coelicor*
22 *Mycobacterium tuberculosis*
23 *Caenorhabditis elegans*
24 *Leishmania donovani*
25 *Mycoplasma pneumoniae*
26 *Mycoplasma genitalium*
27 *Methanococcus jannaschii*

Figure 2.4 (cont. 2 of 2): Primary sequences of adenine phosphoribosyl transferases aligned around the region about D127 and D128 of CHO APRT.

```
            100         110         120         130         140         150
B motif     |           |           GX-KVvivDDllATGGTlXAA|         |
1       TV-SASY-ALEYGKAE-LEIQKDALE--PGQ-KVVVVDDLLATGGTMCAACELLGQLQAEVV
2       TV-SASY-SLEYGKAE-LEIQKDALE--PGQ-RVVIVDDLLATGGTMFAACDLLHQLRAEVV
3       TI-SASY-ALEYGKAE-LEIQKDALE--PGQ-RVVIVDDLLATGGTMFAACDLLHQLRAEVV
4       TV-SASY-SLEYGKAE-LEIQKDALE--PGQ-RVVIVDDLLATGGTMFAACDLLHQLRAEVV
5       TV-SASY-SLEYGKAE-LEIQKDALE--PGQ-KVVIVDDLLATGGTMCAACELLSQLRAEVV
6       TV-SASY-ALEYGKAE-LEIQKDALE--PGQ-KVVIVDDLLATGGTMCAACQLLGQLRAEVV
7       TL-SASY-ALEYGKAE-LEIQKDALE--PGQ-RVVIVDDLLATGGTMCAACELLNQLRAEVV
8       TL-SASY-ALEYGKAE-LEIQKDALE--PGQ-RVVIVDDLLATGGTMCAACELLNQLRAEVV
9       TL-WASY-SLEYGKAE-LEIQKDALE--PGQ-RVVVVDDLLATGGTMNAACELLGRLQAEVL
10      VV-SVEY-KLEYGSDT-FELQKSAIK--PGQ-KVVVVDDLLATGGSLVAATELIRKVGGVVV
11      VV-SVEY-QLEYGSDT-FELQRTAIQ--PGQ-KVVIVDDLLATGGSLLAASELVRKVGGVVL
12      CA-SITFTKLDHEEI--FEMQVEAIP--FDS-NVVVVDDVLATGGTAYAAGDLIRQESLVVV
13      CF-KATY-EKEYGSDL-FEIQKNAIP--AGS-NVIIVDDIIATGGSAAAAGELVEQVGAHIL
14      VI-SEEY-SLEYGTDT-IEMHVGAVE--PGE-RAIIIDDLIATGGTLAAAIRLLERVGVKIV
15      VI-SEEY-SLEYGTDK-IEMHVGAVQ--PND-RVLIVDDLIATGGTLCAAAKLIERVGAKVV
16      VH-RVEY-DLEYGKDS-LEIHQDAVA--PHH-RVLIVDDLIATGGTAKATAELLTKLGCEVL
17      TI-SETY-DLEYGTDQ-LEIHVDAIK--PGD-KVLVVDDLLATGGTIEATVKLIRRLGG-EV
18      VL-AEGY-QTEYGEAF-LEVHADSLC--EGD-SVLIFDDLIATGGTLLAAASLVRRLGARVF
19      VL-SQAY-STEYGEAH-LEIHADSLC--EGD-SVLLFDDLIATGGTLLAAAQLVRRMRANIH
20      TI-SQSY-QLEYGQDT-LEMHVDAIS--EGD-NVLIIDDLLATGGTVEATVKLVQRLGGA-V
21      TL-SQAY-DLEVSS---AEIEVHAEDLTAGD-RVLVVDDVLATGGTAKASLELIRRAG-AEV
22      VL-SEEY-YRAYGAAT-LEILAEGIE-VAGR-RVVIIDDVLATGGTIGATRRLLERGG-ANV
23      TI-EASY-VKEYGEDR-VEIQEGAIK--NGD-IVFLIDDLLATGGTLRAATDLVVKAGGKVG
24      LIRSEPYEKEY-KEAAPEVMTIRYGSIGKGS-RVVLIDDVLATGGTALSGLQLVEASDAVVV
25      LA-RETY-DLEYRQNSILEMRVDALENCK---RCVIVDDLLATAGTVAAIDKLIARLG SQT
26      LI-SASY-DLEYRKHAVLEMSTTSLIQANNAKRCVIVDDVLATAGTVAAIDQLLKQLN-GET
27      IP--VFQSTGYSKG-QLYLNGIE-----KGD-KVVIIDDVISTGGTMIAIIDALKRAG-AEI
```

Figure 2.5 (1 of 2) : Aligned phosphoribosyl transferase sequences in the region of the so-called B motif. Note the extremely conserved asparate and glycine residues near the middle of the tract. These residues are conserved between families of nucleotide-binding proteins, as evidenced by the sequences at the bottom of the figure, taken from functionally and specifically diverse sources.

Figure 2.5 (cont. 2 of 2) :

Escherichia coli	**APRT**	GTDQLEIHVDAIKPGDKVLVVDDLLATGGTIEATVKLIR
Streptomyces coelicor		SSAEIEVHAEDLTAGDRVLVVDDVLATGGTAKASLELIR
Saccaromyces cerevisiae		HEEIFEMQVEAIPFDSNVVVVDDVLATGGTAYAAGDLIR
Leishmania donovani		APEVMTIRYGSIGKGSRVVLIDDVLATGGTALSGLQLVE
Drosophila melanogaster		GSDTFELQKSAIKPGQKVVVVDDLLATGGSLVAATELIR
Arabadopsis thaliana		GTDTIEMHVGAVEPGERAIIIDDLIATGGTLAAAIRLLE
Homo sapiens		GKAELEIQKDALEPGQRVVVVDDLLATGGTMNAACELLG
Vibrio harveyia	**HPRT**	RDVRILKDLDDDIKGKDVLIVEDIIDTGNTLNKIREILS
Trypanosoma brucei		GQVRMLLDTRHSIEGHHVLIVEDIV-TALTLNYLYHMYF
Tritrichomonas foetus		NLTI-SKDLKTNIEGRHVLVVEDIIDTGLTMYQLLNNLQ
Toxoplasma gondii I		QLTVLS-DDLSIFRDKHVLIVEDIVDTGFTLTEFGERLK
Schistosoma mansoni		EPILTGLGDPSEYKDKNVLVVEDIIDTGKTITKLISHLD
Plasmodium falciparum		TLEIVS-EDLSCLKGKHVLIVEDIIDTGKTLVKFCEYLK
Homo sapiens		DIKVIGGDDLSTLTGKNVLIVEDIIDTGKTMQTLLSLVR
Trichoderma reesei	**OPRT**	DHGEGGNIVGASLKGKRVLIVDDVI-TAGTAKRDAIEKIT
Rhizobium trifolii		GHGRNAQIEGNMPEGSRVLVIEDLT-TAGGSMFKFIDAVR
Coxiella burnetii		THGKQNQIEGRIRKGQRALIVEDLI-STGKSALAAGLALR
Escherichia coli		DHGEGGNLVGSALQG-RVMLVDDVI-TAGTA-RESMEIIQ
Dictyostelium discoides		AYGTKQLIEGRFKEGDNVLVVEDLV-TSGASDLETVRDLN
Podospora anserina		DHGEGGNIVGAPLKGKRVLIVDDVI-SRCTAKREAIAKIE
Saccharomyces cervisiae 5		DHGEGGIIVGSALENKRILIIDDVM-TAGTAINEAFEIIS
Sordaia macrospora		DHGEGGNIVGAPLKGKRILIVDDVI-TAGTAKREAIAKIE
Rattus norvegicus I	**PRPPSYN**	EVDRM--VLVGDVKDRVAILVDDMADTCGTICHAADNLL
Homo sapiens		EVDRM--VLVGDVKDRVAILVDDMADTCGTICHAADKLL
Caenorhabditis elegans		EVEKM--TLVGSVEGKVAILVDDMADTCGTICMAADKLV
Saccharomyces cerevisiae I		EVSRM--VLVGDVTDKICIIVDDMADTCGTLAKAAEILL
Saccharomyces cerevisiae II		-----LITLVGNVRGRSAIILDDMIDRPGSFISAAEHLV
Rattus norvegicus II		------ITVVGDVGGRIAIIVDDIIDDVESFVAAAETLK
Salomonella typhimurium		ANVSQVMHIIGDVAGRDCVLVDDMIDTGGTLCKAAEALK
Bacillus subtilis		PNVAEVMNIVGNIEGKTAILIDDIIDTAGTITLAANALV
Synechocystis sp.		HNVAEVLNVIGDVQGKTAVLVDDMIDTAGTICEGARLLR
Leishmania donovani		AAGKVDTMQVGEVAGFTCIIVDDMIDTGGTLVKACELLK
Azo. brasilense	**ANTHPRT**	AA-RVMRAKAIPVEA-PDGTI-DTCGTGGDGSGTYNIST
Halo. volcanii		GFAQGMRDAALIHRARPAARSSDTAGTGGDDYNTINVLD
Chicken	**GLNPRT**	GVRMKLSAVRGVVEGKRVVMVDDSIVRGTTSRRIVTMLR
Homo sapiens		GVAKKFGVLSDNFKGKRIVLVDDSIVRGNTISPIIKLLK
S. typhimurium	**QNPRT**	AHATVITREDGVFCGKRWV*GDDVRLTWHVDDGDAIHAN
OTHER		
H-ras human		------DPTISEDSYRKQVVIDGETCLLDILDTAGQEE-
h-ras *Saccharomyces*		------DPTI-EDSYRKQVVIDDKVSILDILDTAGQEE-
Thymidine kinase human		PDIVDFCEMMANEGKTVIVAALDGTFQRKAFGSILNL--
gal. kinase *Saccharomyces*		ANMGPGYHMSKQNLMR-ITVVADIMLVLTMAVWIRLPL-
tlrC ATPase *S.fradiae*		GERSRLALAATLASQPELLLLDEPTNDLDDRAVHWLEEH
uvrA M. tuberculosum		GRVKLASELQKRSTGRTVYILDEPTTGLHFDDIRKLLNV
CFTR human		GHKQLMCLARSVLSKAKILLLDEPSAHLDPVTYQIIRRT
eEF-3 Saccharomyces		GQKVKLVLAAGTWQRPHLIVLDEPTNYLDRDSLGALSKA
MDR-1 human		GQKQRIAIARALVRNPKILLLDEATSALDTESEAVVQVA
recA Legionella		EQALEITDMLVRSAAVDVVIIDSVAALTPKAEIEGEMGD

transferases was developed using MEGALIGN (DNASTAR), and is shown in Figure 2.6. From an alignment of the most highly-conserved region from the global alignment (centered around residues 127-128 of CHO APRT), a second phylogeny was reconstructed. In addition, an alignment of phosphoribosyl transferase sequences, with the exclusion of the most conserved sequences was evaluated. The result of these analyses were not substantially different from the phylogenies reconstructed using the entire sequences. In Figure 2.7, a molecular phylogeny of phosphoribosyl transferases in addition to APRT is illustrated.

2.3.4 Secondary structure in phosphoribosyl transferases

In Figure 2.8, results from secondary structure analysis of APRT sequences are shown. Figure 2.9 is a secondary structure comparison between CHO and human APRT and human HPRT. Also indicated is the center for the conserved B motif, corresponding to aspartate residue 128 in CHO APRT.

2.4 Discussion

Coding sequences for *aprt* have been identified in many organisms. An exhaustive computer search of available sequence databases recovers some 20 unique *aprt* sequence entries. The degree to which any specific residue or tract of residues has an essential role in the function or structure of an enzymes can be inferred to a certain extent from comparisons of conserved and divergent sequences within a family of presumptively homologous enzymes. Conserved regions are thought to be catalytically or structurally essential regions which are under selective pressure. Implicit in this approach rests a common ancestry from which the sequences from different organisms must have arisen.

Figure 2.6 : Cladistic analysis of representative APRT primary protein sequences. This molecular phylogeny shows exceptions to the accepted evolutionary history of the species and Kingdoms. The time span of species shown here is by typical estimates, 800 million to one billion years. See Table 2.1 for species notations.

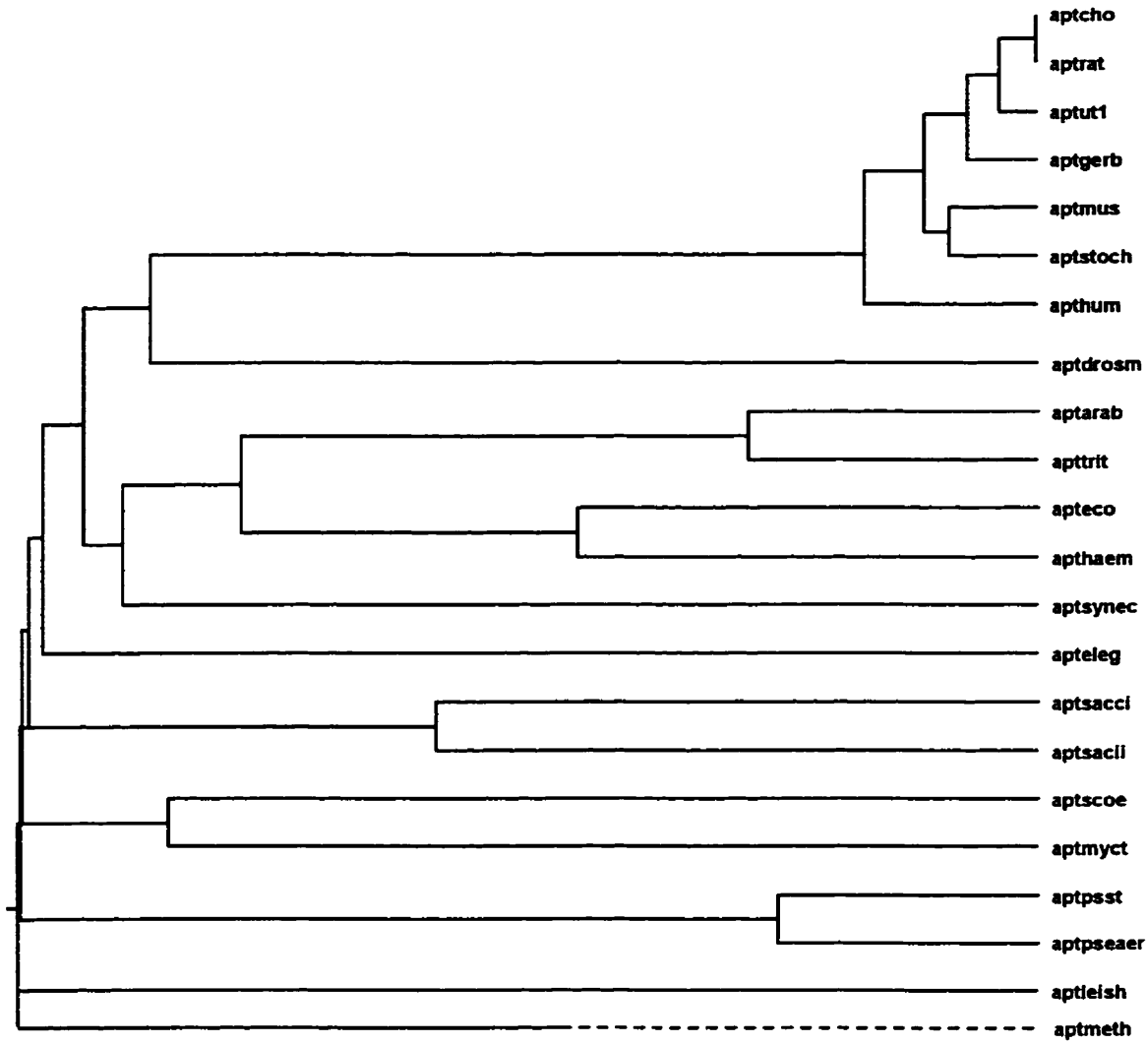


Figure 2.7 : Phylogenetic analysis of representative APRT (apt), HPRT (hpt), OPRT (opt), XGPRT (xpt), and ATP phosphoribosyl amido transferase (atp) primary protein sequences. This phylogeny does not reconstruct evolutionary history of the species and Kingdoms, but with some exceptions, sorts across species boundaries according to substrate. See Table 2.1 for species notations.

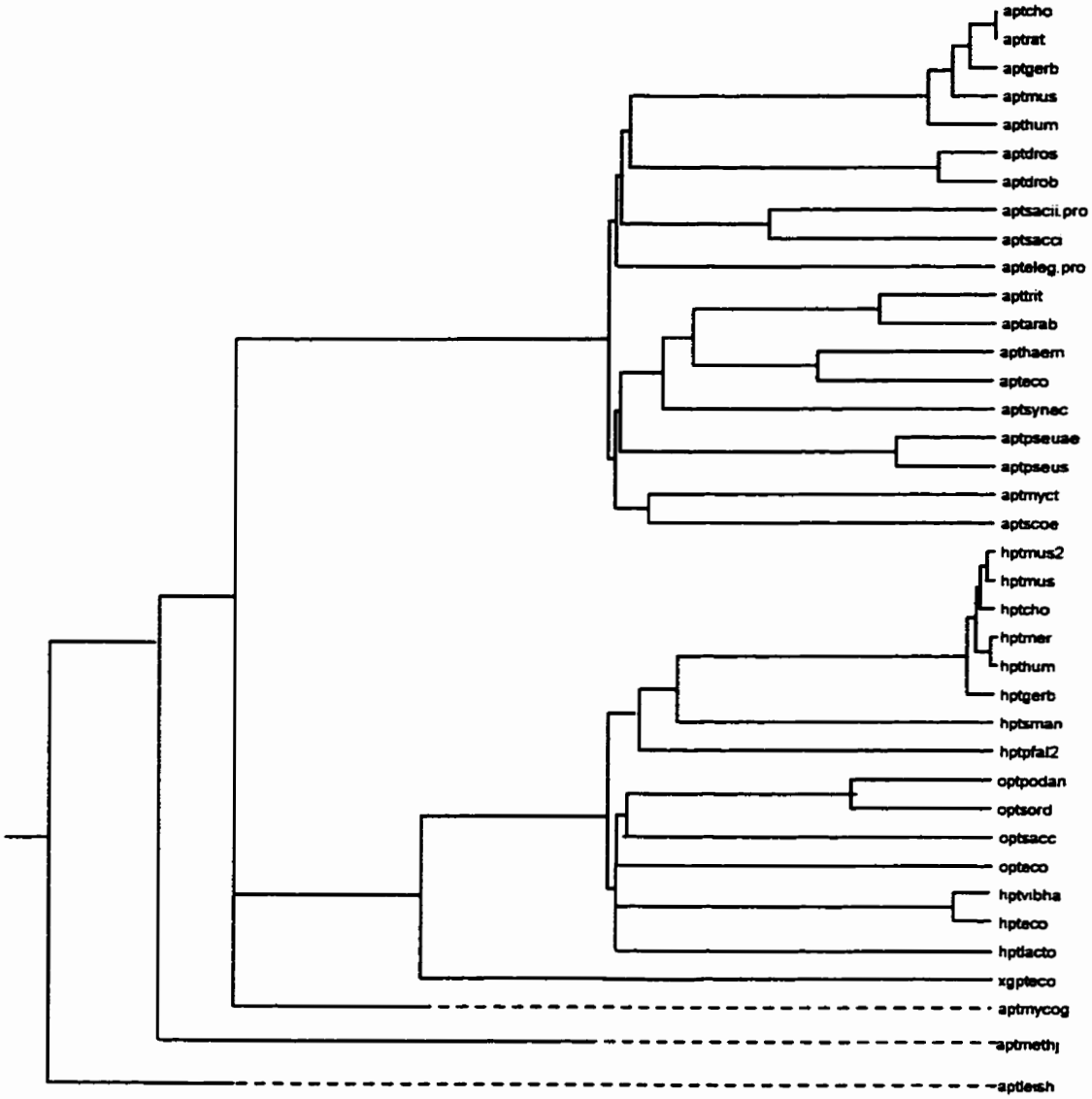


Figure 2.8 : Global secondary structure analysis by the method of Garnier and Robison. Shown here are results for APRT sequence from Chinese hamster (CHO), *Escherichia coli* (ECO), *Haemophilus influenzae* (HAE), and *Saccharomyces cerevisiae* (I) (SACC). These APRT sequences are of similar length (183 ± 5 residues). The summary bars represent alpha helix (A), beta sheet (B), turn (T), and coil (C) predictors.

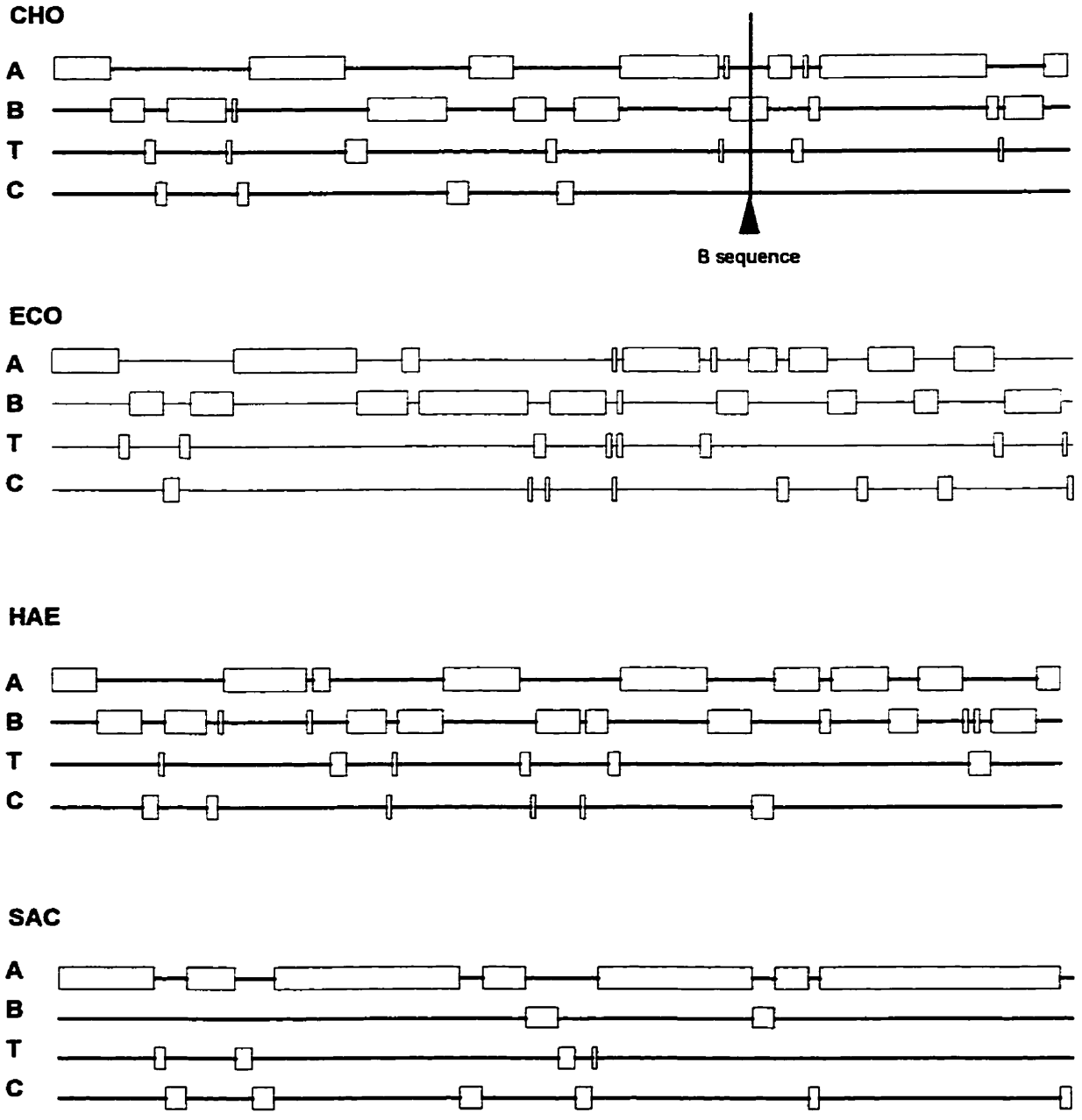
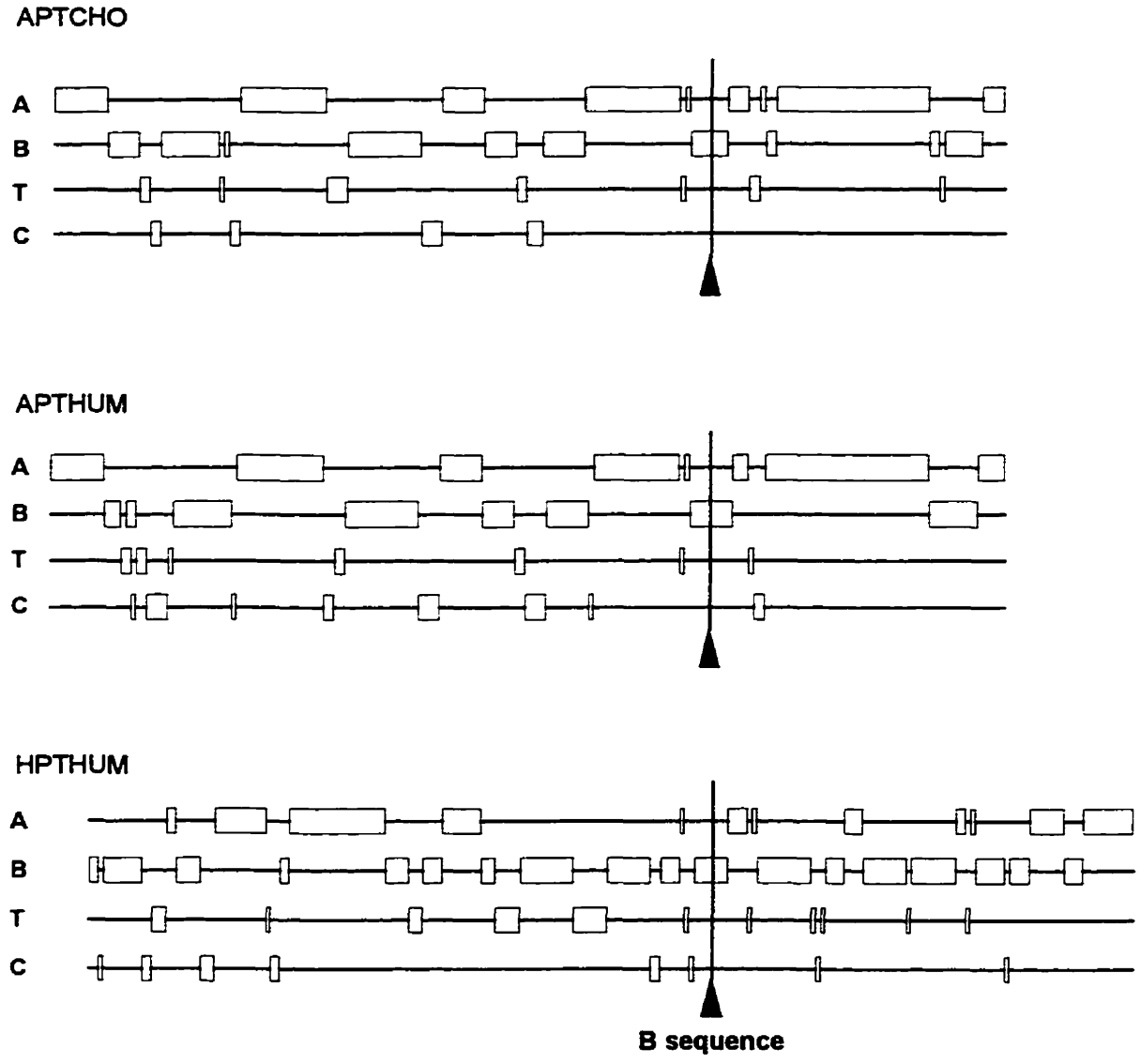


Figure 2.9 : Comparison of secondary structure (Garnier and Robison) motifs of APRT (human, APTHUM, Chinese hamster, APTCHO) and HPRT (human, HPTHUM). Sequences are to scale, aligned relative to the center of the highly conserved B sequence as indicated. Note the very different global secondary structure predictions for the HPRT sequence relative to the APRT sequences. Interestingly, in the region of the B sequence, secondary structure is very similar. Alpha helix; A. Beta sheet; B. Turn, T. Coil: C.



Interestingly, unless we assume that there is only one possible solution to a given kinetic requirement, in this case phosphoryl transfer, even convergent sequences may be informative with respect to structure and function.

In addition, there are many representatives of related enzymes such as HPRT, OPRT, XPRT, UPRT, and some the phosphoribosyl transferases of *de novo* purine or amino acid synthesis. From the aligned sequences, several features are apparent by examination. In the region between residues 120 to 134 (using the numbering scheme for CHO APRT), strong sequence conservation can be identified, even though the flanking sequences appear to be completely unrelated. A consensus sequence for this domain is shown in Figure 2.4. The domain is usually referred to as the PRPP-binding domain, but this identification was originally based upon an inferred role for the region, rather than by direct biochemical or structural data.

Using aligned phosphoribosyl transferase sequences, it is possible to reconstruct a molecular phylogeny of this family. Since it is possible that there are convergences of function and sequence from multiple ancestral phosphoribosyl transferases, it is useful to be able to eliminate such convergences from consideration. It may be necessary to reduce sequences under consideration to the level of the conserved domains, since the rest of the molecule may simply be evolving too fast to make long evolutionary distance possible to measure. Unfortunately, phylogeny reconstruction using parsimony is error prone in the situation where there are more "species" than "characters" available for analysis. In order to facilitate phylogenetic analysis of the phosphoribosyl transferase family, alignments have been constructed using the entire sequences, and using reduced subsamples of the

sequences which have stronger alignments, but are shorter overall. The outcome of phylogenetic analysis of these global and reduced alignments were essentially similar.

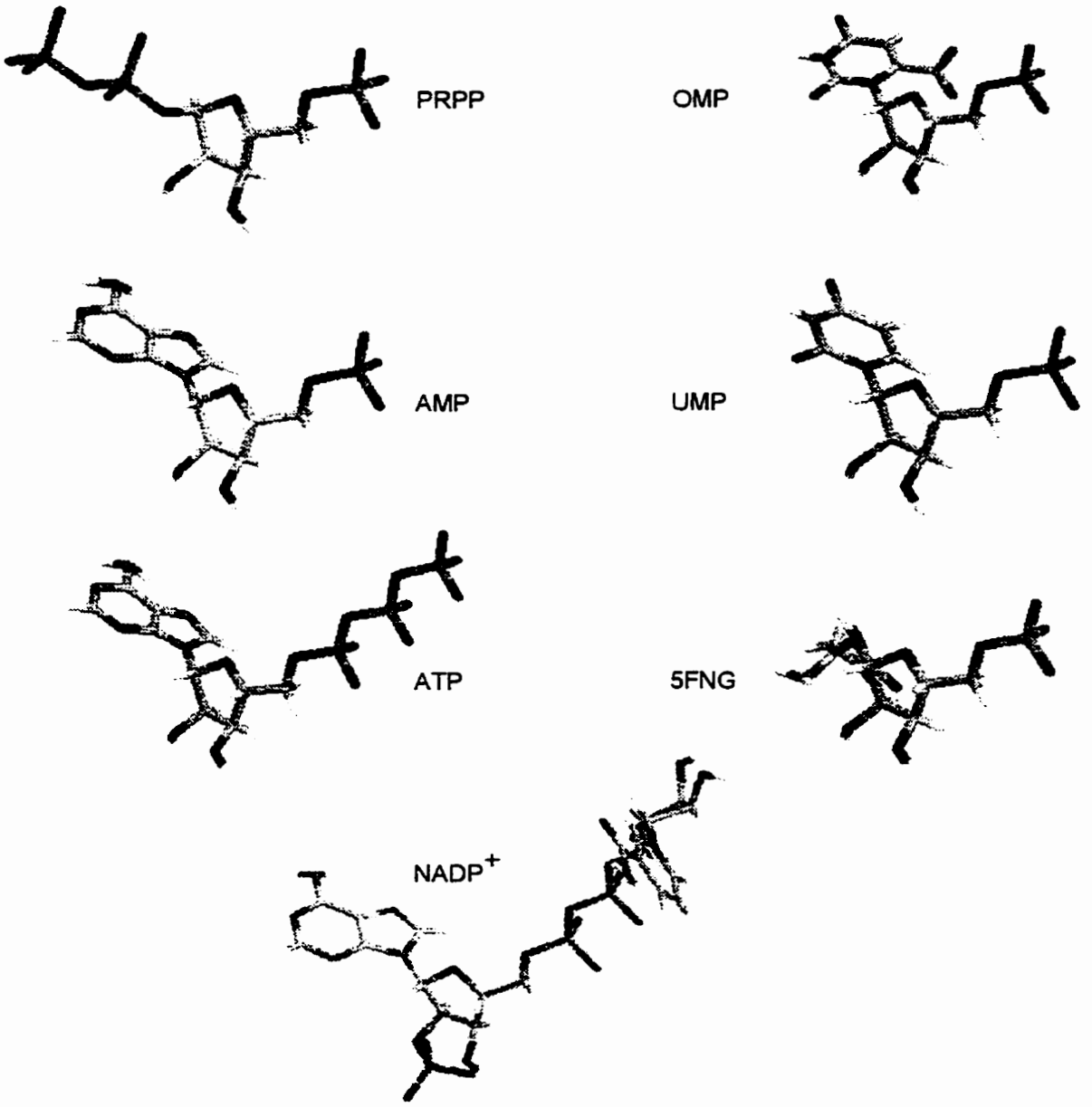
An essential starting point for structure/function analysis is to identify regions of protein sequence which are conserved among functionally equivalent forms of the molecule of interest. These conserved regions localize catalytically indispensable regions. Note that the human enzyme differs from the CHO enzyme by only eighteen amino acids out of 180, thus there is no reason to assume *a priori* that these two enzymes differ in any significant way with respect to tertiary structure. The *E. coli* enzyme however shares only about 20 percent identity and 45 percent similarity with the mammalian enzymes. Thus it is possible that the bacterial enzymes are structurally much different from the vertebrate types. Note also that the sequence of the *Arabadopsis thaliana* enzyme is much more homologous to the vertebrate type than to the bacterial types. It is probable that the main axis of divergence of this enzyme is at the prokaryote/eukaryote boundary. Combined with the observation of the wide distribtuion of APRT activivy throughout plants and archebatcteria, strongly suggest that APRT is an ancient enzyme, orgininating in a distant evolutionary past.

In Figure 2.4, one of the regions of the aligned APRT molecules is shown in detail. This corresponds to the so-called Walker B motif of the nucleotide-binding cassette, as characterized for adenylate kinase, myosin, and ATP-synthases (Walker *et al.*, 1982; Schulz *et al.*, 1986; Liang *et al.*, 1991). In APRT this domain is thought to bind PRPP, based on sequence alignment comparisons to the B domains of other phosphoribosyl transferases. Several reports have suggested that it interacts with the 1-pyrophosphate

moiety of the PRPP molecule (Tso *et al.*,1982; Argos *et al.*,1982;). But nucleotide binding proteins which do not bind PRPP also carry this sequence (Fry *et al.*, 1986; Gorbalenya and Koonin, 1989; Parsell *et al.*, 1991). In addition, the shape and structure of the attached base may vary widely in the case of nucleobase substrates of phosphoribosyl transferases, thus one would not predict that the B sequence interacts strongly with the base substrate. It seems unlikely that this highly conserved motif would serve multiple binding functions, even across species boundaries.

In the structures of the nucleotides which bind to proteins carrying the B motif, the 5'-α-phosphate and the ribose ring are consistent structural features (Figure 2.10). These features are virtually invariant in nucleotides. Since the phosphoribosyl transferase enzymes have similar functions, and the B sequence is very conserved across large evolutionary distances, including the sequences of proteins with different functions from phosphoribosyl transferases, but which also bind nucleotides, it is reasonable to infer that the B motif serves a consistent, but biologically unique function, in most of these molecules. That function should bear on a consistent functional requirement of the divergent proteins, which would appear to be an interaction between the side-chains of the B motif, and either the 5'-α-phosphate or the hydroxyl groups on the ribose ring. Since the conserved core of the B motif is aspartate or glutamate, with negatively-charged carboxyl as their primary functional groups, an intermediate, such as a divalent metal ion must exist between the amino acid carboxyl and the like-charged phosphate or hydroxyl groups. Nucleotide-binding proteins are dependent *in vivo* on the presence of divalent

Figure 2.10: Comparison of the structures of some nucleotides which bind to nucleotide-bnding molecules, including phosphoribosyl transferases at or near the B sequences. 5FNG; 5-phosphoribosyl-N-formylglycinamidine.



metal cations (magnesium in the case of phosphoribosyl transferases). This hypothesis has not until recently, been supported by solved crystal structures of nucleotide binding proteins, which have usually revealed no intimate relationship between the B-motif and the hydroxyl or 5'-phosphate. A new crystal structure for OPRT from *Salmonella*, the first with bound PRPP rather than with bound nucleotide, reveals that the hydroxyl groups of PRPP do indeed lie close enough to the side chains of the conserved aspartates to interact (Scapin *et al.*, 1993, 1995), whereas those of bound OMP do not.

From sequence comparison and phylogenetic analysis of the phosphoribosyl transferase family, it is clear that the sequences of the different family members are quite divergent. Such patterns of divergence are considered indicative of a sequence which is more useful for evaluating recent evolutionary relationships rather than more ancient ones (Kagan *et al.*, 1997;). It is interesting that these sequences cluster according to substrate, rather than according to species, phylum, or kingdom. This is what might be anticipated if the different (by substrate) family members had limited sequence similarity (Brown *et al.*, 1997). While the various alignment or phylogenetic approaches have limitations, the sequence similarity of the aligned data is apparent by examination, and it is unlikely that the result is arbitrary. In addition, the same phylogeny arises when the most conserved elements of the sequences, the B motifs, are aligned and separately analyzed. Thus we infer that the clustering of phosphoribosyl transferase sequences according to substrate is not merely an artefact of low sequence similarity. Rather, this sorting by substrate implies that there were already several substrate-specific members of the phosphoribosyl transferase family extant in the last common ancestors of mammals, bacteria, and plants. Such an

approach has been recently exploited in an analysis of tyrosine and tryptophan tRNA sequences (Brown *et al.*, 1997). Conservative estimates of the span of time involved range from 800 million to 1 billion years.

The divergence of sequences has implications beyond phylogenetic relationships. The folded structure of the active enzyme is an explicit function of the primary structure of the protein (for a review see Holzmann, 1995). Although there are still no reliable global methods for determining the final tertiary structure of any given sequence (Bohm, 1997; Rost and O'Donoghue, 1997), it is clear from functional similarities, and the few existing solved phosphoribosyl transferase crystal structures, that there must be strong overall three-dimensional similarities among these proteins (Busetta, 1988; Scapin *et al.*, 1995; Vos *et al.*, 1996). Yet these similarities of function and structure are founded upon sequences so divergent in some regions as to be apparently completely unrelated. From Figure 2.8, the secondary structural analysis of APRT from four species across three Phyla and two Kingdoms, clear structural similarities can be seen. The alternating alpha-beta-alpha theme seen in Chinese hamster APRT is well aligned to that in *E. coli*, and *Haemophilus*, but less well aligned to that of *Saccharomyces*, a eukaryote which is ostensibly much more closely related in evolutionary terms to Chinese hamster, than hamster is to *E. coli*. This suggests that the *Saccharomyces* sequences have diverged rapidly from other eukaryotic APRT sequences, which have retained their strucural similarity to the bacterial protein.

If we compare the secondary structural predictions of hamster APRT to human APRT, as shown in Figure 2.8, these are plainly more similar than those across distant

evolutionary boundaries. But as shown in Figure 2.9, the human HPRT molecule has a much different secondary structure relative to the APRT sequences. In the region of the B motifs however, the predicted structures of the proteins are very similar; a beta sheet flanked by short helical and turn segments. Somehow, this core of conserved sequence, actually too short to be useful in itself as a predictor of secondary structure, has a conserved structure across purine substrates (and hence across long evolutionary time).

Chapter 3 Expression of CHO APRT in *Escherichia coli*

3.1 Introduction

Adenine phosphoribosyl transferase is a member of a family of related enzymes which transfer the phosphoribosyl moiety of 5'-phosphoribosyl-1'-pyrophosphate (PRPP), to specific cosubstrates. This family includes the clinically important hypoxanthine-guanine phosphoribosyl-transferase, as well as enzymes of histidine, phenylalanine, and *de novo* purine biosynthesis (for a review of phosphoribosyl transferases see Musick, 1981). Deficiency of APRT in human beings has a highly variable clinical presentation, from asymptomatic, to potential renal failure due to accumulation of dihydroxyadenine stones (Fujimora *et al.*, 1985). APRT is also a potential target enzyme in the chemotherapy of certain neoplastic and parasitic diseases (Smith and Matthews, 1957; Queen *et al.*, 1989; Smolenski *et al.*, 1991).

Extensive biochemical work has described the overall chemistry of the reactions catalyzed by APRT and related proteins such as hypoxanthine phosphoribosyl transferase (HPRT)(Hori and Henderson, 1966a,b ; Gadd and Henderson, 1969; Ali and Sloan,1983). Further, some speculations about the implications of the observed inversion of stereochemistry about C-1 of the phosphoribosyl cosubstrate (Spector, 1982), and a proposed tertiary structure of the catalytic or nucleotide binding domain (Bussetta, 1988) have been published. De Boer and Glickman (1991) have reported a retrospective analysis of structure-function relationships implied by the residual APRT activity of mutants in cultured CHO cells. However, little is known about the molecular basis of the reaction, or about the nature of the determinants of specificity and catalytic activity of these enzymes.

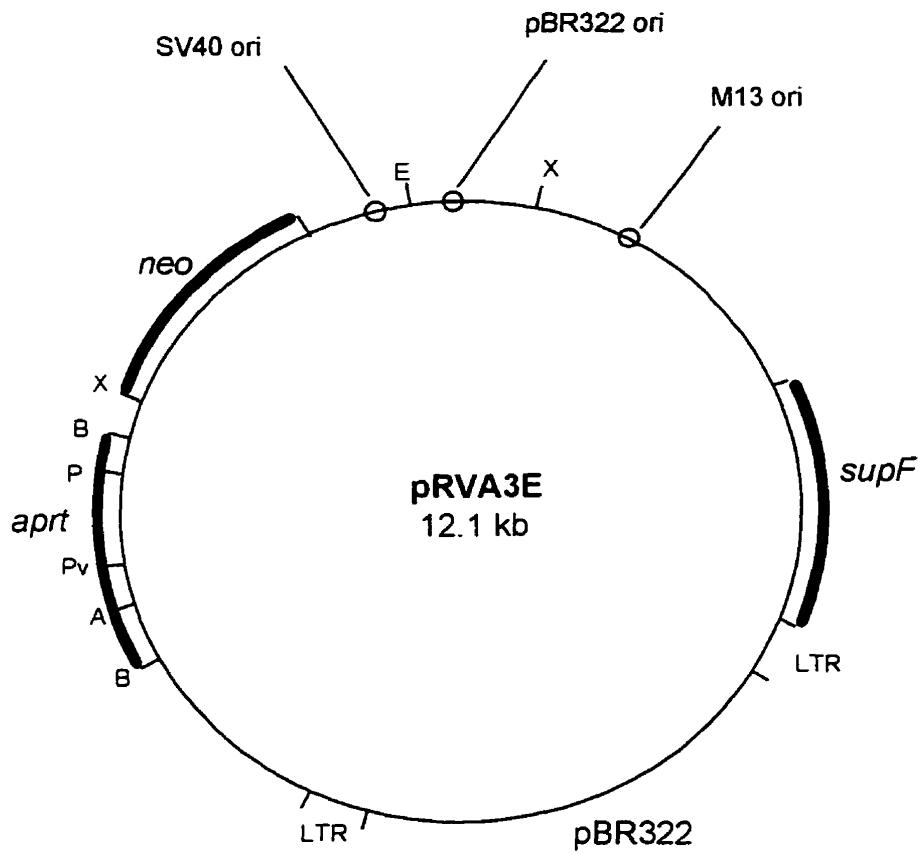
The goal of our study is to analyze structure-function relationships in APRT using an integrated approach which exploits biochemistry and molecular biology. This combined approach requires the ability to express native APRT, and to readily manipulate the gene *in vitro* for mutagenesis and cloning procedures. We have cloned a retrovirally-processed CHO *aprt* cDNA into a derivative of the bacterial expression vector, pKK223-3 (Pharmacia) under the control of the hybrid P_{tac} promoter (Amann *et al.*, 1983; Amann *et al.*, 1988) , and expressed the constructs in an *aprt*$^-$ *E. coli* host strain.

3.2 Materials and Methods

3.2.1 Isolation and characterization of CHO *aprt* from pRVA3E

Plasmid pRVA3E (Figure 3.1) a gift from Dr. Elliot Drobetsky, is a bacterial-mammalian shuttle vector based on Moloney murine leukemia virus (Cepko *et al.*,1984). From a genomic clone previously described, a 5kb BamHI fragment containing the entire aprt genomic sequence was cloned in to the MCS of pZIPNeo (Dr. P.de Jong, pers.comm.). After transforming into *E. coli*, appropriate constructs were transfected into CHO cell line D422, where the plasmid can be maintained by G418 selection. The *aprt* sequence has been retrovirally-processed, which causes removal of intronic sequences, but the poly-A signal and tail of a normal coding sequence or cDNA is absent (Drobetsky *et al.*, 1989). In addition, the CHO regulatory sequences for the *aprt* gene on pRVA3E are functional, such that *aprt* expression may be detected based on adenine utilization or analogue resistance in the mammalian host cells (*i.e. aprt* $^-$). These regulatory sequences are however, non-functional in *E. coli*, thus it is necessary to replace the mammalian

Figure 3.1 : Structure of plasmid pRVA3E, gift of Dr. E.A. Drobetsky. pRVA3E is a derivative of pZIPNeoSV(X), containing a retrovirally-processed *aprt* coding sequence from Chinese hamster. This vector is derived from the unprocessed plasmid pRVA3, constructed by Dr. P.J.de Jong, which contains a 2.4 kb BamHI fragment with the entire *aprt* genomic locus, including the poly-adenylation signal and approximately 500 bp of the upstream region. Restriction enzyme digestion sites are indicated as follows: A;ApaLI (3' end of coding sequence, single site in *aprt*, multiple unmapped sites in vector), B; BamHI, E;EcoRI, P;PstI, Pv;PvuII (marks 5' end of coding sequence), X;XhoI. ori;origin of replication. LTR;long terminal repeat of SV40. The *neo* gene confers resistance to Kanamycin in *E.coli*, and to G418 in mammalian cells. *supF* is a tRNA mutant suppressor useful in *E.coli*.



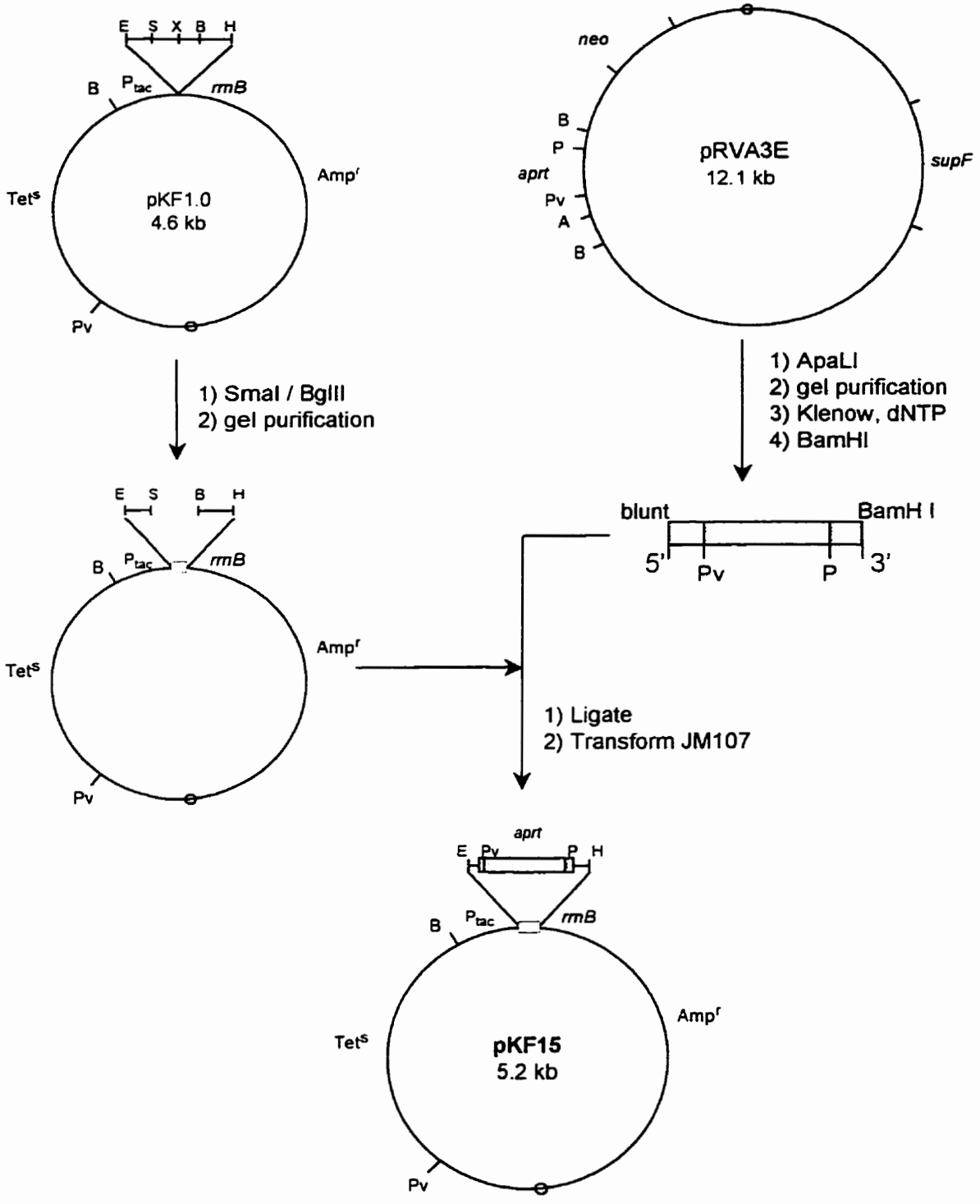
sequences with a controllable bacterial promoter-operator in order to modulate expression in *E. coli*.

Except where noted, the manufacturer's supplied buffers were used according to the package instructions for all enzymatic treatments. A sample of pRVA3E was digested with 10 U ApaLI, BamHI, or both enzymes. 6X agarose loading dye was then added to a final concentration of 1X (10% glycerol, 1 μg/ml xylene cyanol, 1 μg/ml phenol blue, 10 mM Tris HCl pH8.0, 1 mM EDTA). The samples, along with an undigested sample and a DNA sizing standard (1 kb ladder, Gibco-BRL) were loaded onto a 1.0% low EEO (electro endo-osmosis) agarose (Dalton Chemical Laboratories) containing 1X TAE and 10 μg/ml ethidium bromide and electrophoresed at 20 V/cm until the dye front marker reached the anode end of the gel. The gel was photographed with ultraviolet transillumination using a Polaroid MP-4 photodocumentation camera. The resultant photographs (and all others herein) were digitized on a Hewlett-Packard IIcx scanner for later use.

3.2.3 Cloning of CHO *aprt* into pKF1.0

An ApaLI-BamHI fragment from pRVA3E carrying the retrovirally-processed *aprt* coding sequence, was cloned into pKF1.0 (see Appendix I), as follows (see Figure 3.2). 20 μg pRVA3E was digested with 10 U ApaLI (New England Biolabs) at 37°C overnight, then electrophoresed in an 1% low EEO agarose gel (Dalton Chemical Laboratories) in 1X TAE buffer. The band containing the *aprt* sequence was excised and purified by crushing the gel by extrusion through a 25 gauge needle attached to a 1 ml syringe. The crushed gel was then soaked twice 1 volume of TE (10 mM Tris HCl, pH 8.0, 1 mM

Figure 3.2 : Cloning *aprt* from pRVA3E into pkF1.0 (derived from pKK223-3, see Appendix I), as described in text. A;ApaLI (3' end of coding sequence, single site in *aprt*, multiple unmapped sites in vector), B; BamHI, E;EcoRI, H; HindIII, P;PstI, Pv;PvuII (marks 5' end of coding sequence), S; SmaI, X; XbaI.



EDTA). The mixture was centrifuged after soaking, and the supernatants pooled into a fresh tube. 5M NaCl was then added to a final concentration of 50 mM, followed by two volumes of 95% ethanol. The suspension was mixed by inversion, and centrifuged at 14,000 x g for 30 minutes. The supernatant was decanted, and the pellet briefly dried at room temperature. The pellet was then resuspended in 10 µl TE. The isolated fragment was treated with 10 U of Klenow fragment (Pharmacia) in the manufacturer's buffer, at 37 °C for 20 minutes, in the presence of 1mM dNTPs, in order to fill-in the ApaLI ends. This was followed by a 10 minute incubation at 65 °C to inactivate the Klenow enzyme. The mixture was digested with 10 U BamHI overnight at 37 °C, followed by 10 minutes at 65 °C.

In parallel, 10 µg pKF1.0 was digested with 10 U each of SmaI and Bgl II (New England Biolabs), in a total volume of 50 µl, at 37 °C overnight, followed by a 10 minute incubation at 65 °C. The Bgl II cut site provides a complementary sequence overlap to that of a BamH I cut site, but after ligation, the resultant sequence is resistant to cutting by either BglII or BamHI. The blunt end-BamHI fragment from pRVA3Ewas ligated into the digested pKF1.0, by mixing 10 µl of prepared pRVA3E fragment, with 20 µl of the pKF1.0 digest in a fresh tube. 5 µl of 10 mM ATP were added, followed by Nanopure water to a volume of 48 µl. 2 µl (about 20 U) of T4 ligase (New England Biolabs) was added, and the reaction incubated at 15 °C overnight.

The ligation products were transformed into *E. coli* B2325 (*HfrH thi lac deoD8 purD apt110*, derived from K12, Kocharyan *et al.*, 1978) following the method of Hanahan (1983) and plated onto solid LB-ampicillin (10g/l tryptone (Gibco), 5 g/l yeast

extract (Gibco), 10g/l NaCl, 15 g/l agar (Difco), 100 μg/ml ampicillin (ICN)). Several isolated ampicillin-resistant clones were picked and cultured overnight in liquid LB-ampicillin. From these cultures plasmid DNA was isolated and purified using the alkaline lysis method (Birnboim and Doly, 1979). The plasmid DNA was then digested with EcoR I and Hind III as a preliminary screen for the presence of an insert the desired insert.

3.2.3 Characterization of pKF15

Based on preliminary insert sizing estimates, one clone, pKF15, was chosen for detailed characterization by extensive restriction site mapping analysis, followed by Southern blotting (Southern, 1975) to Hybond-N membrane (Amersham) and hybridization with an *aprt* probe obtained from Dr. Mary Mazur-Melnyk, and finally by DNA sequencing (Sanger *et al.*, 1977).

The *aprt* probe was labeled using a random-priming kit (Boehringer Mannheim), according the manufacturer's instructions. Prehybridization of the filter was done at 65°C in a buffer of 1% SDS, 10% dextran sulphate, 1 M NaCl. Hybridization to the heat-denatured probe was done at 65°C in the same buffer as used for the prehybridization. After 18 hours, the membrane was washed twice in 6 X SSC (20 X: 3.0M NaCl, 0.3M NaCitrate), 1% SDS at room temperature, twice in 0.2 X SSC at room temperature, once in 0.2 X SSC at 65 °C, and finally in 0.1 X SSC at room temperature. The membrane was placed into an autoradiography cassette (Fisher Scientific) with Kodak X-AR film for 24 hours at -75 °C. In addition, the promoter end of the insert was sequenced by the dideoxy method (Sanger, 1977), using a primer complementary to an internal segment of the promoter region, 40 bp upstream of the EcoRI site, to verify the promoter-ATG distance

and the 5' end of the inserted sequence. The sequence of this oligonucleotide was 5'-TGTGACA ATTAATCATCGGCTCGTATAATGTGTG-3'.

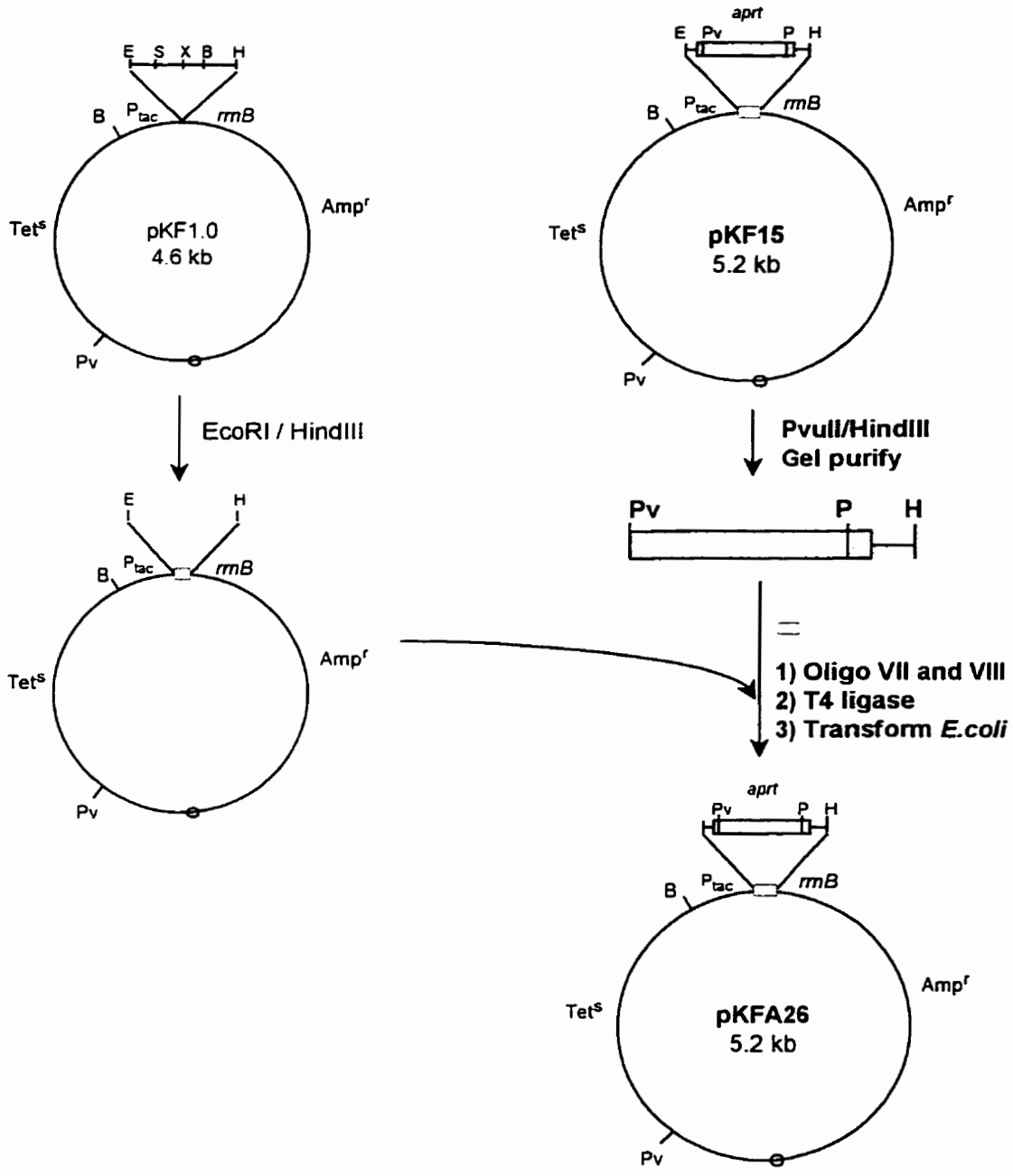
3.2.3 Modification of pKF15 for improved expression

From pKF15, the PvuII-HindIII fragment containing most of the *aprt* coding sequence was subcloned into pKF1.0 digested with EcoRI and HindIII, as in Figure 3.3. pKF15 was digested with PvuII and HindIII (New England Biolabs) overnight at 37 °C. The approximately 600 bp fragment containing most of the *aprt* sequence was gel-purified as previously described, and resuspended in a final volume of 10 µl TE. 10 µg pKF1.0 digested with EcoRI and HindIII using the usual method, was ethanol precipitated and resuspended in 40 µl TE. 20 µl of this was mixed with 10 µl of the PvuII-HindIII digest of pKF15, along with 50 µl of a 1:1 mixture of Oligo VII and Oligo VIII, approximately 25 ng each (see Figure 3.3 for the oligo sequences). 5 µl 10 mM ATP was added, followed by Nanopure water to a volume of 48 µl. 2 µl (about 20 U) of T4 ligase were then added, and the entire reaction incubated at 15 °C overnight. Ligation products were transformed into B2325 competent cells, and plated onto LB-ampicillin as described above. Ampicillin-resistant clones were picked and grown in liquid LB overnight, and plasmid DNA prepared by alkaline lysis. Plasmid DNA was digested with BamHI and HindIII, then electrophoresed to detect an insert of about 600 bp.

3.2.4 Characterization of pKFA26

Among ampicillin-resistant clones resulting from this experiment, the construct pKFA26 was shown in preliminary size screening to have an insert of approximately 600 bp size. pKFA26 was further characterized by diagnostic restriction digests with BamHI,

Figure 3.3 : Construction of pKFA26 from pKF15. The *aprt* coding sequence from pKF15 was subcloned as shown. The final sequence of the promoter-ATG codon region is shown at the bottom of the figure. Sequences of olignucleotides VII and VIII are shown in italics. * denotes the location of the PvuII blunt cut site. The 3' end of the promoter and the ATG codon are in bold.



Final Promoter-ATG codon structure:

. . . CAC**AGGA**AACAG*AATT**ATG**GCGGAATCTGAGTTGCAG*CTG . . .

. . . GTGTCCTTTGTCTTAA*TACCGCCTTAGACTCAACGTC*GAC . . .

HindIII, PvuII, EcoRI, and double digests with BamHI-PvuII, HindIII-PvuII, and BamHI-HindIII. Digests were electrophoresed and photodocumented as described above. The structure of the promoter-ATG region was subsequently confirmed by dideoxy sequencing (Sanger *et al.*, 1977).

3.3 Results

3.3.2 *aprt* from pRVA3E

The a*prt* cDNA contains no ApaLI site, thus the fragment containing the entire coding region must be larger than 540 bp (the size of the full-length *aprt* sequence). The BamHI digest of pRVA3E should release a fragment of approximately 1.1 kb, since the introns of the *aprt* gene have been removed by passaging through the mammalian host cells. From diagnostic restriction digests of pRVA3E, with BamHI, ApaLI, and BamHI-ApaLI double digests, it was determined that the *aprt* coding sequence, contained on the 1.1 kb BamHI fragment, would be contained within the band of approximately 600 bp size, as indicated in Figure 3.4 (see Figure 3.5 for results of a Southern blot including pRVA3E cut with BamHI). This band was subsequently excised and gel-purified for use in subcloning the *aprt* sequence into pKF1.0.

3.3.1 Structure of pKF15

Among the ampicillin-resistant clones, pKF15 was found to have an insert of appropriate size. A series of restriction enzyme digests was performed and the results shown in Figure 3.5. Note the size of the fragment cut from the MCS by BamHI and HindIII. Note also that pKF15 is restricted by EcoRI as indicated in Lanes E and E/H. From this gel, a Southern blot (Southern, 1975) was performed using a full length

Figure 3.4 : Restriction digest of pRVA3E. U; uncut pRVA3E. A; ApaLI digest. A/B; ApaLI/BamHI double digest. B; BamHI digest. Along the right side is indicated the size, in kilobase pairs, of the DNA sizing standard (1 kb ladder, BRL).

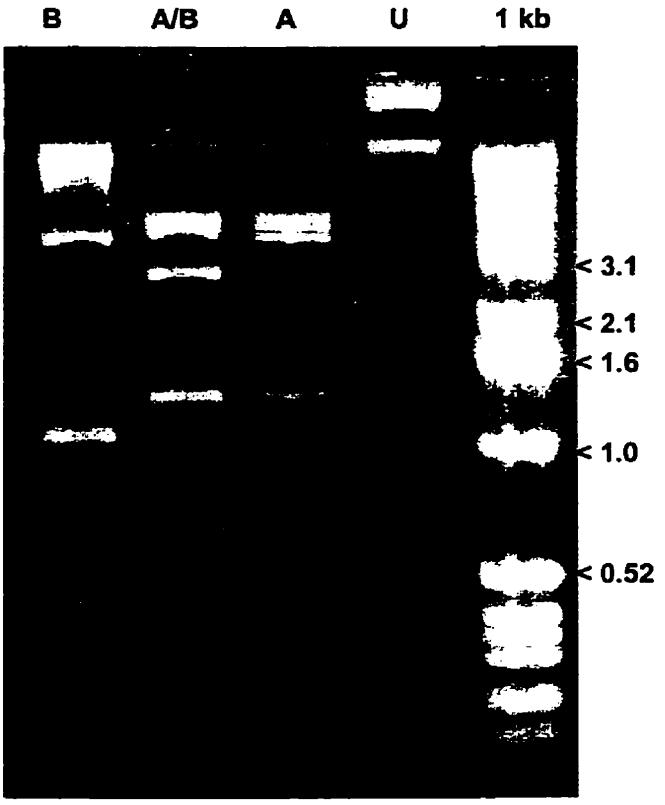
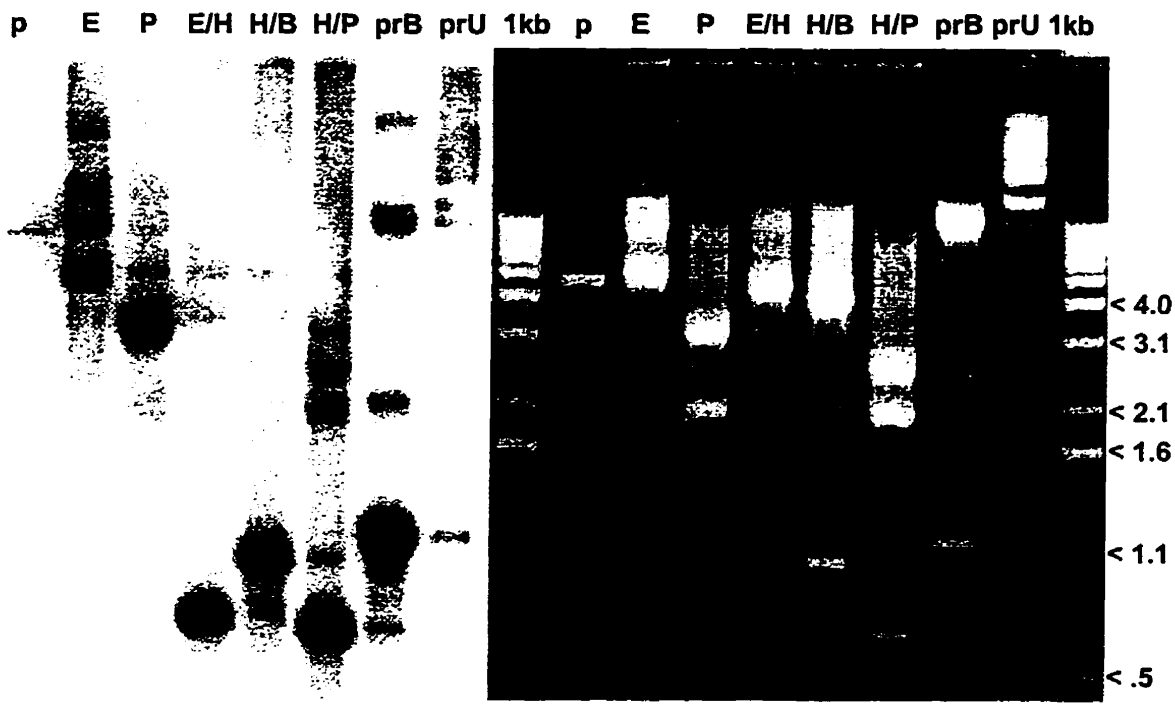


Figure 3.5 : Diagnostic restriction digests and Southern blot hybridization of pKF15. On the left frame is the autoradiograph of the Southern hybridzation, on the right, the agarose gel from which the blot was taken. Along the right edge is indicated the sizes in kilobases of the DNA sizing standard fragments from the 1 kb standard (BRL). p; undigested pKF1.0. Digests of pKF15: E;EcoRI. P; PvuII. E/H; EcoRI/HindIII double digest. H/B; HindIII/BamHI double digest. H/P; HindIII/PvuII double digest. prB; pRVA3E BamHI digest. prU; undigested pRVA3E.

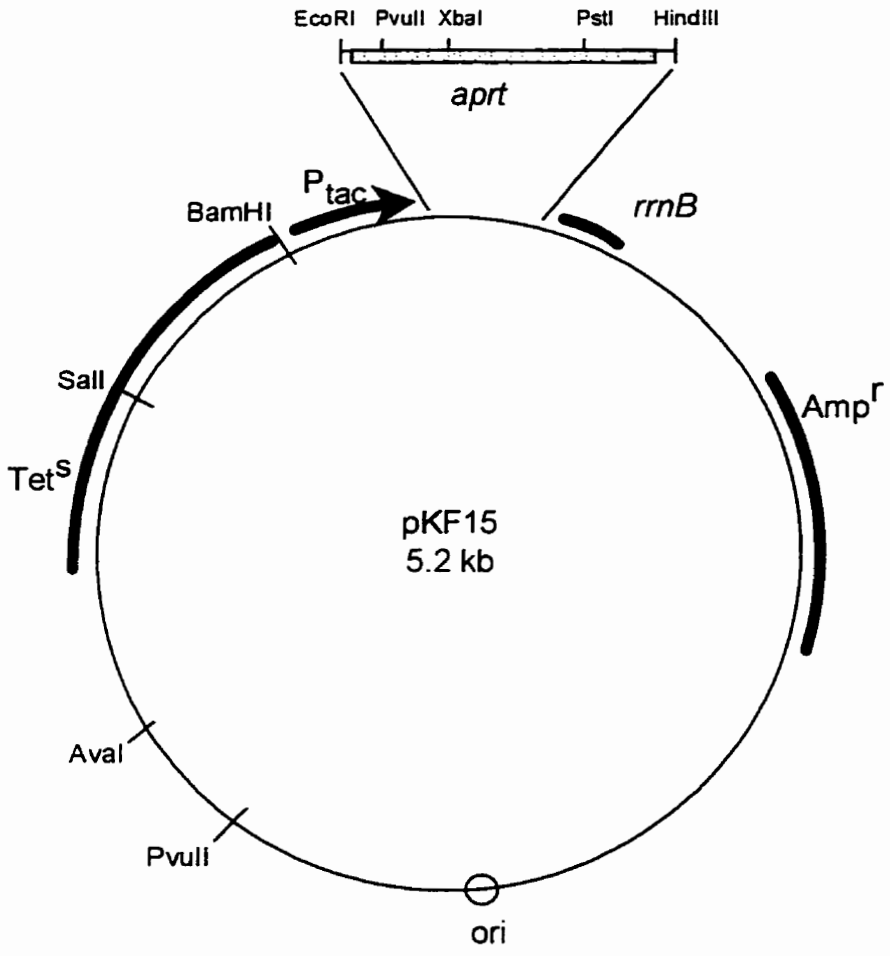


pRVA3E *aprt* fragment as a probe. Bands from the gel which showed hybridization correspond to sizes anticipated for a full-length aprt coding sequence insertion. Finally, the structure of the EcoRI junction at the 5' end of the coding sequence was verified by dideoxy sequencing (Sanger et al.,1977). The final promoter to ATG codon spacing was 15 basepairs. The final structure of pKF15 is shown in the schematic diagram in Figure 3.6

3.3.2 Characterization of pKF15

Once the structure of the clone pKF15 was confirmed, efforts to amplify the plasmid and induce APRT activity were attempted. pKF15 was able to express APRT at levels sufficient to allow growth of the host *aprt*$^-$ strain B2325 (Kocharyan *et al.*, 1975), which is *aprt*$^-$ and defective in *de novo* synthesis of purines. On adenine-containing medium, in the presence of 1mM IPTG. Using a previously described APRT assay (Okada et *al.*, 1986), it was determined that the plasmid-encoded APRT expression was somewhat lower than the endogenous APRT activity of the *aprt*$^-$ parent *E. coli* strain, B796. All plasmid-bearing (ampicillin-resistant) hosts were also able to grow on M9-adenine in the presence of IPTG, but activity in crude extracts was not significantly inducible by IPTG in rich medium at 37 °C. This level of expression was sufficient to demonstrate that Chinese hamster ovary APRT could functionally replace the *E. coli* enzyme, but was not sufficient to facilitate expression and isolation of novel mutant proteins. The limitation of expression was partly the result of a somewhat long (15 bp) Shine-Delgarno sequence to initiation codon spacing in pKF15, as shown in Figure 3.5. Preliminary SDS-PAGE (Laemmli, 1970) gels failed to reveal any significant new band of protein in comparison to vector-only controls.

Figure 3.6 : Structural diagram of pKF15, a derivative of pKF1.0 (see text) containing the CHO *aprt* coding sequence under the control of the P_{tac} promoter. Note the presence of the EcoRI site just 5' to the coding sequence. In pKF15 the promoter-ATG distance is 15 bp.



3.3.4 Structure of pKFA26

In response to the poor inducibility and low overall yield of APRT from pKF15, it was decided to modify pKF15 in order to increase expression levels if possible. pKF15 was modified as described in Methods. From the subsequent transformation, the construct pKFA26, among others, was identified and fully characterized as to restriction enzyme site pattern, as shown in Figure 3.7. The EcoRI site is removed as demonstrated in Lane E. The sequence of the promoter to ATG region was also verified, also indicated in Figure 3.7. The final promoter-ATG spacing was 8 bp, within one basepair of the distance empirically determined to be ideal for expression under the control of the hybrid P_{tac} promoter (de Boer *et al.*, 1983; Amman *et al.*, 1988). A schematic of the final structure of pKFA26 is shown in Figure 3.8

3.3.5 Characterization of pKFA26

After verification of the structure of pKFA26, the plasmid was grown under rich medium conditions in LB at 37 °C. Under these conditions, APRT activity was inducibl by IPTG at 1 mM, yielding crude extract activities about 3-5 fold greater than that observed with pKF15 under the same conditions. In an attempt to further enhance yield, the plasmid was amplified with chloramphenicol according to the method of Neidhardt *et al.* (1988). Under these conditions no viable cells could be subcultured from the culture flask. In contrast, plasmids from host cells with pKF15 or the pKF1.0 vector, could be amplified with chloramphenicol and remained viable.

An alternative method of amplification, amino acid supplementation in the presence of ferric chloride (Angelov and Ivanov, 1989) proved effective, with final yields

Figure 3.7 : Agarose gel and Southern blot hybridization of pKFA26. The agarose gel on the right is in register with the Southern hybridization autoradigraph on the ieft. Sizes of fragments in the 1 kb ladder sizing standard are indicated along the right-most edge. The restriction enzyme digestions performed were: U; uncut . B; BamHI. H; HindIII. P; PvuII. E; EcoRI. BP, HP, and BH, are double digests with the indicated enzymes.

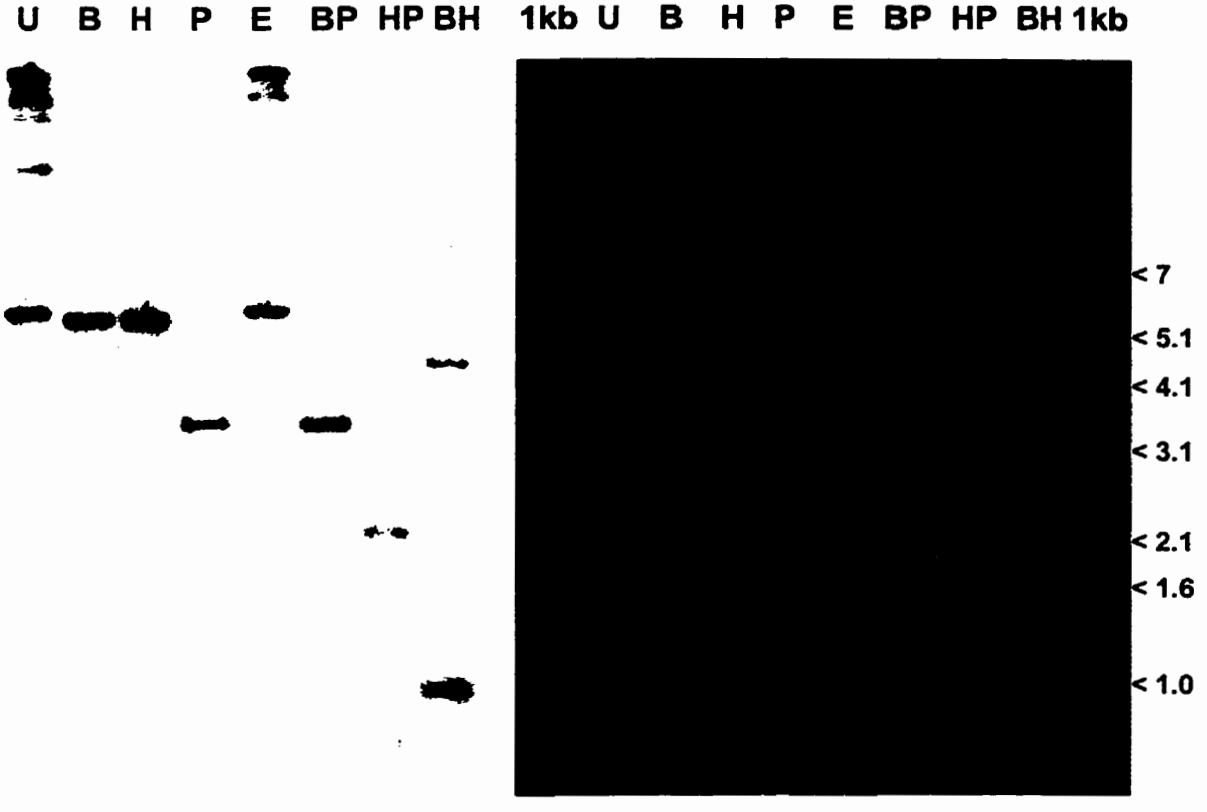
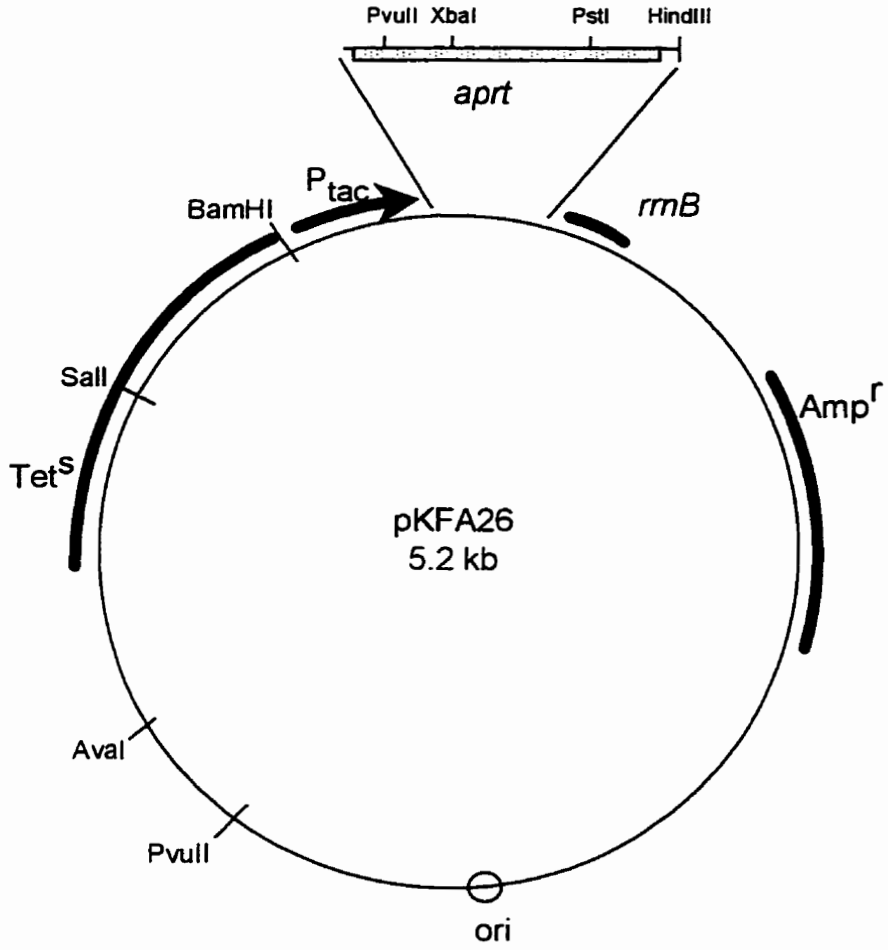


Figure 3.8 : Structural diagram of pKA26, a derivative of pKF15 (see text) containing the CHO *aprt* coding sequence, with a reduced S/D-ATG spacing relative to pKF15. The EcoRI restriction site just 5' to the ATG codon is no longer present, so that the S/D-ATG distance is 8 bp.



abut 15-30 fold higher than those observed with pKF15. Some variability in yield was seen using this method. It was also observed that different brands or lots of tryptone and yeast extract could also affect yields substantially. SDS PAGE gels of the crude and clarified extracts showed a new band at the molecular weight corresponding to 20 kilodaltons, the approximate molecular weight of the CHO APRT monomer.

3.3.5 Kinetics of APRT from pKFA26

Using the assay method of Okada, gross activity and kinetic parameters were evaluated for APRT (pKFA26) purified from rich medium cultures. Double-reciprocal plots of the adenine and PRPP kinetics data, mean of three replicates, is shown in Figures 3.9 and 3.10 respectively. The K_M for adenine, 9.1 μM, in the presence of saturating MgPRPP, is in reasonable accordance with reported values for this enzyme in the literature (Hershey and Taylor, 1978; Arnold and Kelley, 1978). Differences in reported values are likely attributable to differences in the source of the extracts tested (*e.g.* erythrocyte versus liver versus recombinant product) and the mode of analysis. The K_M for PRPP, 3.7 μM, is also close to reported literature values. A chart of the results, and values from previous publications is shown in Table 3.1. Notable is the large difference in K_M of the *E. coli* APRT for PRPP, which is more than 10-fold greater than that reported for mammalian cells (Hochstadt-Ozer, 1971).

3.4 Discussion

In order to readily generate mutations and characterize their effects in the Chinese hamster *aprt* gene, a bacterial cloning and expression system more amenable to manipulation than that of the native mammalian cells was developed. Mammalian tissue

Figure 3.9 : Lineweaver-Burke plot of velocity versus concentration of adenine for APRT (pKFA26). K_M for adenine from this example is approximately 9.1 μM.

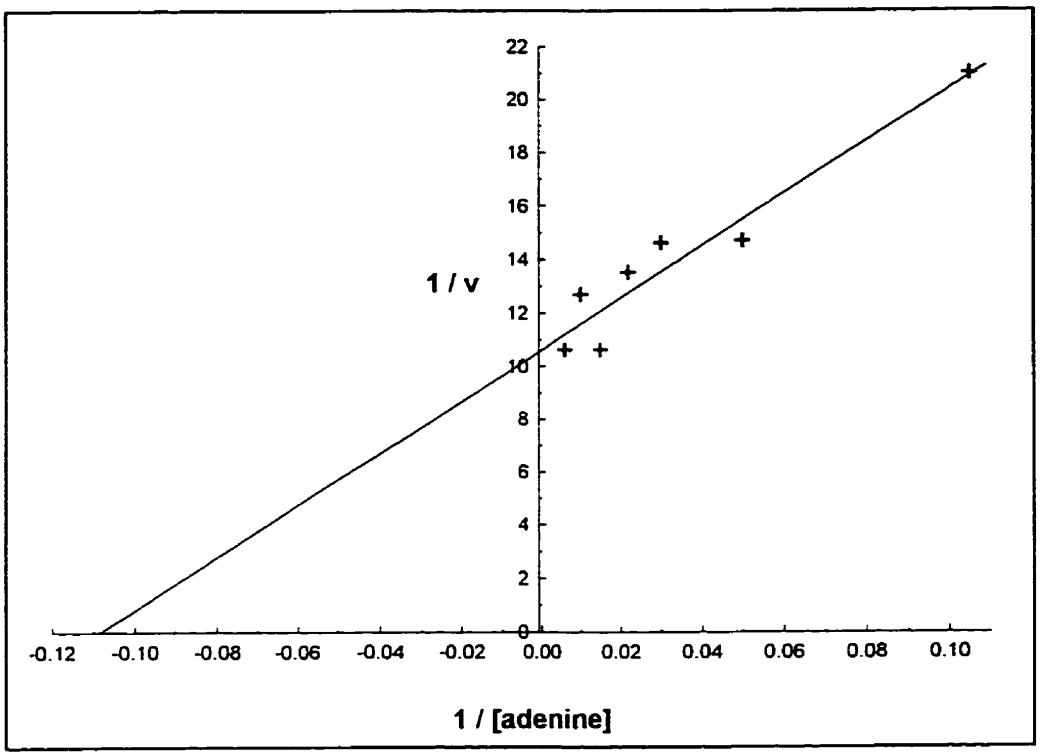


Figure 3.10 : Lineweaver-Burke plot of velocity versus substrate PRPP. Km for PRPP from this plot is approximately 3.7 μM. Assay method described in text.

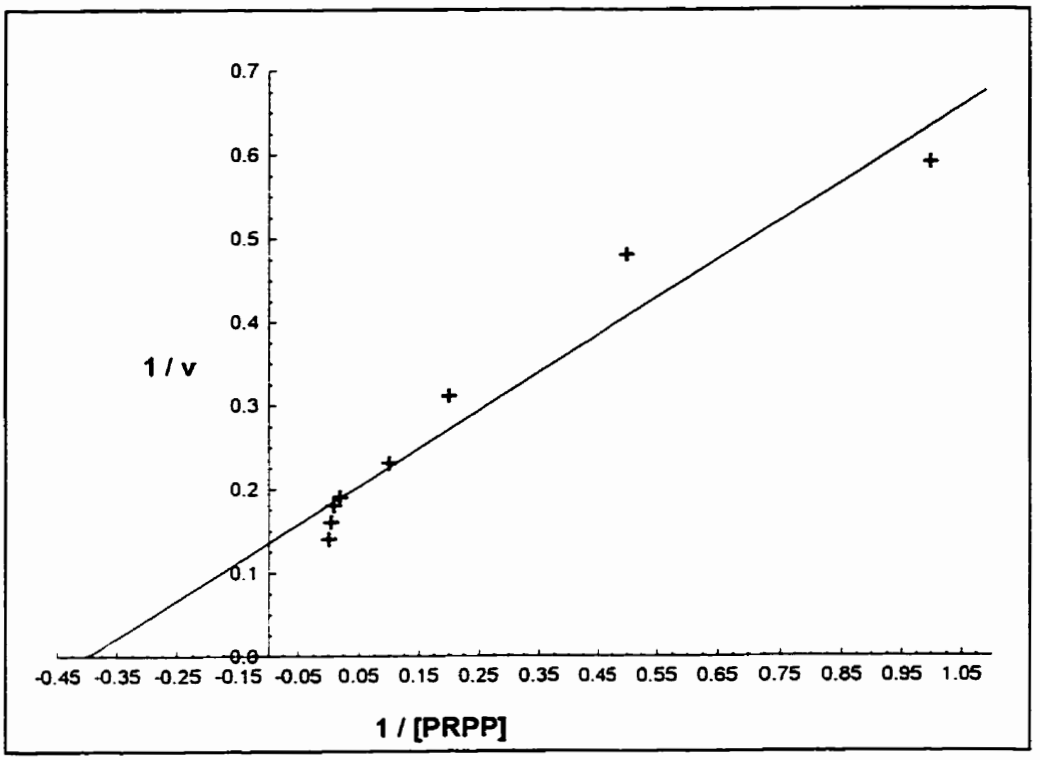


Table 3.1 : Kinetic parameters of APRT extracts from *E.coli* carrying pKFA26, and from reports in literature of the native enzyme from other sources. Mouse, rat, and human APRT differ by fewer than 5 amino acids from CHO APRT.

Source	Km		Reference
	adenine	PRPP	
pKFA26	3.7 ± 1.2 μM	9.1 ± 2 μM	this work
CHO erythrocytes	1.0	3.0	Hershey and Taylor, 1978
Mouse FM3A cells	6.6	1.2	Okada *et al.*, 1986
Rat liver	1 ± 0.3	5 ± 2	Kenimer *et al.*, 1975
Human erythrocytes	1.4-2.7	6.3-9.3	Dean *et al.*, 1968
Human erythrocytes	1.0-2.6	3.6-8.1	Henderson *et al.*, 1968
E.coli	10	150	Hochstadt-Ozer, 1971

culture is difficult and expensive, and it is difficult to isolate most proteins in quantity. Further, there is evidence that mutations lead to a detectable APRT-deficiency phenotype, which does not always engender complete loss of enzyme activity. The phenotype observed is due not to loss of enzyme activity *per se*, but rather to the presence of an altered form of the enzyme with biochemical properties different from the wildtype APRT. The phenotype of analogue resistance in these cases may also be due to rapid degradation of the mutant protein relative to the wildtype, or to reduced expression of mutant sequences due for example, to deviations from optimal codon sequences.

It is also possible that some mutations confer an intermediate level of function, which is not detected as a wildtype phenotype under typical selective conditions. Modulation of selective agent concentration has been previously used to allow selection of heterozygous *aprt*$^{-}$ *aprt*$^{+}$ cells (Adair *et al.*, 1980). The effect of reducing selective agent concentration changes is to change the concentration of cytotoxic nucleotide analogue produced, thus allowing survival of cells with reduced levels of APRT activity. By extension, altering the intracellular concentration of nucleotide analogue by altering the properties of the enzyme may allow rescue of mutants with residual APRT activity. It is likely that a single set of selective conditions within a single experiment may exclude some mutants with residual function.

The pKFA26 heterologous expression system in *E. coli* using relatively inexpensive media and laboratory methods is cost-effective. In addition, cloning and subcloning in *E. coli* are well-described, along with readily available tools for mutagenesis and mutant characterization. Chinese hamster adenine phosphoribosyl transferase does

not have significant post-translational modifications in its native environment nor does the native protein have intrachain or interchain disulfide bonds (Hochstadt-Ozer and Stadtman, 1971; Kenimer *et al.*, 1975; Holden *et al.*, 1979; Okada *et al.*, 1986). In such a system, using suitable amplification methods, even poorly stable or marginally functional mutant enzymes can be detected and isolated.

Chapter 4 The putative nucleotide binding cassettes of APRT

4.1 Introduction ; Is there a generalized nucleotide binding cassette in APRT?

Current models of enzyme structure and function suggest that domains of strong conservation probably represent tracts in which amino acid residues are under selective pressure. The substrate binding domain of many nucleotide-binding proteins contain tracts of highly conserved residues, constituting the nucleotide binding cassette (NTBC). These regions, called the A and B domains, corresponding to their proximity to the 5' end of the gene, have been described in several reviews of nucleotide binding proteins (Rossmann *et al.*, 1974; Walker *et al.*, 1982; Saraste *et al.*, 1990; Bork *et al.*, 1992). Figures 4.1 (A sequence) and 4.2 (B sequence) are alignments of these regions from several phosphoribosyl transferase sequences. Particularly in the case of the B sequences, it is clear by examination of the highly conserved residues, that this region has not changed significantly through long evolutionary times. The divergence of the Eukaryota from Prokaryota is estimated at 1.6-1.8 x 10^9 years (Hori and Osawa, 1979), yet the core of this region, amino acid residues 125 to 133 of the CHO APRT sequence, is essentially unchanged. Beyond this region these proteins are divergent (see Figure 2.2), to the extent that conventional computerized algorithms cannot easily align the sequences.

Figures 4.3 and 4.4 represent alignments of some of the phosphoribosyl transferases to nucleotide-binding protein sequences in the regions most similar to the A and B motifs of Figures 4.1 and 4.2. These are not exhaustive, but demonstrate that there is strong sequence similarity in these domains among and between nucleotide binding enzymes. The consensus sequences are hypothesized to bind phosphate and nucleobase or

Figure 4.1 (1 of 2) Alignment of primary sequences in the A region of phosphoribosyl transferases from various sources. Highly identical residues (greater than 80%) are indicated with a **x**, moderately indentical (50-80%) are indicated *. Finally, residues with considerable conservation of character (eg. hydrophobicity) but not necessarily of sequence, are indicated with ! . Numbering follows the pattern for Chinese hamster APRT.

Figure 4.1 (cont. 2of 2)

```
                                                    70         80
                                                   !!*!   xx   x! *!! ! !!  !
Cricetulus longicaudatus                APRT       IAGLDS-RGFLFGPSLAQELGLGCVLI
Mus pahari                                         IAGLDS-RGFLFGPSLAQELGVGCVLI
Mus spicelegus                                     IAGLDS-RGFLFGPSLAQELGVGCVLI
Rattus norvegicus                                  IAGLDS-RGFLFGPSLAQELGLGCVLI
Homo sapiens                                       IAGLDS-RGFLFGPSLAQELGLGCVLI
Drosophila melanogaster                            IVGLDS-RGFLFNLLIATELGLGCAPI
Saccharomyces cerevisiae I                         IAGIEA-RGLLFGPSLALALGVGFVPI
Triticum aestivum                                  VAGVEA-RGFIFGPPIALAIGAKFVPI
Synechocystis sp.                                  VVGMES-RGFLFGMPLAYQMNAGFIPV
Escherichia coli                                   VVGTEA-RGFLFGAPVALGLGVGFVPV
Pseudomonas aerogenes                              IGAMDA-RGFLIGSAVAYALNKPLVLF
Haemophilus influenzae                             VLGTES-RGFIFGAPVALALGLPFELV
Streptomyces coelicor                              VVGLEA-RGFILGAPVALRAGLGFIPV
Mycobacterium tuberculosis                         EFKDDA-RGFLVAAAVATRLEVGVLAV
Caenorhabditis elegans                             VAGLEA-RGFLFGPQVAIQLGVPFVPI
Leishmania donovani                                ILGFDA-RGFLFGPMIAVELEIPFVLM
Mycoplasma pneumoniae                              ---PEA-RGFIFGGALASKTKLPLVLV
Methanococcus jannaschii                           LVTAEA-MG---IPLVTTLSLYTDIPY
Homo sapiens                            HPRT      -VALCVL-KGG--YKFFADLLDYIKALN
Cricet. longicaudus                                -VALCVL-KGG--YKFFADLLDYIKALN
Schistosoma mansonii                               -TLMCVL-KGG--FKFLADLVDGLERTV
Plasmodium falciparum                              -HILCLL-KGSRGFFTALLKHLSRIHNY
Trypanosoma brucei                                 -VIVSVL-KGS--FVFTADMVR-ILGDF
Leishmania donovani II                             -YLLCVL-KGS--FIFTADLAR-FLADE
Escherichia coli                                   -VLVGLL-RGS--FMFMADLCR-EVQV-
Vibrio harveyi                                     -VMVGLL-RGS--FVFMADLAR-AIEL-
Rhodobacter capsulatus                             -VVVGLL-RGS--FVFIADLIR-EIGV-
Homo sapiens                            OPRT       -VCGVPY-TA---LPLATVICS-TN-QI
Trichoderma reesei                                 -VFGPAY-KG---IPLCSAITIKLGDVA
Dictyostelium discoides                            -QETPALVCG---VPYTALPIA-TGMSI
Sacc. cerevisiae 10                                -IFGPAY-KG---IPLATAATIKLGQIR
Podospora anserina                                 -VFGPAY-KG---IPLAAIVCVKLAEIG
Sordaria macrospora                                -IFGPAY-KG---IPLATAATDKLAQLD
Coxiella burnetii                                  -VFGPAY-KG---IPLATSTTDKLAELD
Escherichia coli                                   -LFGPAY-KG---IPIATTTAVALAEHH
Lactobacill. plantarum                             -IAGVAT-AG---IPHAAWVAE-LL-NL
Bos taurus                              UPRT       -VCGVPY-TA---LPLATIVCS-TH-EI
Drosoph. melanogaster                              -VCGVPY-TA---LPRATIVSV-QQ-GT
```

Figure 4.2 (1 of 2): Aligned phosphoribosyl transferase primary structures in the region of the so-called Walker B motif. Sequence numbering is according to the scheme for CHO APRT. Key to proteins : APRT; adenine phosphoribosyl transferase (PRT). HPRT; hypoxanthine PRT. OPRT; orotate PRT. UPRT; uracil PRT. PRPPSYN; PRPP synthetase GLNPRT; glutamine phosphoribosyl amido transferase.

Figure 4.2 (cont. 2 of 2)

```
                                  110        120        130        140
APRT                              |          |          |          |
Cricetulus longicaudatus      GKAELEIQKDALEPGQKVVVVDDLLATGGTMCAACELLG
Homo sapiens                  GKAELEIQKDALEPGQRVVVVDDLLATGGTMNAACELLG
Escherichia coli              GTDQLEIHVDAIKPGDKVLVVDDLLATGGTIEATVKLIR
Streptomyces coelicor         SSAEIEVHAEDLTAGDRVLVVDDVLATGGTAKASLELIR
Saccaromyces cerevisiae       HEEIFEMQVEAIPFDSNVVVVDDVLATGGTAYAAGDLIR
Leishmania donovani           APEVMTIRYGSIGKGSRVVLIDDVLATGGTALSGLQLVE
Drosophila pseudoobscura      GSDTFELQRTAIQPGQKVVIVDDLLATGGSLLAASELVR
Drosophila melanogaster       GSDTFELQKSAIKPGQKVVVVDDLLATGGSLVAATELIR
Arabadopsis thaliana          GTDTIEMHVGAVEPGERAIIIDDLIATGGTLAAAIRLLE
HPRT
Cricetulus longicaudatus      DIKVIGGDDLSTLTGKNVLIVEDIIDTGKTMQTLLSLVK
Homo sapiens                  DIKVIGGDDLSTLTGKNVLIVEDIIDTGKTMQTLLSLVR
Vibrio harveyia               RDVRILKDLDDDIKGKDVLIVEDIIDTGNTLNKIREILS
Trypanosoma brucei            GQVRMLLDTRHSIEGHHVLIVEDIV-TALTLNYLYHMYF
Tritrichomonas foetus         NLTI-SKDLKTNIEGRHVLVVEDIIDTGLTMYQLLNNLQ
Toxoplasma gondii I           QLTVLS-DDLSIFRDKHVLIVEDIVDTGFTLTEFGERLK
Schistosoma mansoni           EPILTGLGDPSEYKDKNVLVVEDIIDTGKTITKLISHLD
Plasmodium falciparum         TLEIVS-EDLSCLKGKHVLIVEDIIDTGKTLVKFCEYLK
XGPRT
Escherichia coli              NQRELKVLKRAEGDGEGFIVIDDLVDTGGTAVAIREMYP
OPRT
Trichoderma reesei            DHGEGGNIVGASLKGKRVLIVDDVI-TAGTAKRDAIEKI
Rhizobium trifolii            GHGRNAQIEGNMPEGSRVLVIEDLT-TAGGSMFKFIDAV
Coxiella burnetii             THGKQNQIEGRIRKGQRALIVEDLI-STGKSALAAGLAL
Bacillus subtilus             AHGKGNQIEGAVQEGQKTVVIEDLI-STGGSVLEACAAL
Escherichia coli              DHGEGGNLVGSALQG-RVMLVDDVI-TAGTA-RESMEII
Dictyostelium discoides       AYGTKQLIEGRFKEGDNVLVVEDLV-TSGASDLETVRDL
Podospora anserina            DHGEGGNIVGAPLKGKRVLIVDDVI-SRCTAKREAIAKI
Saccharomyces cervisiae 5     DHGEGGIIVGSALENKRILIIDDVM-TAGTAINEAFEII
Sordaia macrospora            DHGEGGNIVGAPLKGKRILIVDDVI-TAGTAKREAIAKI
UPRT
Haemophilus influenzae        ------QKLASDLEERLSIVVDPMLATGGSMIATLDLLG
PRPPSYN
Rattus norvegicus             EVDRM--VLVGDVKDRVAILVDDMADTCGTICHAADNLL
Homo sapiens                  EVDRM--VLVGDVKDRVAILVDDMADTCGTICHAADKLL
Caenorhabditis elegans        EVEKM--TLVGSVEGKVAILVDDMADTCGTICMAADKLV
Saccharomyces cerevisiae I    EVSRM--VLVGDVTDKICIIVDDMADTCGTLAKAAEILL
Saccharomyces cerevisiae II   -----LITLVGNVRGRSAIILDDMIDRPGSFISAAEHLV
Rattus norvegicus             ------ITVVGDVGGRIAIIVDDIIDDVESFVAAAETLK
Escherichia coli              ANVSQVMHIIGDVAGRDCVLVDDMIDTGGTLCKAAEALK
Salomonella typhimurium       ANVSQVMHIIGDVAGRDCVLVDDMIDTGGTLCKAAEALK
Bacillus subtilis             PNVAEVMNIVGNIEGKTAILIDDIIDTAGTITLAANALV
Synechocystis sp.             HNVAEVLNVIGDVQGKTAVLVDDMIDTAGTICEGARLLR
Leishmania donovani           AAGKVDTMQVGEVAGFTCIIVDDMIDTGGTLVKACELLK
GLNPART
Escherichia coli              SVRRKLNANRAEFRDKNVLLVDDSIVRGTTSEQIIEMAR
Bacillus subtilus             GVRMKLSAVRGVVEGKRVVMVDDSIVRGTTSRRIVTMLR
Gallus gallus                 GVAKKFGVLSDNFKGKRVVIIDDSIVRGNTISPIIKLLR
```

Figure 4.3 : Alignment of "A" sequences from phosphoribosyl transferases and more distantly related nucleotide binding proteins. The highly conserved "GKS" of the so-called phosphate-binding loop ("SRG" in Chinese hamster APRT) (Saraste *et al.*, 1990) is bolded. Numbering is for CHO APRT.

Organism	Protein	Substrate	Sequence (60 | 65)	Reference
Chinese hamster	APRT	AMP	IDYIAGL--**DSRGF**	this work
Yeast	APRT	AMP	IDFIAGI--**EARGL**	this work
Human	HGPRT	GMP/IMP	HHIVALC--**VLKGG**	this work
Toxoplasma	HPRT	IMP	LHIICIL-**KGSRGF**	this work
Salmonella	ATPRT	ATP	RLRIAIQ-**KSGRLQ**	1
Porcine	AK1	AMP/ATP	IIFVVGGPGS**GKGT**	2
Bovine mito.outer	AK1	AMP/ATP	RAVLLGPPGA**GKGT**	2
E.coli	F_1ATPase B	ATP	KIGLFGGAGV**GKTV**	2
Human	CFTR(N-term)	ATP	RVGLLGRTGS**GKST**	3
E.coli	OPPF	ATP	TLGVVGESGC**GKST**	3
Human	MDR(N-term)	ATP	TVALVGNSGC**GKST**	3
Neurospora	H^+-ATPase	ATP	TVAIPGAFGC**GKTV**	4
Carrot	H^+-ATPase	ATP	TCAIPGAFGC**GKTV**	4
Human	PKC-B	ATP	TDFNFLM-VL**GKGS**	5
Human	h-AKT	ATP	NEFEYLK-LL**GKGT**	5
E.coli	TK	TMP	LYFYYSAMNA**GKST**	6
Human	TK	TMP	IQVILGPMFS**GKST**	6
Vaccinia virus	TK	TMP	IQLIIGPMFS**GKST**	6
Herpes simplex	dTK	dTMP	RVYIDGPHGM**GKTT**	7
Epstein-Barr	dTK	dTMP	SLFLEGAPGV**GKTT**	7
Salmonella	MUTS	ATP	---ITGPNMG**GKST**	8
Human	h-RAS	GTP	--VVVGAGGV**GKSA**	9
Horse	PGK	ATP	-PFLAILGG-**AKVA**	9
Drosophila	EF-11	ATP	-IVVIGHVDS**GKST**	9
Arabadopsis	EF-1A	GTP	-IVVIGHVDS**GKST**	9
Human	Myosin	ATP	--LITGESGA**GKTV**	9
E.coli	FGART	ATP	ELVSDGG-RG**GKFE**	10
E.coli	EF-Tu	GTP	NVGTIGHVDH**GKTT**	11
Yeast	CDC-42	GTP	KCVVVGDGAV**GKTC**	11
Yeast	RNAhelicase	ATP	DLMACAQTGS**GKTA**	12

Reference Key:

1 Argos *et al.*, 1983
2 Schulz, 1987
3 Hyde *et al.*, 1990
4 Gogarten *et al.*, 1989
5 Bellacosa *et al.*, 1991
6 Bockamp *et al.*, 1990
7 Balasubramaniam *et al.*, 1990
8 Haber and Walker, 1991
9 Saraste *et al.*, 1990
10 Sampei and Mizobuchi, 1989
11 Bourne *et al.*, 1991
12 Jamieson *et al.*, 1991

Figure 4.4 (1 of 2) : Alignment of sequences in the "B" region of phosphoribosyl transferases and other nucleotide binding proteins from various sources. Numbering follows the pattern for Chinese hamster APRT.

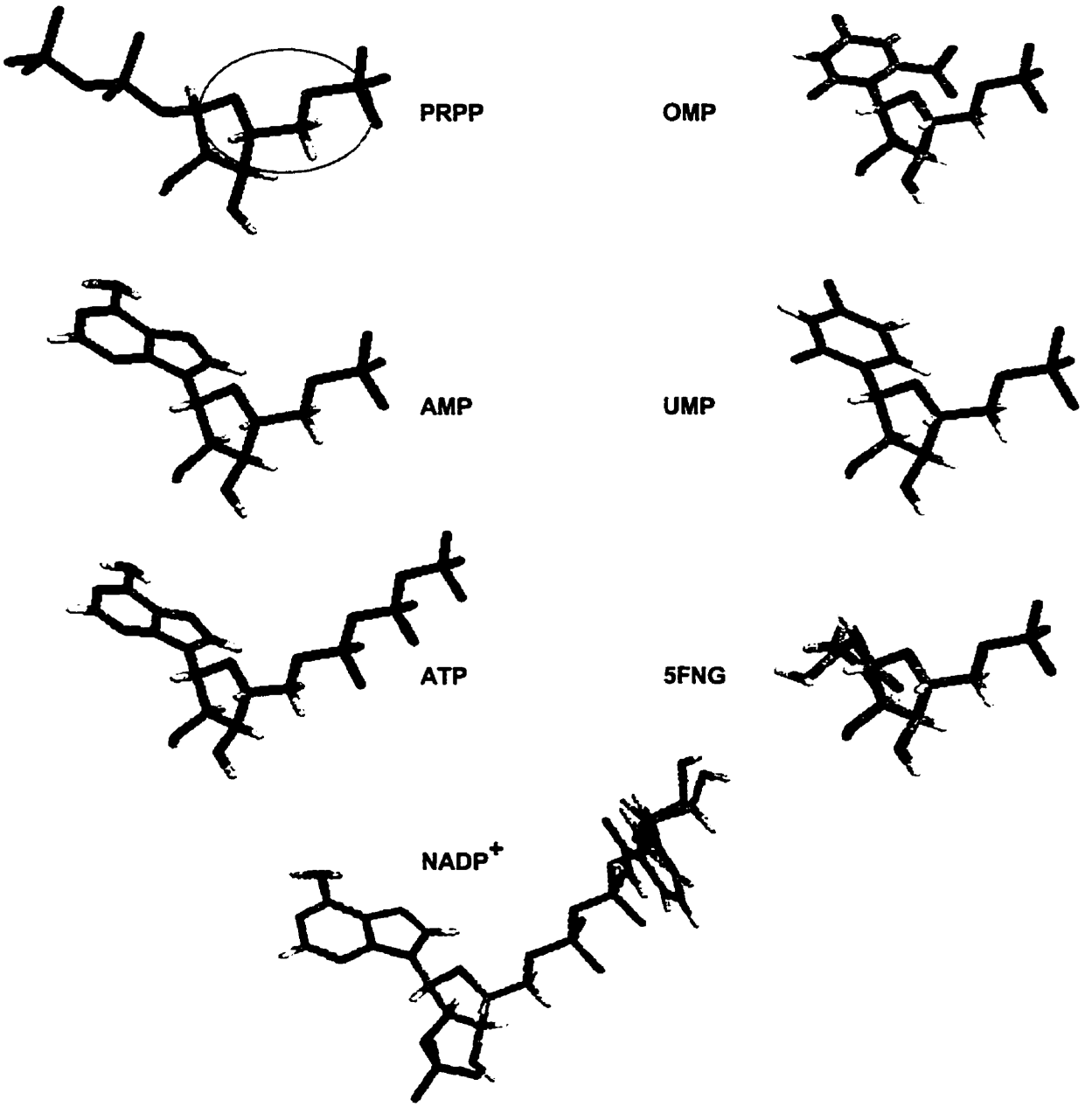
Figure 4.4 (cont) :

Species	Protein	120 130 140
Cricetulus longicaudatus	**APRT**	--PGQ-KVVVVDDLLATGGTMCAACEL
Mus pahari		--PGQ-RVVIVDDLLATGGTMFAACDL
Rattus norvegicus		--PGQ-KVVIVDDLLATGGTMCAACEL
Homo sapiens		--PGQ-RVVVVDDLLATGGTMNAACEL
Drosophila pseudoobscura		--PGQ-KVVIVDDLLATGGSLLAASEL
Saccharomyces cerevisiae I		--FDS-NVVVVDDVLATGGTAYAAGDL
Triticum aestivum		--PND-RVLIVDDLIATGGTLCAAAKL
Synechocystis sp.		--PHH-RVLIVDDLIATGGTAKATAEL
Escherichia coli		--PGD-KVLVVDDLLATGGTIEATVKL
Pseudomonas stuzeri		--EGD-SVLLFDDLIATGGTLLAAAQL
Haemophilus influenzae		--EGD-NVLIIDDLLATGGTVEATVKL
Streptomyces coelicor		LTAGD-RVLVVDDVLATGGTAKASLEL
Mycobacterium tuberculosis		-VAGR-RVVIIDDVLATGGTIGATRRL
Caenorhabditis elegans		--NGD-IVFLIDDLLATGGTLRAATDL
Leishmania donovani		IGKGS-RVVLIDDVLATGGTALSGLQL
Mycoplasma genitalium		QANNAKRCVIVDDVLATAGTVAAIDQL
Methanococcus jannaschii		--KGD-KVVIIDDVISTGGTMIAIIDA
Mastom. unguiculatus	**HPRT**	ALT-GKNVLIVEDIIDTGKTMQTLLSL
Homo sapiens		TLT-GKNVLIVEDIIDTGKTMQTLLSL
Cricet. longicaudus		TLT-GKNVLIVEDIIDTGKTMQTLLSL
Plasmodium falciparum		CLK-GKHVLIVEDIIDTGKTLVKFCEY
Trypanosoma brucei		DIR-GKHVLVLEDILDTALTLREVVDS
Tritrichomonas foetus		NIE-GRHVLVVEDIIDTGLTMYQLLNN
Leishmania donovani II		SVE-NRHILIVEDIVDSAITLQYLMRF
Toxoplasma gondi		IFR-DKHVLIVEDIVDTGFTLTEFGER
Escherichia coli		DIR-GKDVLIVEDIIDSGNTLSKVREI
Rhodobacter capsulatus		IIG-GRDVLVVEDIIDTGHTISKVMEM
Homo sapiens	**OPRT**	TINPGETCLIIEDVVTSG-SSVLETVE
Trichoderma reesei		ASLKGKRVLIVDDVITAG-TAKRDAIE
Dictyostelium discoides		RFKEGDNVLVVEDLVTSGASDLETVRD
Sacc. cerevisiae [5]		ASLEDKRVLIIDDVMTAG-TRINEAFE
Podospora anserina		SALENKRILIIDDVMTAG-TAINEAFE
Rhizobium trifolii		NMPEGSRVLVIEDLTTAGGSMFKFIDA
Escherichia coli		SALQGR-VMLVDDVITAG-TARESMEI
Bos taurus	**UPRT**	AVNPGDTCLIIEDVVSSG-SSVWETAE
Drosoph. melanogaster		IFNAGDTCLIVEDVVTSG-SSILDTVR
Azo. brasilense	**ANTHPRT**	KAIPVEAPDGTIDTCGTGGDGSGTYNI
Halo. volcanii		LIHRARPAARSSDTAGTGGDDYNTINV
Chicken	**GLNPRT**	GVVEGKRVVMVDDSIVRGTTSRRIVTM
Homo sapiens		DNFKGKRIVLVDDSIVRGNTISPIIKL
S. typhimurium	**QNPRT**	GVFCGKRWV-GDDVRLTWHVDDGDAIH
OTHER		
H. sapiens	H-ras	SEDSYRKQVVIDGETCLLDILDTAGQE
Saccharomyces	H-ras	-EDSYRKQVVIDDKVSILDILDTAGQE
H. sapiens	thymidine kinase	ANEGKTVIVAALDGTFQRKAFGSILNL
Saccharomyces	gal. kinase	KQNLMR-ITVVADIMLVLTMAVWIRLP
S.fradiae	*tlrC* kinase	TLASQPELLLLDEPTNDLDDRAVHWLE
M. tuberculosum	UvrA	KRSTGRTVYILDEPTTGLHFDDIRKLL
H. sapiens	CFTR	SVLSKAKILLLDEPSAHLDPVTYQIIR
Saccharomyces	EeF-3	GTWQRPHLIVLDEPTNYLDRDSLGALS
H. sapiens	MDR-1	ALVRNPKILLLDEATSALDTESEAVVQ
Legionella	REC A	VRSAAVDVVIIDSVAALTPKAEIEGEM
Saccharomyces	RNA helicase	ISLANVKYLVLDEADRMLDMGFEPQIR
Xenopus	RNA helicase	IGLDFCKYLVLDEADRMLDMGFEPQIR
Bos	UDP glucoronosyl transferase	KLQRGKFEVLLSDPVFPCGDIVALKLG

nucleoside, based on commonalties of substrate structures. But all of these proteins exploit the high-energy bonds of the 5'-phosphates of the bound nucleotide. In the case of phosphoribosyl transferases, the energy of the reaction comes from the 1' pyrophosphate group, and the 5'-phosphate remains unreacted. The phosphoribosyl transferase reaction is at the opposite end of the bound substrate, and is chemically quite different from the reactions of the other nucleotide binders.

In Chapter 2 it was hypothesized that the B sequences are responsible for interactions with substrate, specifically with PRPP, yet this sequence shares strong similarity with many nucleotide binding proteins which do not bind PRPP. Previous reports have suggested, based on inferred structural models, that this region interacts with the 1'-pyrophosphate of the PRPP (Argos *et al.*, 1982; Aghajari *et al.*, 1994). However, as can be seen in Figure 4.5, the structure of the pyrophosphate moiety of PRPP is rather different from that found in nucleotides such as ATP. It seems unlikely that this highly homologous primary structural element represents a single unique solution to binding pyrophosphate, since that would require quite different tasks in different NTBC proteins. In addition, phosphoribosyl transferase sequences corresponding to the A tract, thought to be involved in the reaction mechanism, are not as clearly defined as in kinases, polymerases, and dehydrogenases, whereas the B tracts, due to the homologous function of binding the common structural elements of nucleotides, are highly conserved among nucleotide binding proteins where it can be identified. Since phosphoribosyl transferases bind nucleotides either as substrates or products of the reaction, it is possible that the difficulty of identifying the Walker A sequence in phosphoribosyl transferases may be due

Figure 4.5: Comparison of the structures of nucleotides which bind to nucleotide-bnding molecules, including phosphoribosyl transferases at or near the B sequences. 5FNG; 5-phosphoribosyl-N-formylglycinamidine. The encircled atoms of the PRPP molecule are invariant in nucleotides.



to a variant of the so-called nucleotide-binding cassette, which might still be identifiable by comparisons to consensus sequences or to individual members of the nucleotide-binding cassette family.

4.1.1 A nucleotide binding cassette in phosphoribosyl transferases

Busetta (1985) and others (Argos *et al.*, 1993; de Boer and Glickman, 1991) have attempted by sequence analysis and structural prediction to identify specific residues of the phosphoribosyl transferases which might serve in binding the respective substrates. Figure 4.6 is a rendition of de Boer and Glickman's proposed model of the APRT nucleotide-binding fold built upon structural modeling and comparison to other nucleotide-binding proteins.

Figure 4.7 is a representation of the same region from the crystal structures of OPRT (Salmonella typhimurium, Scapin *et al.*, 1995) and HGPRT (human, Eads *et al.*, 1995), obtained by digitization in Rasmol 2.6 (Sayle, 1995) of the Protein Databank entries for accession numbers 1STO and 1HMP respectively. The regions corresponding to Busetta's predictions are indicated in highlights. It is notable that in both of the above examples, showing bound product nucleotide, that the most highly conserved aspartate residues of the B sequence, have no close interaction with the product. Recently however, a crystal structure for Salmonella OPRT (Accession code 1OPR) has become available, which contains the unreacted PRPP and orotate molecules (Scapin *et al.*, 1995). In this molecule, the conserved aspartates are clearly in close proximity to the ribose hydroxyls, near the region of the 5'-phosphate group. This observation is at variance with both the Busetta and De Boer models, and with previously published phosphoribosyl

Figure 4.6 : de Boer and Glickman's proposed model of CHO APRT, used with permission. This model shows that the conserved aspartates, D127 and D128, interact with the pyrophosphate moiety of the PRPP. Busetta's model is essentially similar. Note that G133, also highly conserved, is found in a loop rather distant from the reactive center, although this may be modified by a folding of the loop and helix over the adjacent sheet. Numbers indicate residue positions, dots along the chain indicate some of the mutations recovered from *in vivo* mutagenesis experiments.

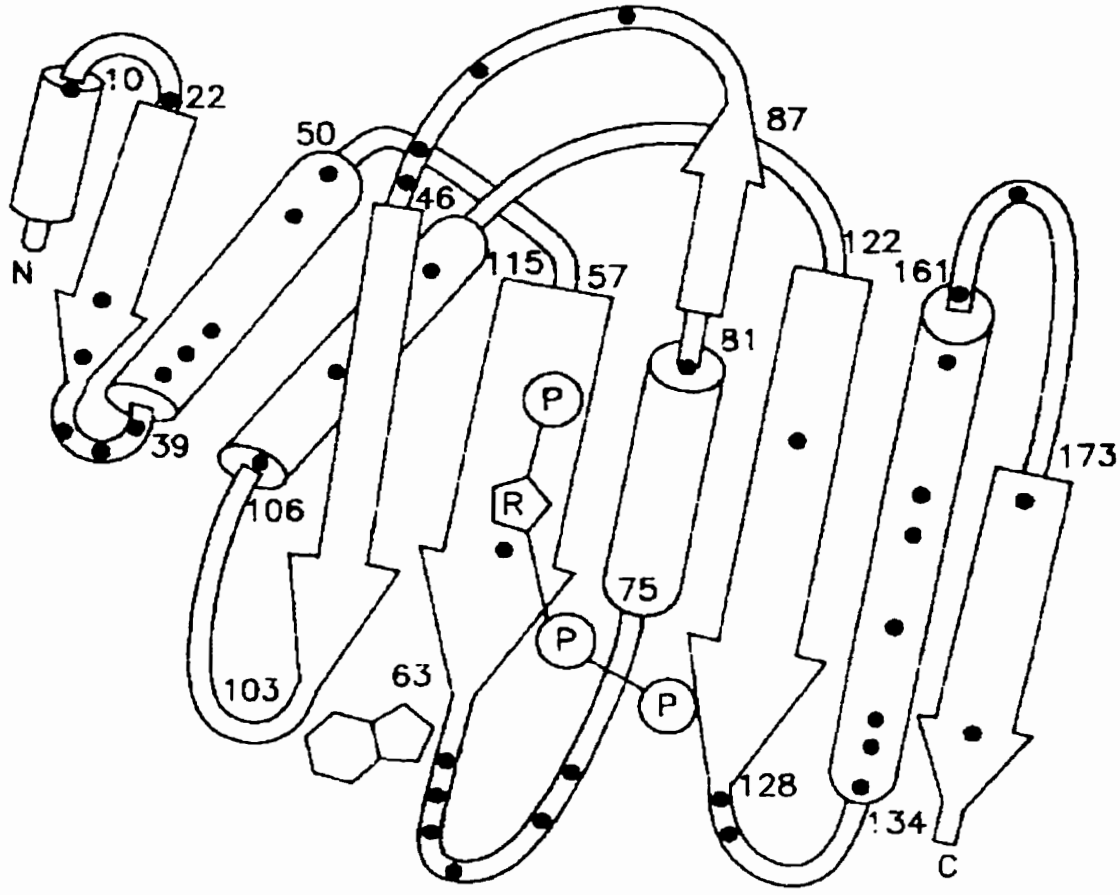
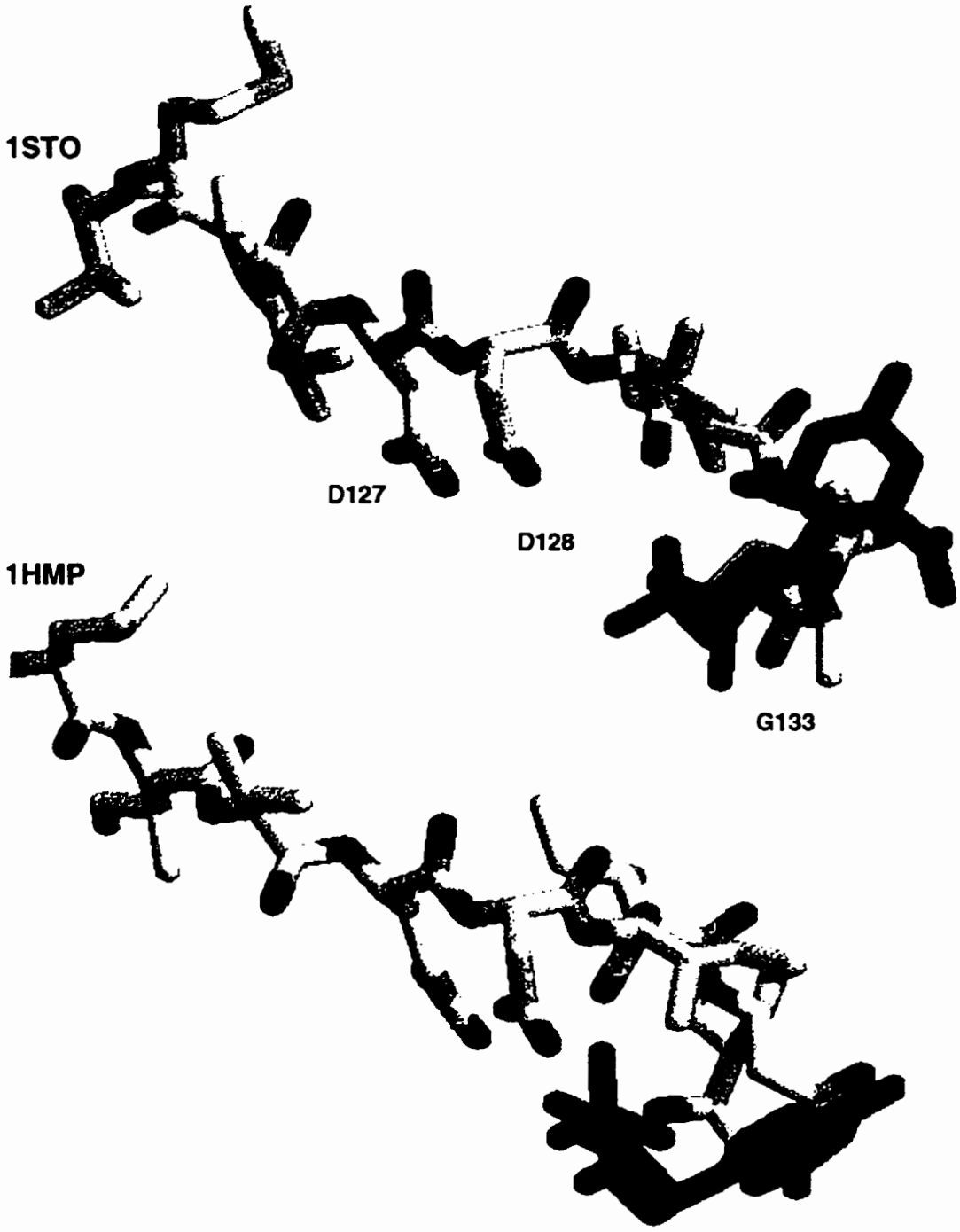


Figure 4.7 : Comparison of the B sequence structure around the bound products in OPRT (1STO) and HPRT (1HMP). The highly conserved region corresponding to Asp127 to G133 of CHO APRT is numbered in the 1STO structure.



transferase structures. In Figure 4.8 the structures of 1STO and 1OPR in the same region are rendered with Rasmol 2.6. Note that the orotate base is in essentially the same position in both frames, and the aspartate D127 (CHO numbering) is also relatively fixed. The remainder of the domain however, is substantially altered. The hydrophobic residues 5' to the conserved aspartates are seen to interact with parts of the ribose ring of the substrate, which is not apparent from the earlier bound product crystal structures. Clearly then, the catalytic site of OPRT, and likely that of the related phosphoribosyl transferases, undergoes a substantial change in shape to accommodate the different components of the catalytic event. This is consistent with Spector's "surface walk" hypothesis (Figure 4.9) , which, in attempting to explain the observed inversion of stereochemistry about C-1, also predicts a substantial translocation of either active site groups or substrate atoms.

4.1.2 Predictions based on sequence comparisons of the putative substrate binding regions of APRT.

Alignments of functionally and evolutionarily related sequences may reveal regions of strong sequence conservation, or conversely, regions in which sequences seem to be variant over evolutionary time. These sequences may represent elements of the catalytic site, critical structural elements, or essential residues involved in substrate binding which may or may not be catalytically or structurally relevant. It may also be possible that some residues within the coding sequence are resistant to change over long spans of time, due to some intrinsic feature of the DNA sequence, such as being contained within nucleosomal or secondary structures (Schweiger *et al.*, 1995), control elements such as enhancers or methylation signals (Park and Taylor, 1988; Magin *et al.*, 1992, Macleod *et al.*, 1994; She

Figure 4.8 : Comparison of the B sequence structure around the bound substrates before (1OPR) and after (1STO) the OPRTase converison of orotate to OMP. The substrate is in blue. Residues are numbered according to the CHO APRT sequence. The local alignment is :

```
1sto       G R V M L V D D V I T A G T A I R E
CHOaprt    Q K V V V V D D L L A T G G T M C A
                      127         133
```

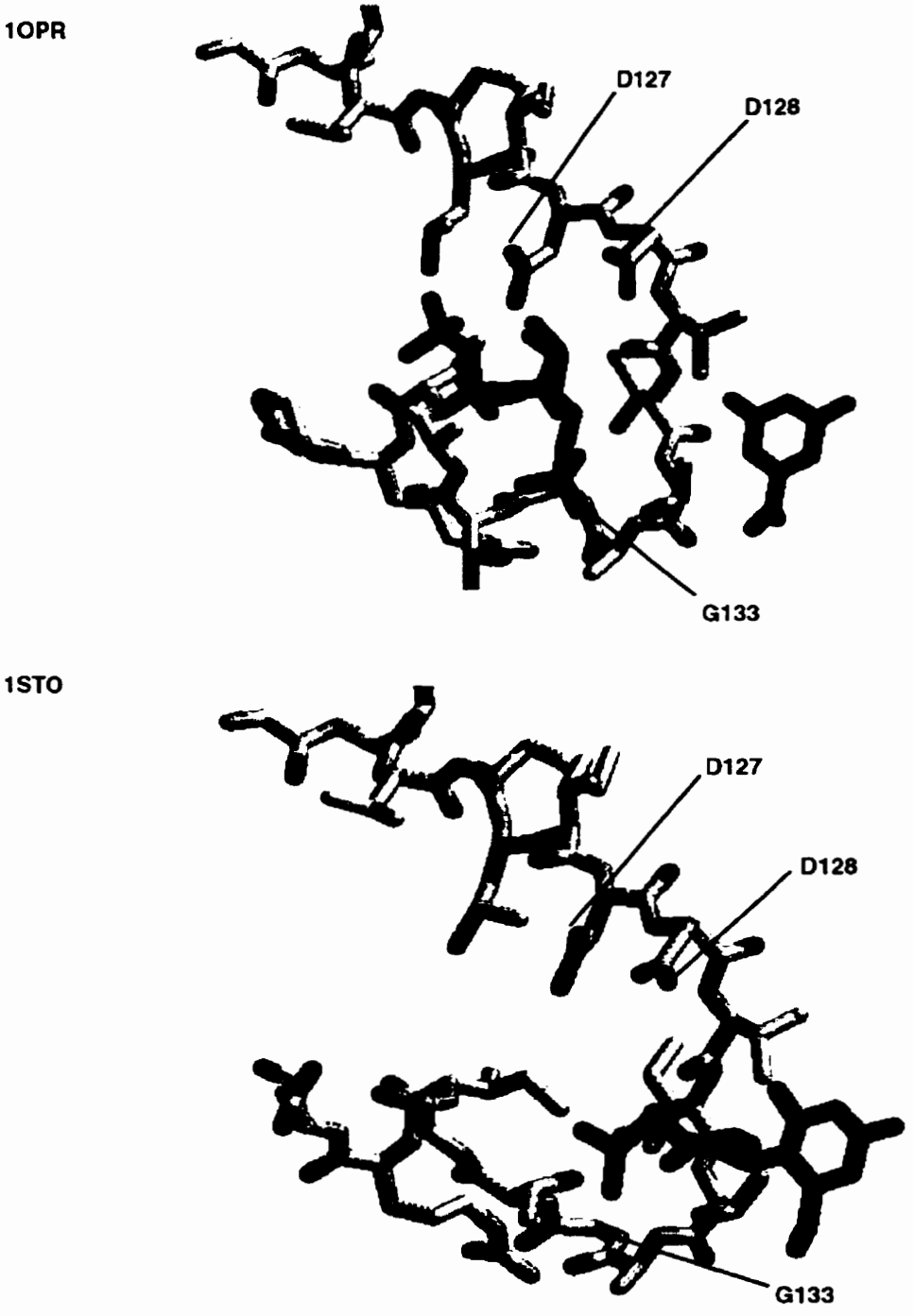
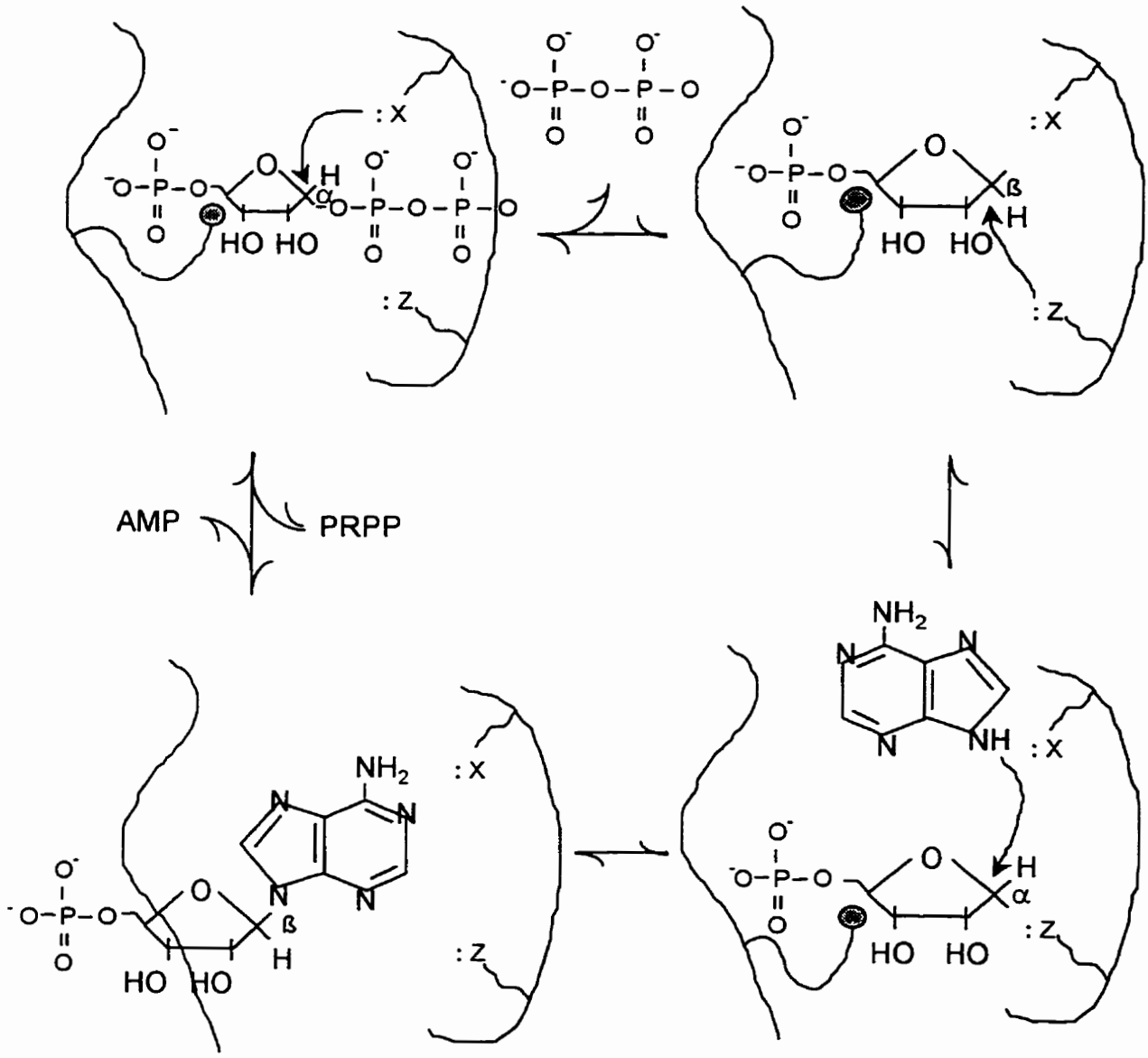


Figure 4.9: Adaptation of Spector's (1988) model of the catalytic events leading to inversion of stereochemistry around C-1 of the ribose in APRT. Note that the C-5 ribose region (invariant in nucleotides) is likely to be involved in substrate binding but not in the catalytic event. This model of molecular translocation during catalysis is consistent with the observed differences in binding of precursor (PRPP) and product in recent OPRT structures (see text).



et al., 1995), or tracts of bases required for correct splicing of introns from messenger RNA (Chen *et al.*, 1993; Kessler *et al.*, 1993). Intron splice sequences especially may be conserved, since incorrectly spliced mRNA molecules have a low likelihood of yielding functional protein products.

It is interesting to speculate on whether sequence tracts under selective pressure beyond the level of protein function (*e.g.* gene regulation, methylation, introns), form sequence cores which may be propagated within the genome by virtue of their effect on gene expression, rather than by any direct effect on protein function, even though the selection pressure is upon the protein. Thus regions of similarity or identity within the coding region are useful but limited indicators of functional importance. Sequence conservation is however a starting point for further analysis, such as kinetic or stability variations of mutants. From alignments of APRTs and related proteins (HPRT, XPRT, OPRT), as seen in Chapter 2, we can identify certain regions of interest. Since it is impractical to construct and characterize all possible mutants at all positions in the coding sequence, a subset of sites, among them the most locally conserved, were selected.

4.2 Materials and Methods

4.2.1 *In vitro* mutagenesis of the *aprt* NTBC

All restriction and modification enzymes, including buffers and cofactors, except as noted, were from New England Biolabs. From the vector construct pKFA26, containing the *aprt* coding sequence from CHO, a BamHI-HindIII fragment containing the expression cassette (operator, promoter, coding sequence) was directionally subcloned into M13mp19 using standard methods (see Chapter 3). The clone mp19A26 was

characterized with respect to the locations of restriction sites for BamHI and HindIII. The mp19A26 construct was further characterized by dideoxy DNA sequencing (Promega, fMol kit) to confirm the orientation of the inserted sequence.

Using mp19A26 as a substrate, the T7-GEN *in vitro* mutagenesis kit (USB, Batt *et al.*, 1990) was used to introduce site specific changes into the *aprt* sequence. Figure 4.10 is a flow diagram of the typical experimental design. In brief, the mutagenic oligonucleotide is annealed to the single-stranded substrate in the provided buffer, by briefly heating to 70 °C, and cooling at ambient temperature (20 °C). The oligonucleotides used and the corresponding mutations they cause are shown in Table 4.1. T7 DNA polymerase, 2 U, are next added, along with deoxynucleotides, including 5-methylcytidine triphosphate. After 30 minutes, T4 DNA ligase (2 Units) and ATP to 1 mM are added. After a further 30 minutes, HhaI and MspI (1 U each) are added. After 10 minutes, 4 U of Exonuclease III is added. After a further 10 minutes, the entire reaction if heated to 65 °C for 10 minutes to stop enzymatic reactions. The HhaI and MspI nick, rather than cut, DNA at their respective cut sites where 5-methyl-cytosine is incorporated. Thus only the non-mutagenized parental strand is nicked, and becomes a substrate for Exonuclease III digestion. The effect of the protocol is to reduce the frequency of non-mutagenized plasmids which may form plaques at the transformation step.

The mutagenesis reaction products were transformed into *Escherichia coli* strain SDM-2 (USB) following the method of Hanahan (1983), then plated in several dilutions onto log-phase lawn cells of the same strain (OD_{600} approximately 0.6, in Luria Broth),

Figure 4.10: T7-GEN *in vitro* mutagenesis flow diagram. Construction of mp19A26, and the sequences of primers used, are described in the text. The *aprt* notation is for the entire expression cassette from pKFA26, including the Ptac promoter, which is excised as a single 1 kb fragment in a BamHI-HinDIII double digest. After Exo III digestion, the newly synthesized mutagenized strand is transformed into *E.coli* SDM-1.

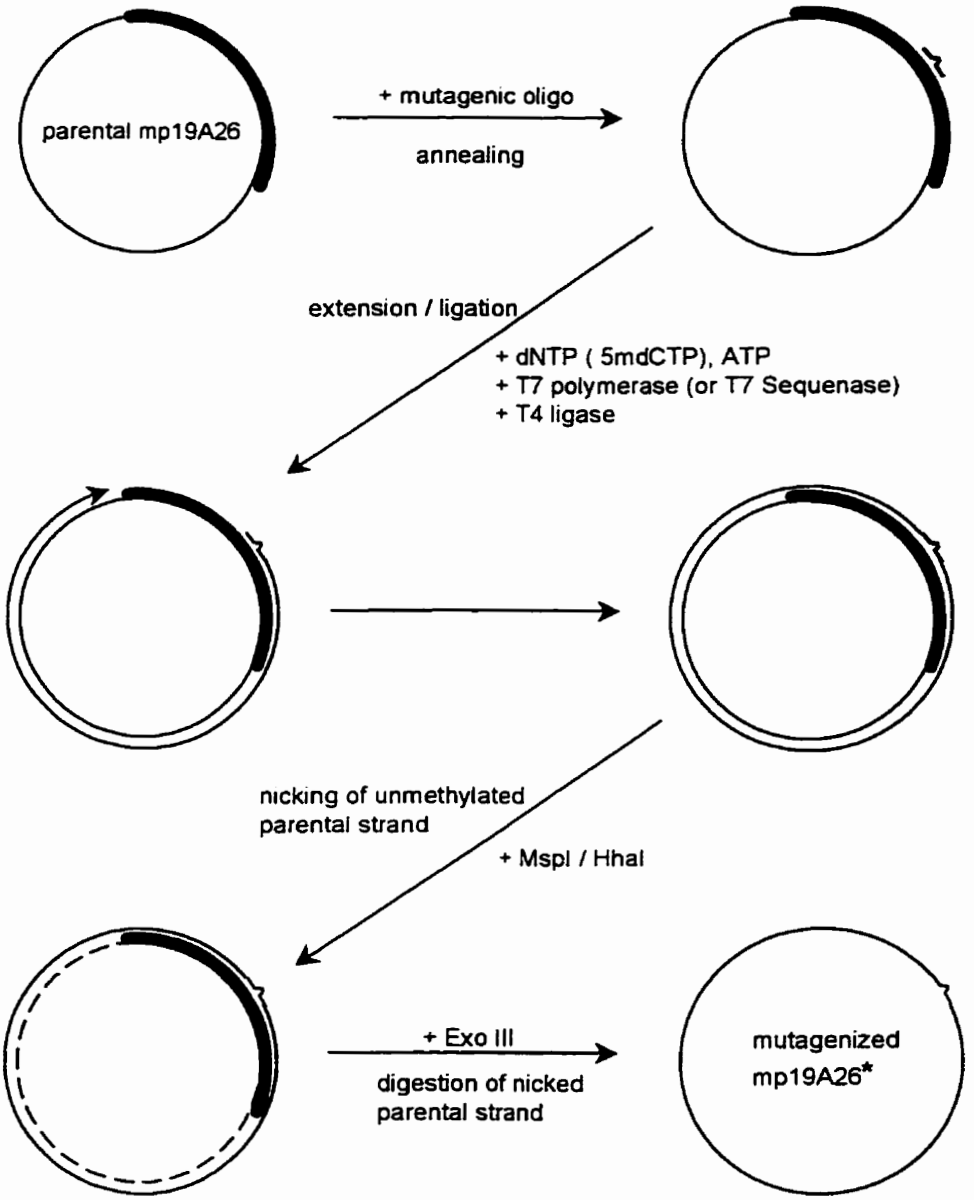


Table 4.1 : Oligonucleotides for site-specific mutagenesis of N-terminus (bolded capital letters denote positions of base-pair mismatch).

Amino Acid	Oligo Sequence	Change	Mutant Name
26	ggcgtgctg**XX**tagggatatc	Phe > random	F26ran
27	cgtgctgttt**GC**aggatatctcgcc	Arg > Ala	R27A
67	cctagactcc**GC**gggattcttg	Arg > Ala	R67A
68	gactccag**C**gattattgttt	Gly > Ala	G68A
93	gggaagctg**XXX**ggcccc	Phe > random	P93ran
101	gcctcc**GG**tgctctcgag	Tyr > Gly	Y101G
103	tcctatgct**AA**cga**CC**a**G**ggcaaggct	L>N,E>D,Y>Q	LEY-NDQ
111	gctgaacta**A**aaatccagaaa	Glu > Lys	E111K
125	gtggtt**A**ttgtagaggatctcc	Val > Ile	V125I
125	gtggtt**A**t**C**gtaga**A**gatctcctggcc	V>I, D>E	
125	ggc**A**a**A**aa**C**gtggtt**A**ttgtaga**G**gat**A** tc**A**t**C**g**A**cactggagg	multiple	HPT-APT
127	gttgtaga**G**gatctcctgg	Asp > Glu	D127E
127	gttgta**A**a**C**gatctcctgg	Asp > Asn	D127N
127	gttgta**C**a**G**gatctcctgg	Asp > Gln	D127Q
127	gttgtag**T**tgatctcc	Asp > Val	D127V
128	tgtagatga**G**ctccggcc	Asp > Glu	D128E
128	tgtagat**A**a**C**ctcctggcc	Asp > Asn	D128N
128	tgtagat**C**a**G**ctccttggcc	Asp > Gln	D128Q
128	gttgtagatg**T**tctcctgg	Asp > Val	D128V
127	ggttgttgtaga**G**ga**G**ctcctgg	D>E, D>E	DEDE
131	gatctcctgg**A**cactggagg	Ala > Asp	A131D
133	gccactg**C**aggaaccatg	Gly > Ala	G133A
134	gccactggag**A**caccatgtgc	Gly > Glu	G133E

to enable plaque formation and detection. From plates with 50 to 100 plaques, 6 plaques for each mutant were chosen for DNA sequencing to determine whether they were mutagenized. In the event that none of the six chosen were mutants, an additional 10 were picked and sequenced.

4.2.2 Identification of *aprt* mutants

The USB T7-GEN protocol yields a high frequency of mutant sequences. Identification of mutant sequences from plaques on agar plates was accomplished by coring with a sterile Pasteur pipette, six plaques per plate from plates containing 50 to 100 well-isolated plaques. These plaque cores were stored at 4 °C until used for DNA preparation. For DNA preparations, a log phase culture of SDM-2 was inoculated with an aliquot of the liquid from a cored plaque suspension. After overnight growth at 37 °C, 1.5 ml of culture was transferred to a fresh microcentrifuge tube, and the cells pelleted at 14000 x G (30 seconds) in a microcentrifuge at room temperature. The cell pellet was frozen for later extraction of double-stranded DNA.

The supernatant, containing the phage particles in suspension, was stored on ice until single-strand DNA could be isolated. In brief, from an autoclaved phage precipitation solution (2.5 M NaCl, 30 % polyethylene glycol 8000), 250 µl were added to 1.0 ml of phage suspension. This mixture was left at room temperature for 20 minutes, then centrifuged at 14000 G for 20 minutes. The resulting supernatant was discarded, and the white, flaky, phage pellet resuspended with 200 µL phenol extraction buffer (10 mM TrisHCl, pH 8.0, 1 mM EDTA, 50 mM NaCl) and vigorous pipetting. To this was added 6M sodium perchlorate to a final concentration of 600 mM. To this suspension was then

added 500 µL of a prepared suspension of glassfibre (Schleicher and Schuell Type 106, gift of Dr. C. Birnboim, 0.5 g, powdered, in 50 ml phenol extraction buffer). This suspension was mixed gently for several minutes. The glass fibre suspension was then precipitated by a short centrifugation at 14000 G, after which the supernatant was discarded. The pellet of glass fiber was resuspended with 200 µL sterile Nanopure water, allowed to equilibrate for 10 minutes at room temperature, then precipitated again with another short centrifugation. The supernatant layer was transferred to a fresh microcentrifuge tube, and 5 M NaCl was added to a final concentration of 50 mM. Two volumes of 95% ethanol were added, and the DNA pelleted by a 30 minute centrifugation at 14000 G. After removing the supernatant, the resulting pellet was washed with a further two volumes of 70% ethanol, briefly dried in room air, and resuspended with 50 µL of TE (10 mM TrisHCl, pH 8.0, 1 mM EDTA). DNA prepared in this fashion was stored at 4°C for up to six months.

Single-stranded DNA prepared as noted above was sequenced by the Sanger dideoxy method (Sanger *et al.*, 1977), using either the Promega fMol cycle sequencing kit. Labeling of the DNA was by incorporation of ^{35}S-α-dATP (3000 Ci/mMol, NEN) during the sequencing reaction. The thermal cycler sequencing program used was 1) 94 °C x 60 seconds 2) 94 °C x 30 seconds 3) 37 °C x 15 seconds 4) 72 °C x 60 seconds. Steps 2-4 were repeated for 35 cycles, followed by a 4 °C chill step. Sequenced samples were stored at -20 °C for up to 48 hours before sequencing gels were run. Immediately before loading onto the gel, samples were denatured in the thermal cycler at 94 °C for 2 minutes. Sequencing reactions were electrophoresed in 6% polyacrylamide gels (Dalton Chemical

Laboratories), with 7M urea, at a constant power of 75 watts. After development, gels were dried on a platform drier (Bio-Rad), and transferred to autoradiography cassettes (Fisher Scientific) with Kodak X-AR film. After overnight exposure (or longer if label activity was low), the exposed films were developed using an automated film processor (Kodak). Sequence data was collected and recorded by hand. The entire coding sequence of identified mutants was determined to ensure that no adventitious secondary mutations had been incurred during the procedure.

4.2.3 Subcloning *aprt* mutants for expression

When the mutation in a given single-stranded clone was confirmed by DNA sequencing, it was necessary to transfer the expression cassette containing the promoter and coding sequence into the pKF1.0 expression vector for effective expression and characterization. Double stranded DNA was prepared from the saved cell pellet from the original single stranded DNA preparation. The cell pellet was thawed, and alkaline lysis performed using the Promega Wizard miniprep kit. If the DNA yield was insufficient, a second overnight culture in fresh B2325 cells was prepared and DNA isolated from that. DNA quality and yield was estimated by electrophoresis of 5 µl aliquots in 1.0 % low EEO agarose (Dalton Chemical Laboratories), using ethidium bromide (1 µg/ml) as a DNA visualization stain.

DNA for subcloning, and the pKF1.0 vector DNA, in 20 µl aliquots, were digested with BamHI and HindIII, 5 U each, following the manufacturer's instructions in a total volume of 25 µl. Digestion to completion was verified by agarose gel electrophoresis. After heat treating at 65 °C for 10 minutes to inactivate the restriction enzymes, 15 µl of

digested clone DNA were mixed with 15 µl of digested vector DNA in a fresh microcentrifuge tube. 5 µl of 10X ligase buffer, 5 µl of 10 mM ATP, and 8 µl of sterile Nanopure water were also added. After mixing, 2 µl, approximately 15-20 U, of T4 DNA ligase were added. The ligation reactions were incubated at ambient temperature overnight, up to 18 hours, then used directly in transformations of B2325 (K12 *HfrH lac thi deoD8 purD apt*) competent cells, following the method of Hanahan (1983).

4.2.4 Expression and characterization of ntbc mutants

From frozen stocks of verified subclones containing the mutation of interest, overnight cultures in LB with 100 µg./ml ampicillin were prepared. The next day, large volume (1 litre) cultures of LB, supplemented with $FeCl_3$ and 1X BMG (10X: 20% glucose, 5% yeast nitrogen base without amino acids (Difco), 100 mM glycine) were inoculated with the overnight cultures. Ampicillin (ICN) to 100 µg/ml was also added. These flasks were shaken at 200 rpm overnight at 37°C, up to 16 hours. To the dense culture the next morning, the cultures were induced with IPTG (1 mM), and the incubation with shaking continued for two hours. Cell pellets were collected by centrifugation at 4 °C, in 1 litre Nalge bottles, in a Sorvall RC3 centrifuge at 2000 G. These cell pellets were transferred to 50 ml Corning centrifuge tubes, and weighed. 5 ml of 0.5 X TEGN (10 mM Tris HCl, pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, 0.5 mM EDTA, 5% glycerol) was used to rinse the centrifuge bottle, and this was mixed with the cell pellet, which was subsequently stored at -80 °C until use.

Protein extracts were prepared by thawing a cell pellet on ice for 1 hour, then transferring the pellet to a chilled mortar on ice. A volume of 10X proteinase inhibitor

cocktail (1 mM TPCK, 0.2 mM leupeptin, 1 mM TLCK, 0.5 mg/ml chymotrypsin-trypsin inhibitor (all SIGMA) in sterile Nanopure water) was added to a final concentration of 1X, and β-mercaptoethanol (SIGMA) was added to 0.1 % (*i.e.* 1 µl per 1000 µl extract). A weight (1/10 of mass of cell pellet) of activated alumina powder (SIGMA A-1842, McIlwein, 1948) was added to the mortar, and the pellet was ground to a paste with the alumina powder. After grinding, the paste was transferred to a Corning tube, and centrifuged at 4 °C for 20 minutes to pellet alumina and cell debris. The supernatant was then transferred to clean 35 ml Oakridge tubes, protamine sulfate from a 20 mg/ml stock was added to 2.0 mg/ml, and 100 mM MgPRPP to a final concentration of 0.5 mM, and the solution mixed gently for 20 minutes while on ice. The suspensions were then centrifuged at 20,000 RPM in a Sorvall SS-34 rotor in a Sorvall RC5-C centrifuge (approximately 30,000 x g) for 60 minutes. After centrifugation, the clarified extracts were aliquotted to microcentrifuge tubes. Samples were retained for immediate activity assessment and protein concentration measurements. The aliquots were frozen at -20 °C for up to 2 weeks until used in further characterization experiments. Extracts were stored for longer periods at -80 °C.

For kinetics analysis, extracts with sufficient activity were further purified as follows. The extract was heated for 1 minute to 42 °C, then chilled on ice. After centrifuging at 14,000 x g for 5 minutes at 4 °C, 500 µl aliquots of extract were purified using FPLC (Pharmacia) on a Superose 12 (micro-preparative scale) size-exclusion column with a 30 ml bed volume. The column was eluted with chilled 0.5 X TEGN containing 0.5 mM MgPRPP at 0.5 ml/min. 100 µl eluate fractions corresponding to a

molecular weight of approximately 40,000 Daltons (mass of the APRT dimeric molecule, approximately 19 ml eluted volume) were collected and assessed for APRT activity. Peak activity fractions were pooled and realiquotted as 100 µl samples in fresh tubes for storage at 4 °C. Automated SDS-PAGE (Phast system, Pharmacia, according to the method of Laemmli, 1970), 20 % acrylamide gels, followed by silver staining (Pharmacia Phast-Silver kit), were run on each pool to evaluate relative purification of the samples. Such purified fractions were used within 24 hours for kinetic analysis. Immediately before kinetic assays, the aliquots were desalted on Sephadex G-25 minicolumns (Pharmacia) to remove residual PRPP.

4.2.5 Kinetics of mutant APRT

^{14}C-8-adenine (50 Ci/mMol) was obtained from NEN, and NaPRPP from Sigma. Mg-PRPP was prepared in small batches by adding 2 moles of $MgCl_2$ for each mole of PRPP to a 100 mM solution of NaPRPP, and stored over liquid nitrogen until use. Evaluation of the gross catalytic competence of the mutant enzymes isolated as described above was done using an adaptation of the assay method of Okada *et al.* (1986). In brief, an aliquot of crude or purified extract, less than 1 µg total protein per assay sample, was added to buffer (0.5X TEGN) on ice. In the case of gross activity assays, the buffer also contained 100 µM ^{14}C-8-adenine, and 600 µM MgPRPP. For the normal enzyme these concentrations should be saturating. The reaction was immediately mixed and placed at 37 °C in a temperature controlled sandbath for 20 minutes. For kinetics assays the concentrations of the substrates and length of incubation time was varied as needed. After incubation, 10 µl of 50 mM EDTA pH 8.0 was added with mixing, and the reaction placed

on ice. The total volume of the reaction was transferred by micropipet to DEAE ion exchange glass-fiber filters 2.5 cm in diameter (Whatman DE81). These filters (except total count filters) were washed three times in 10 litres of cold tap water. After the final wash, filters were transferred to dry blotting paper, and dried at 65 °C. Dried filters were transferred to polypropylene scintillation vials, 8 ml of scintillation cocktail (Fisher Scintiverse II) were added, and the sample activities were measured in a Wallac 1410 counter. Counts were recorded on a desktop computer connected to the counter, and transferred as plain text files to Statistica for analysis.

4.3 Results

4.3.1 Mutants of the ntbc of APRT

Putative mutant sequence-containing plaques were sequenced by the dideoxy method (Sanger *et al.*, 1977). A set of representative sequencing gel autoradiographs is shown in Figure 4.11. Each mutant clone was sequenced from 5' of the ATG codon to downstream of the translation stop codon. In each case, no adventitious mutations were found. In the case of residue R67, a single mutant with a deletion of this codon was identified, probably caused by an impurity in the mutagenic oligonucletide. Confirmed mutants were subcloned into pkF1.0. Figure 4.12 is a set of representative agarose gels for screening subclones for inserts of appropriate size, followed by restriction enzyme digestion analysis to verify position and size of the insert, as well as the presence of characteristic restriction sites. Clones with multiple inserts, or apparent alterations in size or restriction pattern were not further characterized, since the insert was likely to be modified from the original sequence.

Figure 4.11: Sequencing gel autoradiographs of some of the mutant clones as described in the text. The final digit of the clone name is for identification purposes only. Arrows indicate site of the changes in the mutant clones. Underlined bases are different from wildtype sequence at that position.

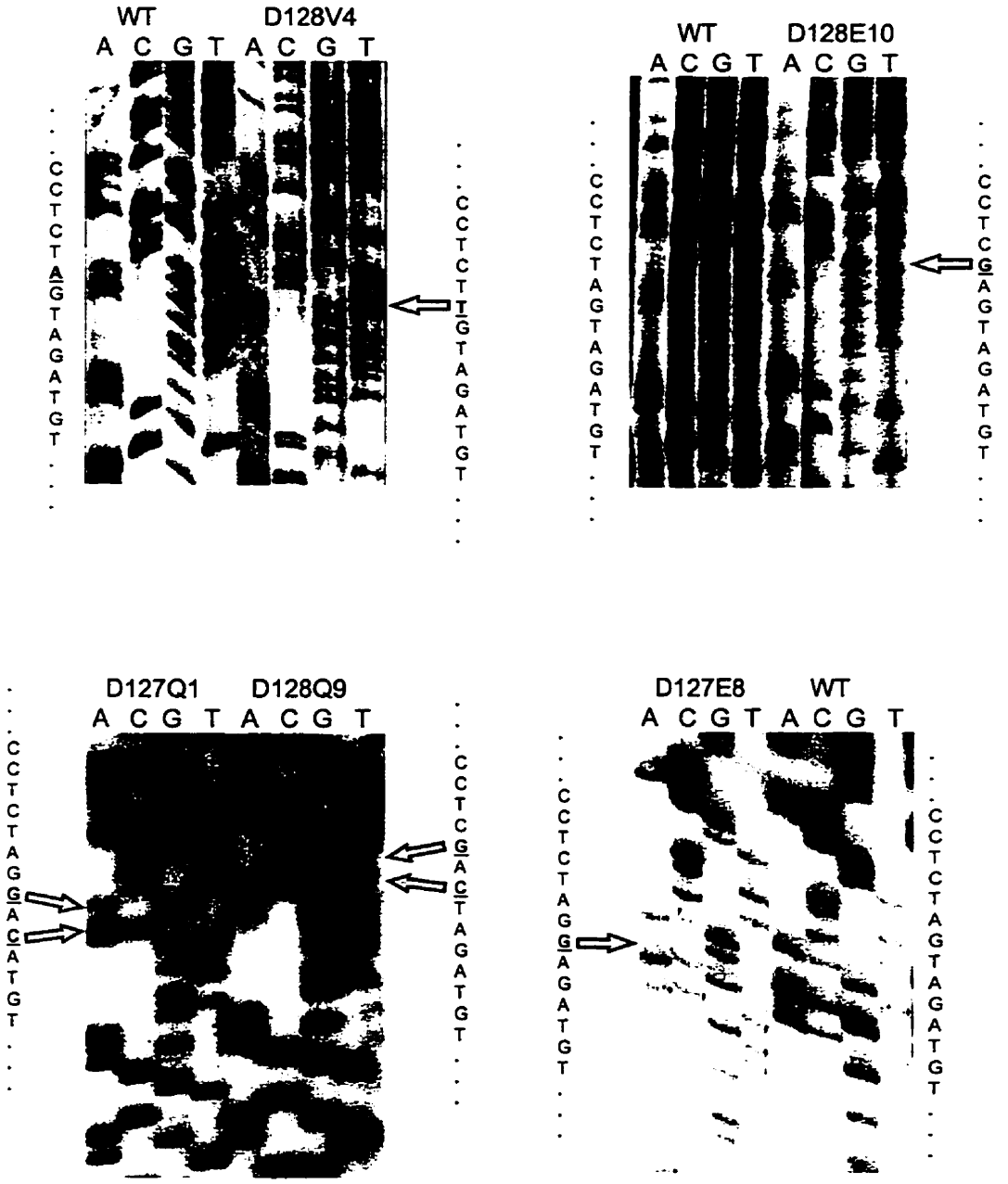
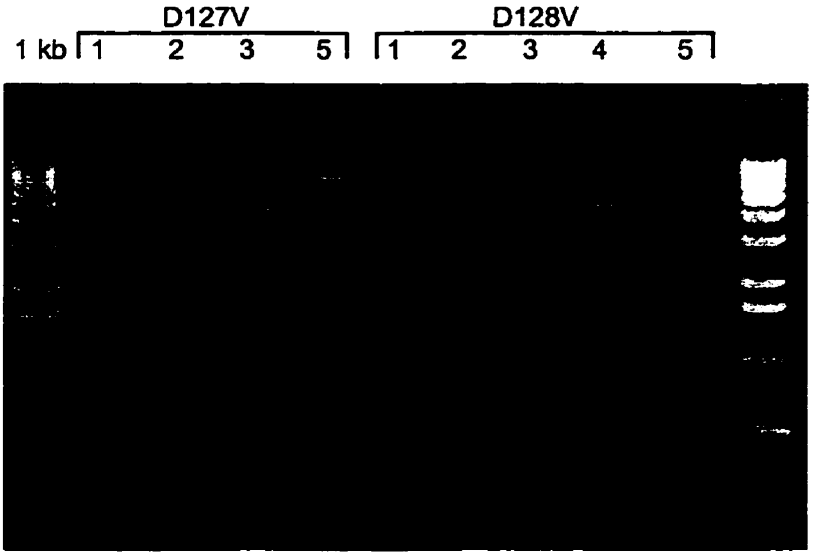
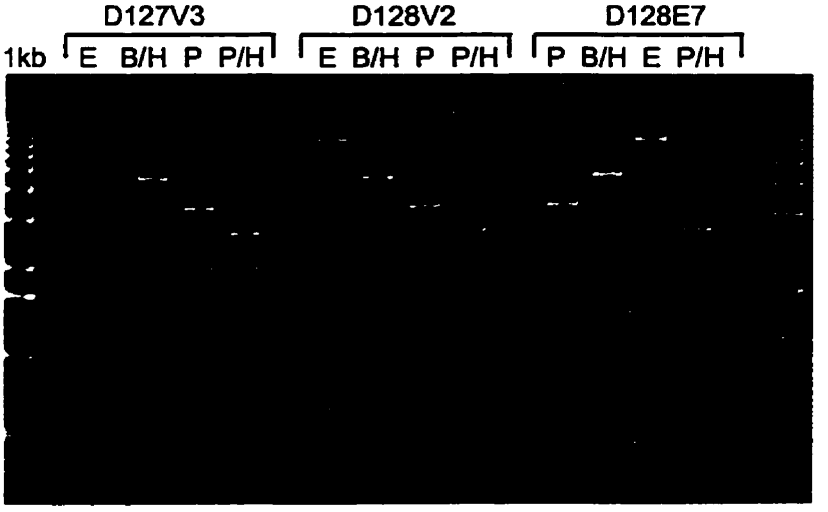


Figure 4.12 : Typical subcloning gels. The isolated plasmid after the subcloning tranformation has been a) digested with BamHI and HindIII, then run to check the size of any insert. As well, in a), some clones have multiple bands of unexpected size, and are not further characterized. Next, as shown in b), the insert-bearing clones are restricted with EcoRI (E), BamHI-HindIII (B/H), PvuII (P), and PvuII-HindIII (P/H). Numbers following clone names are for identification only.

a)



b)



4.3.2 Zero order kinetics of mutant APRT enzymes

Results of saturation kinetics experiments are shown in Table 4.2. Several mutants have activity levels in clarified extracts very similar to that of the wildtype enzyme. Enzymes with little activity in clarified extracts were not further characterized.

4.3.3 Kinetic parameters of ntbc mutants

Results of kinetic assays on mutants with greater than 10% of wildtype activity are summarized in Table 4.3. Note that not all ntbc mutants were tested due to poor stability under storage or assay conditions. This poor stability likely leads to a significant loss of activity during the process of assaying the activity. Thus, in the time it takes to accumulate sufficient product to measure accurately, the enzyme activity has been falling, thus no reliable assessment of kinetic parameters is possible, despite substantial activity by gross activity measurement.

4.4 Discussion

All extracts were prepared and tested in parallel with the wildtype protein. Thus it is reasonable to infer that mutants which have substantially reduced gross activity relative to the normal enzyme extracts have suffered functional or structural change which are deleterious to enzyme activity. Note that not all nucleotide binding cassette mutants were analysed kinetically due to insufficient stability of the extracted mutant protein under storage or assay conditions. This poor stability leads to a significant loss of activity during the process of the assay itself. Thus, in the time it takes to accumulate sufficient product to measure accurately, the enzyme activity has been falling. This causes the reaction rate to be nonlinear, without correspondence to substrate conditions. Thus no reliable

Table 4.2 : Kinetic parameters of wildtype APRT and mutants expressed as native protein from the pKFA26 derivatives. Assay conditions are described in the text. The wildtype APRT activity is stable to freeze-thaw during storage under the described conditions. Kinetics were not done on mutants with less than 10 % of the wildtype activity, or on mutants with very poor stability.

Clone	% Relative Activity	Notes
wildtype	100	
F26V	9	1
R67A	9	1
R67Δ	4	1
G68A	6	1
Y101G	43	1
LEY-NDQ	36	1
E111K	3	1
V125I	0.2	
HPT-APT	3.2	
D127E	20	2
D127Q	6	
D127V	1	
D128E	80	2
D128Q	6	
D128V	19	2
DEDE	3	
A131D	46	2
G133A	76	2
G134E	93	2
L143Δ	100	3
V150Δ	20	3
V157Δ	1	3
S161Δ	80	3
F173Δ	80	3
L176Δ	1	3

Notes:

1 Approximately 90% of activity lost after 1 freeze-thaw cycle

2 Approximately 50% of activity lost after 1 freeze-thaw cycle

3 See Chapter 5

Table 4.3 : Kinetic parameters of wildtype APRT and mutants with greater than 10% of wildtype activity. Assay conditions are described in the text.

Clone	% Relative Activity	K_M adenine	K_M PRPP
wildtype	100	9.6 μM	2.5-3.7 μM
Y101G	43	See Note	>750
D127E	20	9.1	28.6
D128E	80	8.7	228-268
A131D	46	405	28.0
G133A	76	24.9-26.6	34.1
G134E	93	18.7-19.5	9.1

Note: Adenine kinetics could not be done for Y101G, due to solubility limits of PRPP. It is probable that the estimate of K_M for PRPP is low.

assessment of activity nor of kinetic parameters is possible for these mutants. Whether these changes are due to stability or direct loss of catalytic function within a properly folded protein cannot be determined from gross analysis alone. Obviously however, if a given mutant has lost all or most of its activity relative to the wildtype enzyme due to poor stability to freezing or freeze-thaw cycles, this is a strong suggestion that they have poor structural stability. Since most of the residue changes are fairly conservative, the observation of radical loss of stability implies a key role for these residues, or of residues in their near proximity (in the folded structure), in maintaining the folded state.

In general, zero order or saturation kinetics reveal less about the molecular basis of the difference between mutants, but can be indicative of gross functional or stability differences between proteins. In the case of the mutants summarized in Table 4.2, it is notable for example, that individual residue substitutions at the highly conserved "B" motif around D127-D128 are functionally tolerated. Residue D128 is highly conserved in phosphoribosyl transferases, but replacing it with glutamate results in an enzyme with about 80% of the gross activity of the wildtype enzyme, and a substantially increased K_M for PRPP. The K_M for adenine of the D128E mutant is not much different from that of the wildtype enzyme. Conversely, the D127E mutant has a grossly altered K_M for PRPP relative to wildtype, but retains more activity. Other substitutions at these positions exhibit a much greater impact on function. Substitution of either aspartate by glutamine (D127Q, D128Q) causes greater loss of activity, and substitution with valine more still. It is interesting that D127V is virtually completely nonfunctional, even though this site is the less conserved of the two aspartates. These observations suggest that only a carboxyl

group at least one of positions 127 or 128 can satisfy the functional requirement of the enzyme. In the case of D127V, the valine probably causes a local distortion of the chain such that the carboxyl on D128 is no longer available. The likely role of these residues in binding the ribose-phosphate moiety which is invariant in nucleotides is consistent with the observations of bound PRPP in OPRT. It is an interesting evolutionary situation that there appear to be few major variations on the "B" sequence motifs in other nucleotide binding proteins. It is possible that the "B" motif is one of the few solutions to the nucleotide binding problem.

The aspartate core of the highly conserved "B" motif is not the only element of this region with important functions. Subsitutions in APRT of either G133 or G134, also well conserved in phosphoribosyl transferases, but less so in related proteins, cause changes in apparent K_M for both substrates. If these residues were directly involved in the catalytic event one would expect to see a gross loss of catalysis, which is not seen. G133 and G134 probably form part of the catalytic pocket, since glycines are capable of forming tight turns and bends. Substitution of either glycine may simply cause a steric interference within the catalytic site, reducing overall binding of both substrates.

Chapter 5 The carboxyl terminus of adenine phosphoribosyl transferase

5.1 Introduction

5.1.1 Sequence relationships and predicted roles of residues in the carboxyl terminus of adenine phosphoribosyl transferase

The carboxyl terminus of APRT is the region of lowest sequence similarity across species and kingdoms (see Figure 2.1). If we define the carboxyl domain arbitrarily as that region of the APRT protein sequence which lies say 30 residues beyond the core of the highly conserved B sequence (see Chapter 4), the consistent pattern among the sequences of APRTs is that with few exceptions, the carboxyl terminus is not vastly different in length relative to the other members of the family. Consistent with assumptions about preserving functionally essential sequences across long evolutionary time, it is trivial to assert that the carboxyl terminal regions of the various APRT molecules have little sequence content which cannot be substituted. This is in contrast with the inter-species and inter-protein (*i.e.* APRT to HPRT to OPRT) conservation of the B sequences. There is an alternative interpretation of the observed lack of overall sequence conservation. There may be selection for overall length or other generalized properties. In this region there is little pressure on specific residues, but there may be structural elements which can be maintained with many variations in sequence, but exhibit conserved character such as charge, hydrophobicity, or secondary structure. Thus one might predict that the carboxyl region might be susceptible to mutations which are structurally disruptive, or which cause changes in charge or hydrophobic qualities.

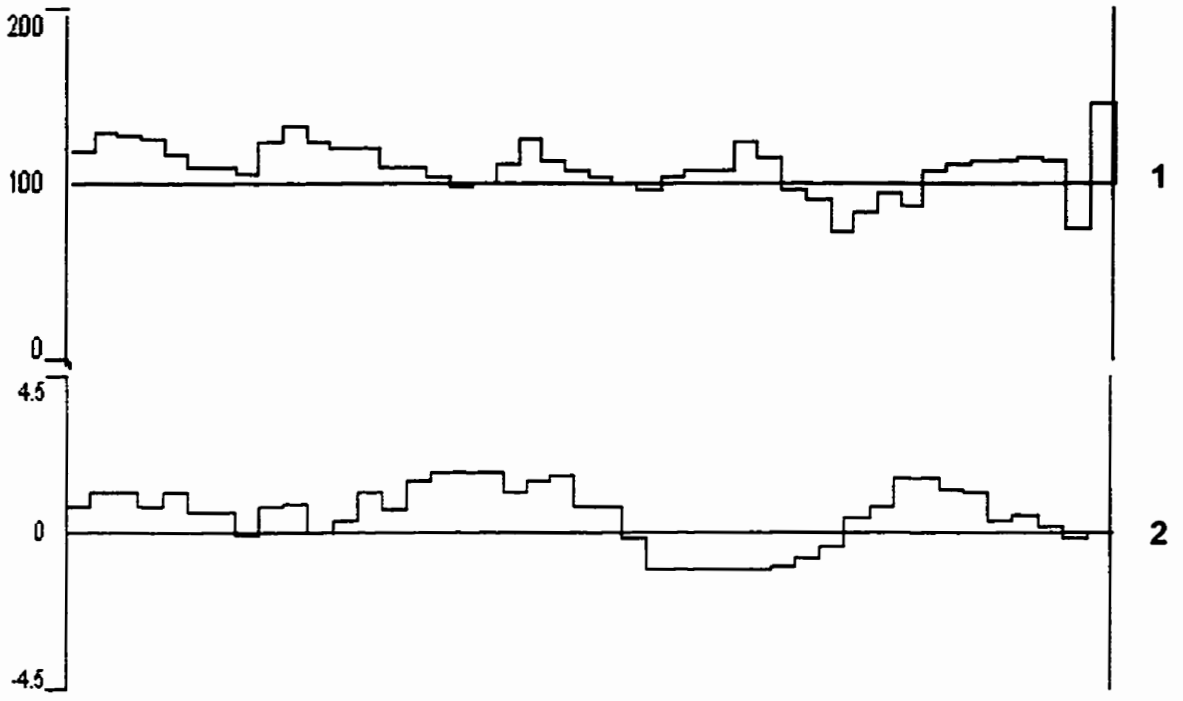
The C-termini of the phosphoribosyl transferases, beyond the catalytic domain, are

not well conserved across species even within the limits of substrates. A surprising ability to preserve function without the native carboxyl terminal domain has enabled research investigating the effect of chimeric protein structures; fusions of one N-terminal domain to a carboxyl region from a different protein. The role of carboxyl termini in enzyme function varies widely from catalytically indispensable (Parissenti *et al.*, 1993; Tsai and Wilson, 1995), interchain interactions (Walker *et al.*, 1992; Katz *et al.*, 1995), maintenance of protein stability (Shortle *et al.*, 1989; Loetscher *et al.*, 1991) to no apparent function at all (Kogure *et al.*, 1996). Some nucleotide-binding proteins have catalytic activity, but altered binding properties in the absence of their C-termini (Zhang *et al.*, 1996). In the case of phosphoribosyl transferases, sequence similarity and structural homologies (or rather, the absence of such) do not suggest essential roles for residues in the carboxyl region distal to the B sequences. Mutations in the C-terminus of APRT which cause residue deletions or substitutions, which would not be predicted to lead to disruption of predicated secondary structure have been identified by phenotypic selection for resistance to adenine analogs in CHO cells (Figure 5.1) (Drobetsky *et al.*, 1989, Phear *et al.*, 1989). These mutant molecules do not competently process the adenine analogs into the cytotoxic nucleotides, but few have been previously characterized for residual adenine-utilization.

One hypothesis is that the C-terminus is critical in structural stability. If the carboxyl domain has function in higher order structure, mutants in this region while not disrupting secondary structure, may alter or impede tertiary or quaternary interactions. If the carboxyl terminus is part of the scaffold of the protein, we would not predict that

Figure 5.1 : Mutants in the region of the carboxyl terminus of APRT. Deletion mutants constructed in this work are shown above the sequence line, mutants from published works below. Termination mutants are indicated *. The underlined gl pair indicates a simultaneous occurrence in a single mutant. Below the sequence figure are shown: 1) the Chou-Fasman alpha helix prediction, 2) Kyte-Doolittle hydrophobicity. These analyses were done using the Protean module of DNAStar (see text).

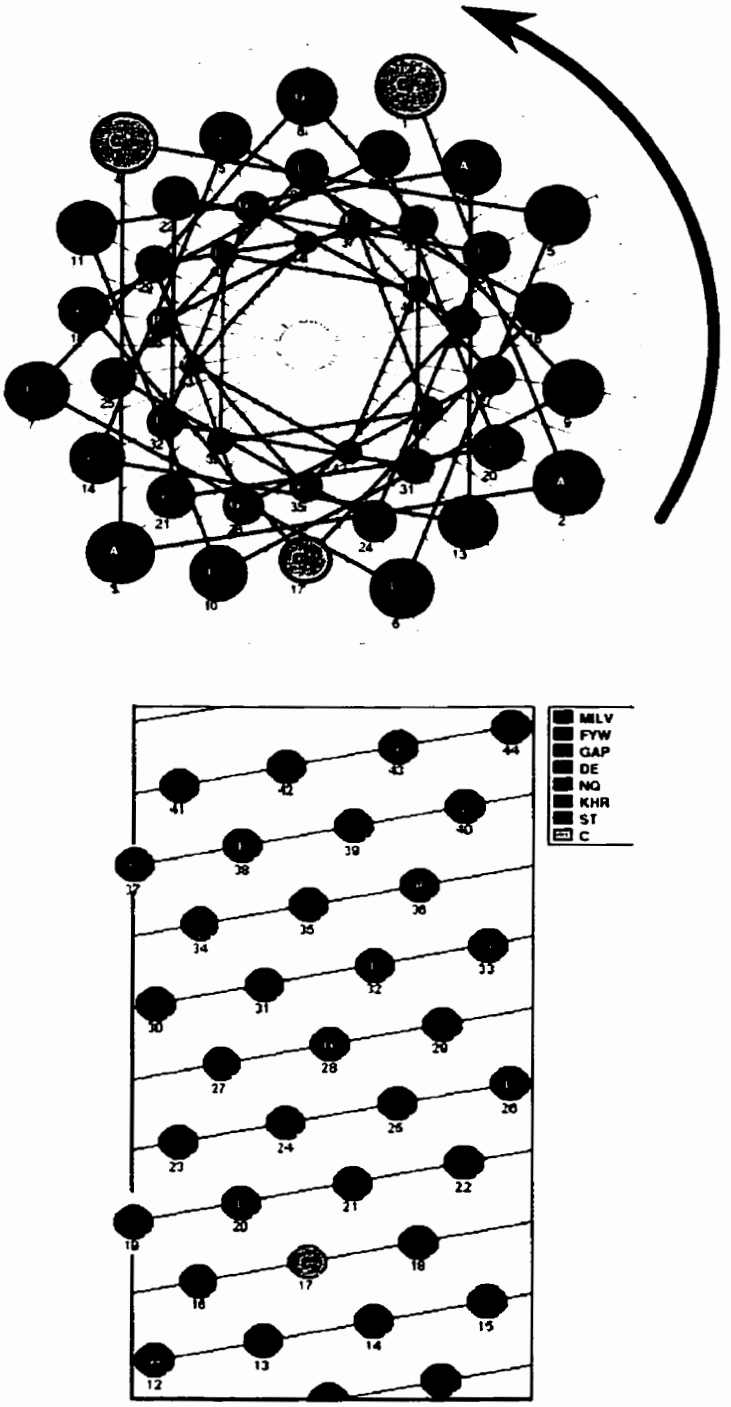
```
 140        150        160        170        180
   | Δ      Δ      Δ   |Δ         |  Δ  Δ    |
AACELLGQLQAEVVECVSLVELTSLKGREKLGSVPFFSLLQYE
qv  qp     p a   ye    k    f d   *rrll  l    cg
v   s       gl              p s          Δ    *
                            a                 l
```



mutations which disrupt secondary structures would lead to functional mutant proteins. Furthermore, since APRT is isolated as a functional dimer, it might be predicted that failure of stabilization of the tertiary structure leads to reduced or absent quaternary (*i.e.* dimeric) structure, with a corresponding loss in enzyme activity. Alternatively, it may be that such mutants cause a structural change leading to rapid proteolytic degradation of the protein, with the result that no detectable activity remains. This observation has been previously reported for specific cases, but a generalized protease sensitivity of mutants is weakly supported in the literature (Kosinski and Bailey, 1992; Davidson *et al.*, 1988). In some cases specific sequence changes, as opposed to generalized structural disruption have been described as responsible for proteolytic sensitivity (Parsell *et al.*, 1989; 1990). It is not clear that protease susceptibility is by any means a satisfactory explanation for apparent loss of activity in mutants causing subtle structural disruption.

Interacting α-helices have been previously described as one type of interaction between polypeptides (Crick *et al.*, 1953; Leon *et al.*, 1997). Such α-helices have been shown to often have a crude heptad repeat structure (Parry, 1982; McLachlin and Dunn, 1997), and ought to exhibit an α-helix with a notably hydrophobic face, such that the interaction between the helices is primarily based upon hydrophobic sidechains (Zhou *et al.*, 1992; McCoy *et al.*, 1997) . From Figure 5.1, we can identify sequences within the carboxyl region which have strongly α-helix-predicting properties. There is within the α-helix region, a distinct surface of hydrophobic residues, as shown in helical-net and helical-wheel projections of Figure 5.2. The hydrophobic surface is within two tracts from 150-159 (6/9 hydrophobic) and 168-177

Figure 5.2 : Predicted alpha helical wheel and net structures of the carboxyl region (Protean, DNAStar). Hydrophobic residues are indicated by black, dark blue, or green colours. The blue arc on the helical wheel indicates the amount of rotation in ther position of an amino acid residue, relative to the helix, when a residue upstream (N-term) is deleted. C-1 corresponds to cysteine 137 of CHO APRT.



(6/9 hydrophobic) respectively. Between these tracts lies a region of α-helix, which has no great hydrophobic nature (1/8 hydrophobic). The more downstream tract contains a proline residue, which may in fact disrupt any helical structure despite the global structural prediction. If this region is involved in dimerization functions, we would predict that single residue deletions within the α-helical segments, which should cause a 120° change in relative position around the helical bundle, would be quite disruptive to the interchain interactions, but might in fact leave the bulk of the protein structure intact.

5.1.2 Identifying alterations in quaternary structure

Our approach to testing this hypothesis is to evaluate whether there are catalytically-competent mutants in possible quaternary interaction domains which may lead to loss of or instability of dimerization functions (Prochownik *et al.*, 1989; Walker *et al.*,1992). This may be detected by either a detectable absence of dimerization under specific conditions, or by abrupt loss of activity under mildly denaturing conditions which do not affect wildtype functions. Dimers may be identified by size-exclusion chromatography, but using this approach for purification has allowed the detection of wildtype molecules with detectable activity at the apparent elution volume of the monomeric molecule (see Chapter 3). It was decided to use a published crosslinking method under solvent conditions similar to those where activities are measured.

5.2 Materials and Methods

5.2.1 *In vitro* mutagenesis of selected carboxyl terminus residues

Using the T7-GEN kit (United States Biochemical), mutants were constructed in several sites corresponding to predicted α-helical segments, either internal or adjacent to

strongly predicting tracts. All restriction and modification enzymes, including buffers and cofactors were from New England Biolabs. From the vector construct pKFA26, containing the *aprt* coding sequence from CHO, a BamHI-HindIII fragment containing the expression cassette (operator, promoter, coding sequence) was directionally subcloned into M13mp19 using standard methods. The clone mp19A26 was characterized with respect to the locations of restriction sites for BamHI and HindIII. The mp19A26 construct was further characterized by dideoxy DNA sequencing (Promega, fMol kit) to confirm the orientation of the inserted sequence.

Using mp19A26 as a substrate, the T7-GEN *in vitro* mutagenesis kit (USB, Batt *et al.*, 1990) was used to introduce site specific changes into the *aprt* sequence. In brief, the mutagenic oligonucleotide is annealed to the single-stranded substrate in the supplied buffer, by heating to 70 °C, followed by cooling at 20 °C. The oligonucleotides used and the carboxyl region deletions they cause are shown in Table 5.1. T7 DNA polymerase, 2 U, are added, along with deoxynucleotides, (including 5-methylcytidine triphosphate). After 30 minutes, T4 DNA ligase (2 Units) and ATP to 1 mM are added. After 30 minutes, HhaI and MspI (1 U each) are added. After 10 minutes, 4 U of Exonuclease III is added. After 10 minutes, the entire reaction if heated to 65 °C for 10 minutes to stop enzymatic reactions. The HhaI and MspI nick DNA at their cut sites where 5-methyl-cytosine is incorporated. Thus only the non-mutagenized parental strand is nicked, and becomes a substrate for Exonuclease III digestion. The effect of the protocol is to reduce the frequency of non-mutagenized plasmids which may form plaques at the transformation step.

Table 5.1 : Oligonucleotides for site-specific mutagenesis of the carboxyl terminus region. Positions of the 3 basepair in-frame deletions are indicated by *.

Amino Acid	Oligo Sequence	Amino Acid Change	Mutant Name
143	gtgagctg*ggccagc	delete L143	ΔL143
150	ggctgag*gtggagtg	delete V150	ΔV151
157	gagcctg*gagctgacc	delete V157	ΔV157
161	gctgacc*cttaaggg	delete S161	ΔS161
173	cagtacca*ttctctctcc	delete F173	ΔF173
176	cattcttctct*ctgcaata	delete L176	ΔL176
115-121	cgccttagaacctggcc	sequencing primer	r345

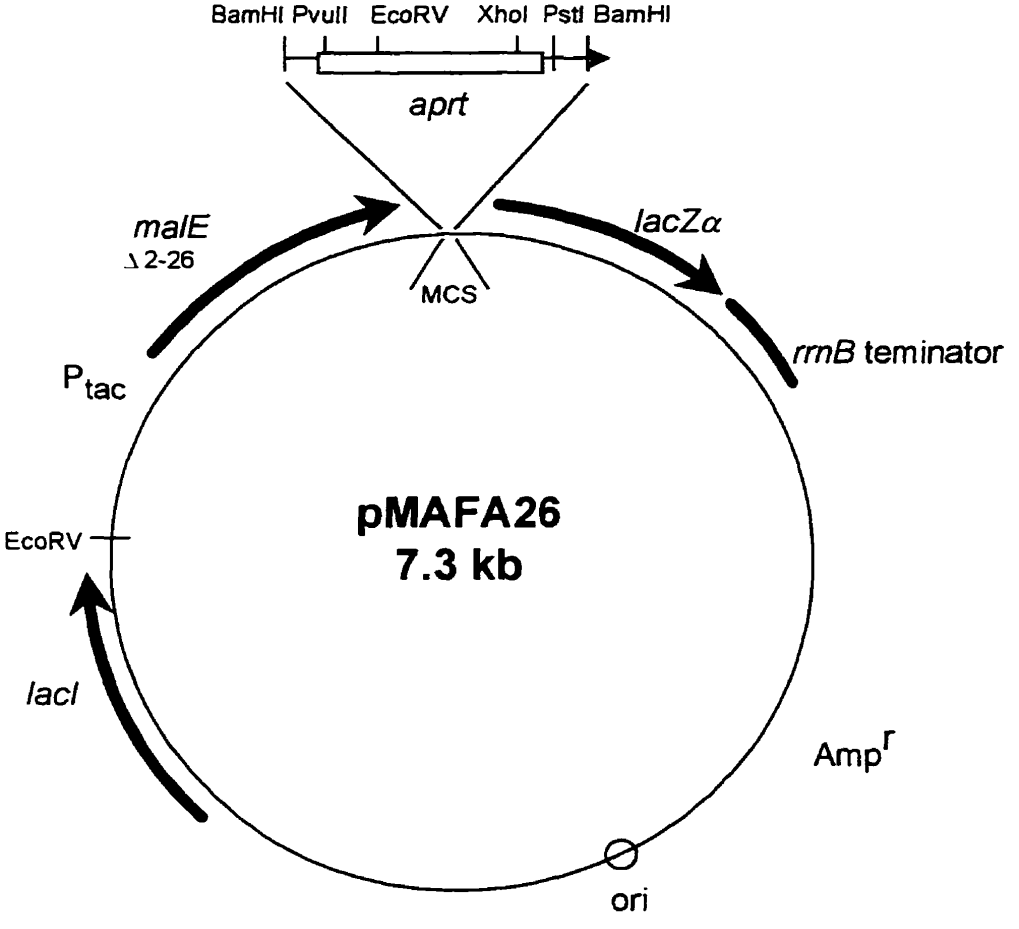
The mutagenesis reaction products were transformed into *Escherichia coli* strain SDM-2 (USB) following the method of Hanahan (1983), then plated in dilutions onto log-phase lawn cells of the same strain (OD_{600} approximately 0.6, in Luria Broth) , to enable plaque formation and detection. From plates with 50 to 100 plaques, 5 plaques for each mutant were chosen for DNA sequencing to determine whether they were mutagenized. In the event that none were mutants, an additional 10 were picked and sequenced.

Mutants confirmed by DNA sequencing were subcloned into pMAF2.6 (Figure 5.3, Appendix II). A XhoI-HindIII fragment containing the carboxyl terminal region, was excised from a double-stranded sample of the mp19A26 mutant clone. The restriction digests were ethanol precipitated, and the DNA pellet after drying, resuspended in 5 µl of Nanopure water. This digested DNA was then cloned into XhoI-HindIII digested pMAFA26, a derivative of pIH1.5 (Appendix 3). Mutant sequences subcloned into pMAFA26 restore the PstI site. Subclones in pMAFA26 were verified by dideoxy sequencing (fMol Cycle Sequencing kit, Promega).

5.2.2 Characterization of carboxyl terminal domain mutants

From clarified extracts prepared by grinding with activated alumina (McIlwein, 1948), aliquots of 1.0 ml, with a protein concentration of approximately 10 mg/ml were applied to amylose-Sephadex resin (New England Biolabs) in 1x20 cm minicolumns (total bed volume 14 ml) , and washed in with 1 column volume of 0.5X TEGNN. The eluate was collected in a single pool, and reapplied to the column, washed in with fresh 0.5X TEGNN. The column was washed with 0.5X TEGNN until the OD_{260} of the eluate

Figure 5.3 : Structure of construct of pMAFA26 (*aprt* (0.6 kb) inserted into pIH902 (6.7 kb)), as described in the Appendix. Note that the maltose binding protein (encoded by *malE*) has a deletion of amino acids 2-26, the membrane-signalling sequence, such that expressed fusion products from pIH902 constructs remain in the cytoplasm. pMAFA26 differs from pMAF15 in that the PstI site at the 3' end of the inserted sequence is removed to facilitate subcloning.



relative to buffer, dropped to below 0.05. 1 ml of 20% maltose was added to the top of the column, then elution with 0.5X TEGNN was continued. Fractions of 20 drops were collected (approx. 0.6 ml) with a fraction collector. OD_{260} was measured for all fractions collected, and detectable peak fractions were pooled into a single aliquot. Pooled eluates were stored on ice until use. From the pooled eluate, 10 ml aliquots were removed as needed and digested with Factor Xa (1 ml/10 ml pooled protein fraction, approximately 250 U). This protease is specific for the linker sequence between APRT and the maltose-binding protein. After digestion overnight at 4 °C, the samples were applied to AMP-Sephadex resin columns of the same format as the amylose resin columns. After washing with 2 column volumes of 0.5 X TEGNN, purified APRT was eluted by addition of 500 µl of 10 mM Mg-PRPP. After initial experiments, AMP-column purification was not done, since the maltose binding protein was not observed to interfere with subsequent assays of APRT. Crude and purified aliquots were analyzed by SDS-PAGE.

SDS-PAGE gels (15% T, 3% C) were run according to the method of Laemmli (Laemmli, 1970; Laemmli and Favre, 1973) on a Bio-Rad minigel apparatus. Separation conditions were 40 mA, 100V, for stacking, and 80 mA, 200 V for resolving. After developing until the dye marker had run out of the gel, the gels were disassembled, and transferred to a staining solution of 0.4% Coomassie R-250 in methanol: acetic acid: water 6:1:3. After staining for 1 hour, the gel was destained by gentle shaking in a tray containing crumpled Kimwipes, in 6:1:3 for up to 24 hours with several changes of solvent (Neuhoff *et al.*, 1988). Gels were subsequently mounted and dried onto cellulose sheets for storage.

5.2.3 Crosslinking of APRT-MAL fusion protein for monomer/dimer differentiation

Crosslinking of purified APRT and mutants was done using the glutaraldehyde method as described by Cheung and Nimni (1982; Wu *et al.*, 1996). To an aliquot (5-10 μg) of protein in a total volume of 100 μl, 13 μl 1 M borate pH 8.9 were added, followed by 8 μl 0.2 M glutaraldehyde (final glutaraldehyde concentration 14 mM). At intervals of 0, 20, 60, 180, 540 seconds, 10 μl of 1 M $NaBH_4$ were added with brisk mixing to stop the reaction and stabilize products. Crosslinking reactions were stored on ice until tracking dye was added and samples prepared for electrophoresis.

SDS-PAGE gels (15% T, 3% C) were run as described above on a Bio-Rad minigel apparatus. Gels were scanned at a resolution of 1200 lines per inch on a Hewlett Packard SJ-1200CX, and images stored for later analysis. These gels were also mounted and dried onto cellulose sheets for storage.

5.3 Results

5.3.1 Deletion mutagenesis of selected carboxyl region residues

The mutants constructed are listed in Table 5.1. Sanger sequencing of the putative mutants confirmed the predicted deletions in each case. Note that the mutations are distributed across the region, and that some are not in strongly helical-predicting segments (Figure 5.1). Figure 5.1 also shows a compilation of other phenotypically-selected mutants in the carboxyl terminal region from several publications (compiled from de Boer and Glickman, 1991). These are aligned to the wildtype sequence, and against the predicted α-helical tracts of the region (Protean, DNASTAR). Also mapped on the wildtype

sequence is the pattern of hydrophobicity and hydrophilicity which is predicted for interchain alpha-helical domains.

5.3.2 Characterization of carboxyl terminal deletion mutants

Relative activities of the isolated and purified wildtype and mutant proteins are shown in Table 5.2. These activities are normalized for protein concentration, from fresh extracts. Note that V150Δ, V157Δ, and L176Δ have very impaired catalytic activity. Mutants L143Δ, S161Δ, and F173Δ have activities in the purified extracts very similar to the wildtype enzyme. Upon cold storage at 4 °C, the activity of the latter three clones was not reduced relative to the wildtype enzyme activity (some activity loss with storage does occur). In frozen cell pellets however, most of the activity of the mutant clones was lost in a relatively short time, whereas the wildtype clones seemed to retain their activity for longer periods in the cell pellet.

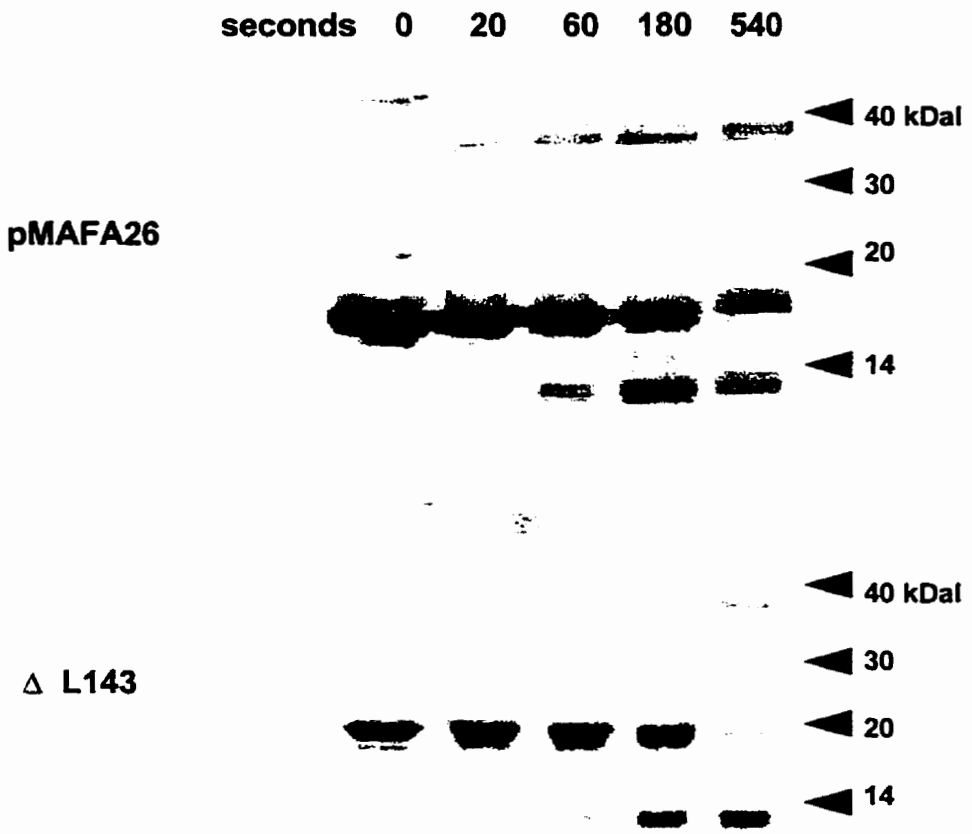
5.3.3 Crosslinked COOH-domain mutants; dimers and monomers

Crosslinking with glutaraldehyde was done for the clone which had substantial activity relative to the wildtype enzyme; L143Δ, S161Δ, F173Δ. Figure 5.3 is a representative series of gel images illustrating the results. Dimeric forms of APRT initially should have an apparent molecular weight of 36-38 kDa. Appearance of such a band, not seen in time zero samples, was taken as evidence of crosslinked dimeric forms of the enzyme. Clone S161Δ exhibited crosslinking kinetics essentially the same as A26 (wildtype), showing detectable dimer crosslinking at the 20 second time sampling. Clones L143Δ and F173Δ exhibit substantially reduced crosslinking to dimeric forms, with no detectable crosslinking until the 540 second samplings. In all cases a band of apparent

Table 5.2 : Relative activities of purified wildtype APRT and mutants expressed as fusion protein from the pMAFA26 derivatives. Assay conditions are as previously described.

Clone	% Relative Activity
wildtype	100
L143Δ	100
V150Δ	20
V157Δ	1
S161Δ	80
F173Δ	80
L176Δ	1

Figure 5.4 : Glutaraldehyde crosslinking PAGE gel images. Time of sample after initiation of the crosslinking is shown in seconds. In the top frame is the wildtype enzyme from pMAFA26. The lower frame is the del L143 mutant.



molecular weight of 12 kDa was seen in the later time samples. The maltose-binding protein, which forms a strong dimeric structure in its own right, was rapidly shifted out of the illustrated range of the gel images, to a molecular weight in excess of 80 kDa. Interestingly, late time samples occasionally exhibited light intensity bands of apparent molecular weight approximately 65-70 kDa. This was interpreted as crosslinking of randomly associating crosslinked dimers (*i.e.* tetrameric forms).

5.4 Discussion

From the above experiments it should be possible to infer that the enzyme is functional as a monomer if APRT activity can be detected under conditions where it can be demonstrated that dimer formation is inhibited or reduced. This begs the question of why APRT (and indeed, other phosphoribosyl transferases) should be polymeric if they can function adequately as monomers. In short, what is the evolutionary advantage of dimeric versus monomeric structure? To answer that there is a stability advantage or kinetic advantage is at best speculative, since there is no strong evidence to support this. It is notable however, that multimeric proteins are by far more common than monomeric proteins.

We hypothesized that APRT is functional as a dimer, but that dimer formation is not necessary or sufficient for catalytic competency. Further, we suggest that the dimerization function may be disrupted by alterations in the carboxyl terminal region, which exhibits little direct sequence conservation among and between APRT molecules from various species. This area, by comparison (using Rasmol 2.6) of the solved crystal structures of *S. typhimurium* or *E. coli* OPRT, human HGPRT, or *B. subtilus* glutamine

phosphoribosyl transferase (GLNPRT) without notable sequence identity in the region, has been shown to be α-helical. In the case of HGPRT, the carboxyl terminal alpha helix lies just adjacent to the interstitial region of the dimeric molecule, whereas in OPRT it is distant from the dimer interface. In the case of glutamine phosphoribosyl transferase, the alpha-helical carboxyl region lies in a groove between adjacent monomers of the tetrameric structure. In *E. coli* XGPRT the carboxyl 20 residues apparently form an unstructured string of residues floating loosely. Thus there seems to be little predictive for the role of the carboxyl region, even though its secondary structure seems to be broadly preserved.

In order to test this hypothesis, *aprt* in a fusion construct with the *E. coli* maltose-binding protein gene has been mutagenized *in vitro* to produce single residue deletions (*i.e.* missing amino acids) in the carboxyl terminal region. These mutants were then expressed, isolated and characterized. Of mutants recovered, some are completely devoid of detectable APRT activity, whereas some are as active as the wildtype enzyme. None of these mutants are in regions of significant sequence conservation (with respect to homologues), thus it was not predicted by this conventional approach that this domain should have an indispensable function. These drastically impaired enzymes are not directly informative about the role of the carboxyl terminus in the overall function of APRT. At most it can be said that some vital function, which might range from dimerization, to proteolytic resistance, to direct catalytic properties depends upon these residues.

From the results it is clear that some of the mutants analyzed have reduced ability to form stable dimers from identical monomers. The selection of mutants is restricted to

the terminal COOH region of the protein for several reasons. First, the domain in question is strongly α-helical in predictive analyses by several algorithms. Alpha helices have been reported to be common components of interchain interactions (Heringa and Taylor, 1997; Hicks *et al.*, 1997; Kammerer *et al.*,1997). Note that there is a rough correspondence of mutant density from mammalian mutant collections and alpha-helix prediction (Figure 5.1). Tracts which are strongly predicting have higher frequencies of mutants, relative to tracts which are less alpha-helical. This observation led us to suspect a structural role for the region. But, compared other APRT primary structures, and other phosphoribosyl transferases, the carboxyl region is poorly conserved. Thus it is likely that the carboxyl domain serves a generalized structural function, rather than a specific catalytic role.

The span of the α-helical region is also characterized by a loose fit to a structure for α-helices involved in interchain domains, as predicted by Crick (1953). The alternation of charged with hydrophobic residues, shown in Figure 5.5, forming a crude repeat element of 7-8 residues over a span of 2-3 repeats fits well with Crick's predictions. This pattern is also identifiable in other APRT sequences, despite the lack of strong sequence similarity in this domain. The implication for deletion mutants of the predicted alpha helical structure at the carboxyl terminus is seen in Figure 5.6. In this figure, the net effects of the chratcerized deletions ΔL143, ΔS161, and ΔF173 are seen. In the case of ΔL143, there is an obvious disruption of a large hydrophobic surface relative to the predicted wildtype structure. The changes caused by ΔS161 and ΔF173 are much less obvious. Of all the COOH domain mutants, ΔL143 has the greatest activity relative to the

Figure 5.5 : Crude heptad repeats in the caroxyl terminus of CHO APRT. The repeats are mainly characterized by properties (charge and size) rather than by sequence identity.

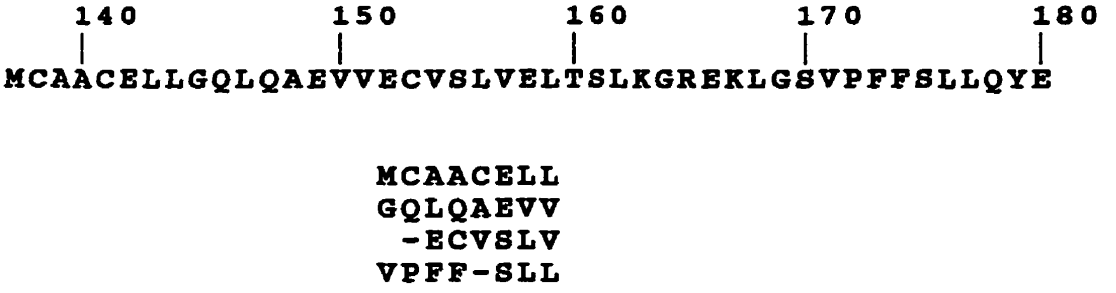
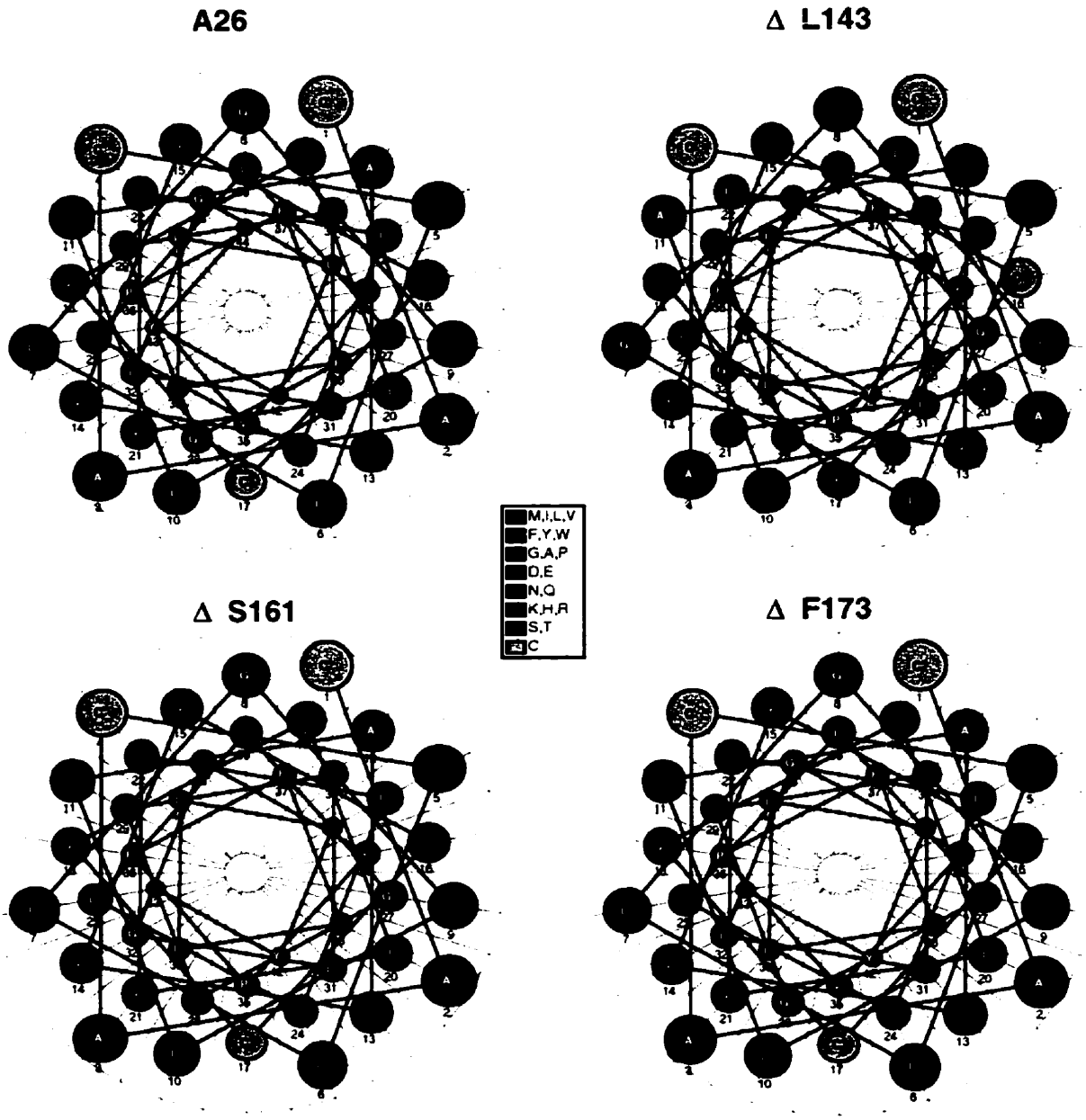


Figure 5.6 : Comparison of COOH terminal helix structure for wildtype A26, Δ L143, Δ S161, and Δ F173 mutants. Note that Δ L143 causes a large rotation in the alpha helix, proximate to the N-terminal end of the helix.



wildtype enzyme, but apparently the poorest capacity to form dimers. ΔL143 causes sufficient local disruption to impede dimerization, but this may be a generalized effect rather than a specific failure of interchain interactions. Whether the carboxyl region helix is involved directly in dimer formation, acting as an interchain binding surface will probably require crystallization and structural analysis of the mutant.

Since it is likely that phosphoribosyl transferases share a common evolutionary ancestor, and most phosphoribosyl transferases are dimeric or polymeric, it is reasonable that they ought to share some conserved structural elements at the secondary structure level as well as at higher levels of organization. The COOH-terminal α-helix may be such an element, and its role in dimer stabilization in APRT may actually represent a relatively recent structural requirement rather than evolutionary heritage. The lack of sequence similarity across phosphoribosyl transferase family boundaries, and the observations in OPRT and HPRT that the carboxyl region alpha helix (which is conserved across families) may have diverse functions supports this proposal.

If the enzyme is functional as a monomer, then what is the structural or kinetic advantage of forming a dimer in solution? In the case of related enzymes, such as HGPRT, higher polymeric states are identified (Holden and Kelley, 1978; Johnson *et al.*, 1979). A gross survey of protein structures form available databases suggests that multimeric proteins are more common than monomeric ones. There may be a distinct stability advantage, in that dimers are less susceptible to denaturation (self-templating for structural form or stability) and subsequent destruction (Green *et al.*, 1995; Shin *et al.*, 1996; Nichols *et al.*, 1997). A speculative extension of this observation is that the

monomers act as co-scaffolds during folding or in solution, such that pre-folded monomers may act as chaperones to new molecules. The hypothesis that dimerization could improve stability is based on the observation that initial protease attack is generally specific rather than generalized (Loetscher *et al.*, 1991; Kosinski and Bailey, 1992; Kosinski *et al.*, 1997). Thus if such sites can be hidden, by dimerizing or polymerizing, then they are substantially less accessible to protease attack, and enhanced stability can be observed. Proteolytic stability is frequently invoked to explain the stability (or lack of stability) of some proteins expressed in heterologous systems (Cheng *et al.*, 1981; Uhlen *et al.*, 1988). There is clear evidence that in *E. .coli* at least, proteolytic susceptibility can be a strong predictor of maximal expression levels (Makrides, 1996).

Chapter 6 The protein filter

6.1 Introduction

6.1.1 The problem of mutant phenotype

In experimental investigations of mutagenesis mechanisms and outcomes, a target gene is often used as a model for extrapolation to more complex situations or as a sentinel marker for a specific target-chemical interaction. A primary goal of mutation research is a taxonomic analysis of mutational data such that for a given genetic target treated with a given mutagen, the spatial distribution, type classification, and relative proportions of mutations may be mapped. Such an analysis is referred to as the mutational spectrum, or mutational fingerprint of the mutagen on that target (Nagao *et al.*, 1996; Sage *et al.*, 1996; Yadollahi-Farsani *et al.*, 1996) . A useful consequence of this type of analysis, is that it should eventually be possible to identify whether a given environmental exposure to a mutagen has caused an accumulation of distinctive mutations in a target sequence (Bartsch and Hietanen, 1996) .

In terms of monitoring human health or exposure to environmental insult, the identification of mutational fingerprints may be a very valuable technique. However, such approaches require the employment of reliable measurement tools for which interassay variation is well understood. Results from individual experiments will need to be compared in order to evaluate responses to mutagens under different conditions of activation, gene expression, or exposure. Such comparisons require certain general assumptions about experimental conditions and target gene responses, including the unproved assertion that under conditions of different selection or expression, subtle

phenotypic differences of individual mutants are not important. It is further assumed that phenotypic differences of mutants selected by different agents will not substantially alter the mutational spectrum of the agent under investigation.

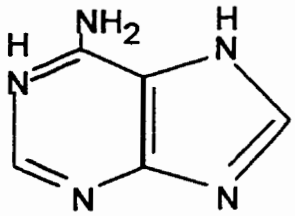
We are interested in the observation that some mutations of *aprt* can be isolated which exhibit phenotypic resistance to adenine analogs, but are still able to utilize adenine. We further question whether some of these mutants may exhibit mixed resistance phenotypes, in that they confer differing degrees of sensitivity to different cytotoxic analogues of the normal substrate. The existence of such mutants must be a consideration in comparisons between experiments, in that the contribution of specific mutants or classes of mutation may be deficient or amplified depending on the experimental conditions.

6.1.2 Structure-function from mutant selection

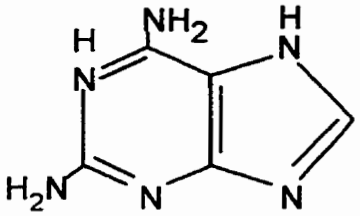
It may also be possible from mixed-resistance mutants to make inferences about the structure or function of residues of the enzyme in the proximity of interaction with the substrates. This approach is of course complicated by the three-dimensional nature of the catalytic or binding site, which is very likely to be composed of many side-chains, several of which when altered, may exhibit steric hindrance of analogue binding, yet not be proximate in the primary structure. For example, if a mutant exhibits sensitivity to 6-chloropurine (6CP) and resistance to 2,6-diaminopurine (DAP), then the basis of that differential sensitivity should involve residues physically close in the normal enzyme to the ring position two (C2) hydrogen on the pyrimidine moiety of adenine. In this case, binding of 2,6-diaminopurine could be restricted sterically by altered side-chains near the active site, interfering with the C2 primary amine of DAP. If we can identify a mutant

Figure 6.1 : Adenine, hypoxanthine and purine analogues used in mPCR selection.

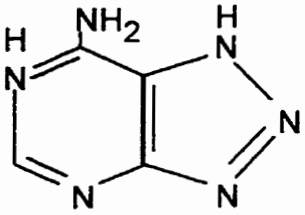
adenine



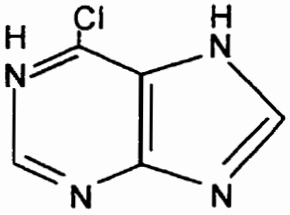
2,6-diaminopurine



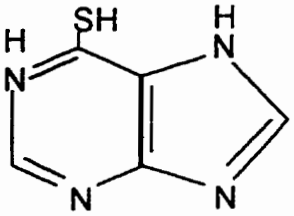
8-azaadenine



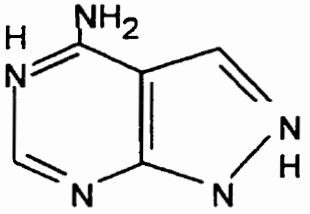
6-chloropurine



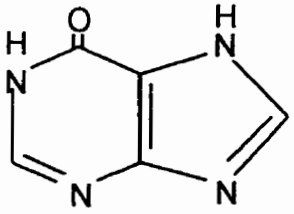
6-mercaptopurine



4-aminopyrazolo(3-4,d)pyrimidine



hypoxanthine



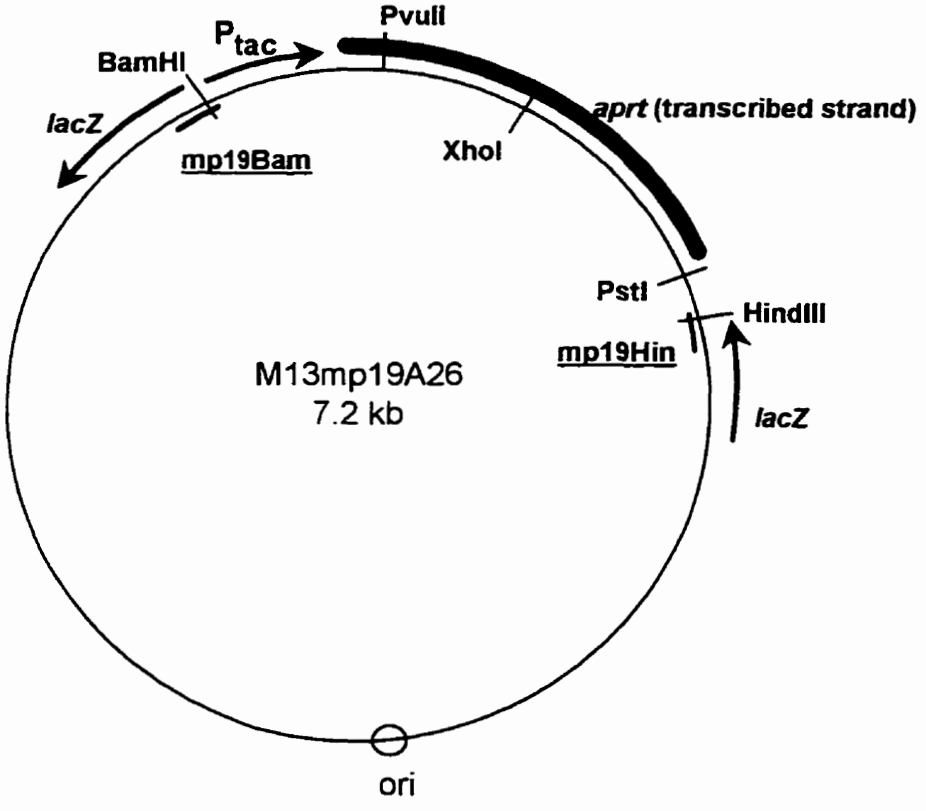
with the inverse behavior, the residues involved, restricting the ring position six (C6) chlorine of 6CP, must be permissive for the C6 amine of adenine and DAP.

6.1.3 Experimental rationale

Polymerase chain reaction (PCR) is a standard method in molecular biology (Mullis *et al.*, 1986). Under typical experimental conditions, fidelity of the Taq DNA polymerase (derived from *T. aquaticus*) in PCR is expected to limit base misincorporations to about $1x10^{-5}$ to $1x10^{-4}$ (Tindall and Kunkel, 1988). In general, unless individual PCR fragments are cloned, any "random" base changes caused by the polymerase will be diluted out against the background of unaltered sequences. There are however, sufficient errors in PCR-amplified material that large scale sequence analyses from PCR DNA are generally done in replicate from original material to reduce reported sequence errors. It is possible to deliberately enhance the frequency of base misincorporations under modified conditions of PCR. The presence of manganese chloride and restriction of one nucleotide create conditions under which misincorporations are common, and may remain unrepaired (Kunkel and Loeb, 1979). Following the method of Leung *et al.*, (1989), we developed a protocol to induce mutations in PCR-amplified *aprt* sequences.

Analogues chosen for assessment, based on structural considerations were 6-chloropurine, 8-azaadenine, 8-aminopyrazolopyrimidine, and 6-mercapto-purine (a presumptive substrate of HGPRT). The structures are shown in Figure 6.1. For each case, the planarity of the structure is preserved. The primary differences between these molecules are substitutions at the six (C6) position of the pyrimidine ring, and in the case of pyrazolopyrimidine or 8- azaadenine, modifications to the nitrogen seven (N7)

Figure 6.2 : M13mp19 subclone of pKFA26 containing the *aprt* expression cassette, including the *tac* promoter. The subclone was constructed by digesting pKFA26 with BamHI and HindIII, then ligating into M13mp19 following standard methods (Appendix III). The position of the primers used in the mutagenic PCR protocol are indicated and labelled as mp19Bam (forward) and mp19Hin (reverse). After mPCR, the XhoI-HindIII fragment with the *aprt* sequence was subcloned into pkF1.0 (Figure 6.2).



or carbon eight (C8) positions of the imidazole ring. Interestingly, 6-mercaptopurine (6MCP) is considered a substrate not for APRT, but for HGPRT (Graf *et al.*, 1976; Pieters *et al.*, 1992; Zu *et al.*, 1997). This position is however not universally accepted. There are many reports of competitive interaction of adenine or adenine nucleotides with 6-mercaptopurine or its nucleotides where hypoxanthine or guanine and their nucleotides have been noncompetitive (Al-Safi *et al.*, 1984; Queen *et al.*, 1989, Dayton *et al.*, 1992). There may be conditional preference of the enzyme for 6-mercaptopurine, based on relative availabilities of cosubstrates. Hypoxanthine, which is not a substrate for APRT, has, like guanine an oxygen at the position corresponding to the primary amine of adenine and to the sulfhydryl group of 6MCP. In DNA, this oxygen forms hydrogen bonds with the amine group of cytosine. Thus the oxygen and amine of hypoxanthine and adenine would be expected to have essentially complementary H-bonding qualities. The group at ring position six may be the primary discriminant for APRT versus HGPRT. But the sulfhydryl group lies between the oxygen and amine groups with respect to its H-bonding qualities. We predict that while 6MCP is indeed a substrate for HGPRT, it may also be an APRT substrate.

6.2 Materials and Methods

6.2.1 Mutagenic polymerase chain reaction

The construction of the APRT-expressing clone has been previously described (see Chapter 3). From the expression construct pKFA26, a 1.1 kbp BamHI/HindIII fragment containing the entire expression cassette was cloned into M13mp19, yielding mp19A26 as shown in Figure 6.2. Purified single-stranded DNA from mp19A26 was

amplified under mutagenic conditions.

Mutations are introduced into the cloned aprt expression cassette under conditions which reduce Taq polymerase fidelity, adapted from the method of Leung et al. (1989) as represented in the flowchart in Figure 6.3. The conditions used for mutagenic PCR are shown in Table 6.1. Note that for each nucleotide (A,C,G, or T), separate reactions are done, which are pooled into a single tube after the PCR reaction is complete. Under these conditions, approximately 1/100 bases may be substituted (Leung *et al.*, 1989).

Primers for PCR were complementary to the extreme ends of the expression cassette, and included the sequences of the restriction sites within the primers such that mutations of the last few bases of the product could be avoided. Oligonucleotide primers for PCR, incorporating intact BamHI and HindIII sites found at the 5' and 3' ends of the expression cassette respectively were as follows: mp19Bam ; GAATTCGAGCTCGGTA CCCG!G ; mp19Hind; ATGACCATGATTACGCCA!AGC, supplied HPLC-purified by Dalton Chemical Laboratories (! Indicates restriction enzyme cut-site).

After PCR, an agarose gel was run to verify the presence of a single band of appropriate size. The PCR product was purified away from excess primer and PCR buffer using a Wizard PCR purification mini-column (Promega). The cleaned-up product was then digested using 2-5 times the recommended amount of enzyme for the amount of DNA in each preparation (15-17 U per μg), then incubated at 37 °C for 18-24 hours at 37 °C. After digestion, the products were heat treated to 65 °C for 10 minutes to inactivate the restriction enzymes, then stored at 4 °C until use.

Figure 6.3 : Flowchart of mutagenic PCR (mPCR) protocol including the subcloning and screening methods. Conditions of the PCR reaction, and selection media are described in the text.

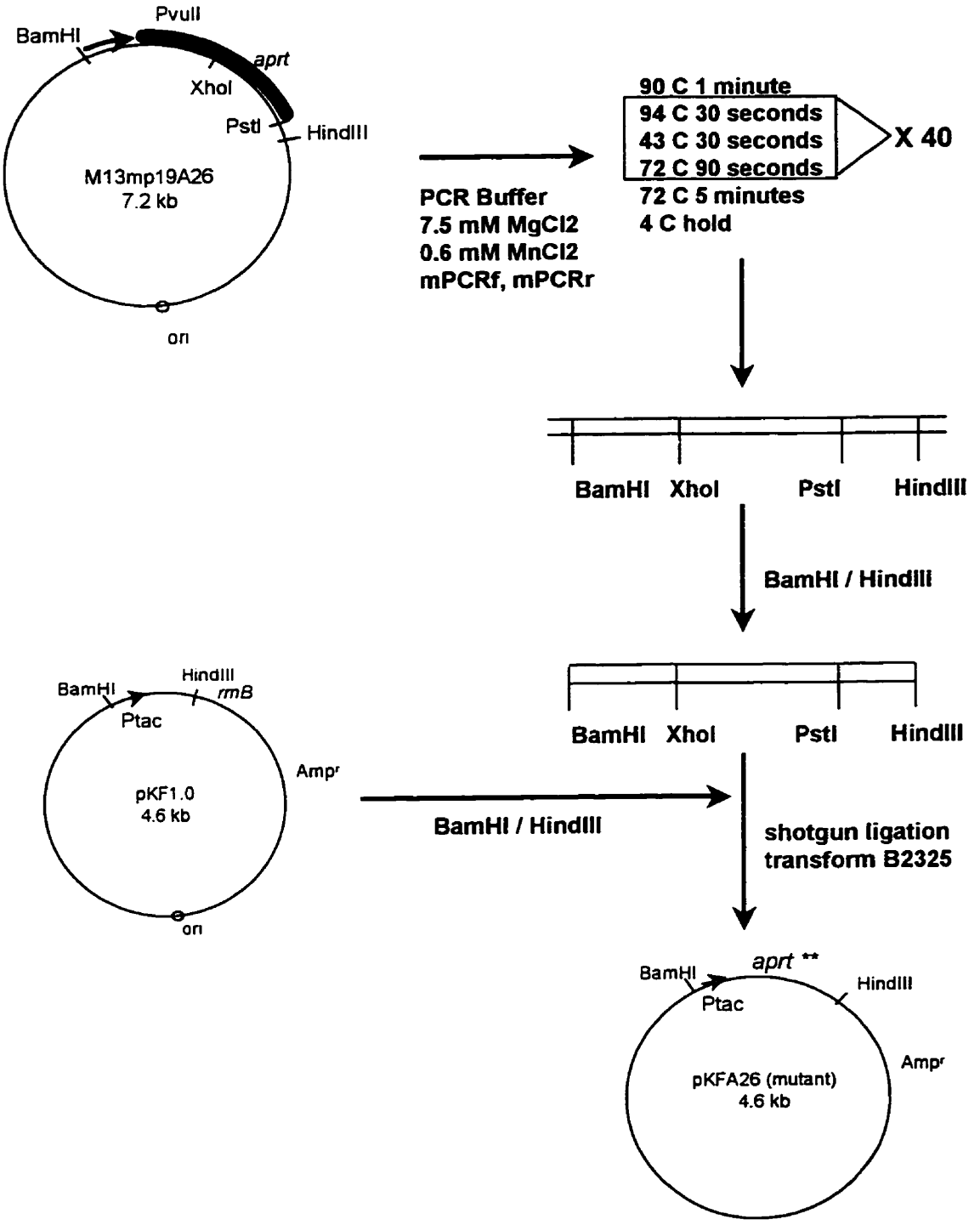


Table 6.1 : Reaction conditions for mutagenic polymerase chain reaction (mPCR). For each experiment, four experimental reactions are run (each deficient in one of the nucleoside triphosphates, dATP, dCTP, dGTP, dTTP), a positive control (no nucleotide deficiencies), and a template-free negative control.

10X mPCR Buffer

10 mM Tris HCl pH 8.3
50 mM KCl
0.01% gelatin
0.01% NP-40
0.01% TritonX100

10X dNTP mixes

	dATP	dCTP	dGTP	dTTP
- A mix	2 mM	10 mM	10 mM	10 mM
- C mix	10 mM	2 mM	10 mM	10 mM
- G mix	10 mM	10 mM	2 mM	10 mM
- T mix	10 mM	10 mM	10 mM	2 mM

Reaction Conditions

1 X dNTP mix (10 µl 10X stock)
1 X mPCR Buffer
10 ng template (mp19A26)
50 picomoles mp19Bam
50 picomoles mp19Hin
7.5 mM MgCl2
0.6 mM MnCl2

PCR Program

90 °C x 1 minute
(begin cycling)
94 °C x 30 seconds
43 °C x 30 seconds
72 °C x 90 seconds
(cycle 40 times)
72 °C x 5 minutes
4 °C hold

6.2.2 Subcloning mPCR mutagenized DNA

Digested PCR products were subcloned into BamHI-HindIII-digested pKF1.0 using standard methods. After 24 hours at 15 °C, the ligation reactions were used to transform B2325 (*aprt- E. coli*) following an adaptation of the method of Hanahan (Hanahan, 1983; Inoue *et al.*, 1990). Transformed cells were plated onto LB agar plates containing ampicillin (100 μg/ml; ICN).

After 24 hours, ampicillin-resistant colonies were picked and gridded onto M9 minimal medium (Difco) containing glucose (0.12 %), 50 μM thiamine, 100 μM adenine, and 100μg/ml ampicillin. Replica-plating was not used, so as to avoid carry-over of rich media or analogues from one plate to the next. 15,000 ampicillin-resistant clones were screened by regridding to selective media from LB plates.

6.2.3 Screening on phenotype selection media

From M9-adenine plates, we picked growing colonies, and regridded adenine-utilizing clones to M9 selection media to look for resistance to adenine analogs. Analogues were suspended in water and solubilized with a minimum volume of 1M NaOH, then added to the media at about 65 °C. The volume of analog in NaOH was insufficient to disturb the final pH of the medium, as verified by pH strip tests on each batch of media. Analogs were used at a final concentration of 4.5 mM (solution limit of 2,6-diaminopurine) in the M9 agar plates. Adenine and hypoxanthine were used at a concentration of 100 μM. Sensitivity or resistance to 6-thioguanine was also evaluated in some clones as a measure of spontaneous background mutant frequencies, indicated as occurrence of resistance to these analogues, substrates of the *E. coli hpt* gene product.

This effect was also monitored in the B2325 *aprt-* strain carrying the vector pKF1.0.

Colonies exhibiting growth within 48 hours were regridded to fresh M9-adenine, and to M9 agar containing either hypoxanthine or adenine, and the various analogues as in Table 6.2. After adenine utilization was confirmed, analogue resistance was assessed by picking cells from growing colonies with a sterile toothpick, and regridding onto selective media, M9 with 100 μM of the purine base (adenine or hypoxanthine), and 4.5 mM of the analogue under test. Because of the possibility of feeding by growing cells proximate to non-growing cells, or of local depletion of analogues, each phenotype was verified by regridding onto new selective plates at low density (approximately 30 clones/plate on a rectangular grid) prior to scoring. Colonial growth was scored for 24 and 48 hours following gridding, based on a rank-order system as indicated in Table 6.3.

6.2.4 Sequence analysis

Mutant phenotype data were tabulated and evaluated by cluster analysis using Ward's method in Statistica. 40 of the clones exhibiting representative growth patterns different by examination from the controls (pKFA26 or pKF1.0 in B2325) were selected for DNA sequencing of the coding region using the Sanger dideoxy method (Sanger and Coulson, 1977) adapted as a kit for thermal cycling (Promega fMol kit).

6.3 Results

6.3.1 Mutants from mutagenic polymerase chain reaction

In total, some 15,000 ampicillin-resistant clones were tested for adenine utilization. After adenine utilization was confirmed, analogue resistance was assessed by picking cells

Table 6.2: Selective growth media used to select mPCR clones, and predicted growth conferred by control plasmids pKF1.0 (no *aprt*) or pKFA26 (*aprt* $^{+}$) . All plates are M9 agar with thiamine, glucose, 100 μM of purine source, 4.5 mM of purine analogue. Mercaptopurine is generally considered a target for HGPRT, for which B2325 is wildtype (*hpt* $^{+}$). Growth predictions indicated by an asterisk may be modified by competition with the purine source.

Code	Purine Source	Analogue	Predicted growth of B2325 with :	
			pKF1.0	pKFA26
ADE	adenine	-	0	3
ADE DAP	adenine	diaminopurine	0	0 *
ADA 8AA	adenine	8-azaadenine	0	0 *
ADE PYP	adenine	pyrazolopyrimidine	0	0 *
ADE 6CP	adenine	6-chloropurine	0	0 *
ADE 6MCP	adenine	6-mercaptopurine	0 *	0
HYPO DAP	hypoxanthine	diaminopurine	3	0
HYPO 8AA	hypoxanthine	8-azaadenine	3	0
HYPO PYP	hypoxanthine	pyrazolopyrimidine	3	0
HYPO 6CP	hypoxanthine	6-chloropurine	3	0
HYPO 6MCP	hypoxanthine	6-mercaptopurine	0 *	0 *
HYPO	hypoxanthine	-	3	3

Table 6.3 Scoring matrix for growth patterns of mutagenic PCR clones which exhibit both adenine utilization and adenine analogue resistance. "Positive control" refers to the strain which would be expected to grow initially under the given selective conditions. For adenine utilization screening, the positive control was B2325::pKFA26 (*aprt*$^-$). For adenine analogue resistance screening , with hypoxanthine as the purine source, the positive control was B2325::pKF1.0.

Observed Growth	Score
Vigourous growth similar to control	3
Slow or little growth relative to control	2
Barely detectable growth	1
No detectable growth	0

from growing colonies and regridding onto selective media with purine base (adenine or hypoxanthine), and the analogue under test. Vigorous growth was estimated relative to control clones with either the wildtype pKFA26 sequence or the vector pKF1.0, depending on the selective conditions. For the purpose of discriminating sensitivity versus resistance to an given chemical, scores of 0 or 1 and 3 were used. Although the ranking of 2 was generally easy to discriminate relative to other ranks, the conservative large-scale difference between 1 and 3 was chosen as the minimum for further assessment.

6.3.2 Phenotypes of *aprt* mPCR mutants

Phenotype data (*i.e.* analogue resistance as indicated by growth scores) for all phentoyped mutants are in Table 6.4. These data are the means of two replicates of the gridding experiments. Because of feeding by live cells proximate to non-growing cells, each phenotype was verified by regridding at low density (approx. 30 clones/plate) prior to scoring. A subset of Table 6.4, mutants for which sequence data was obtained, is given in Table 6.5.

Cluster analysis of phenotype for the subset of mutants from Table 6.5 is shown on Figure 6.4. Due to the ability of Ward's method to smooth differences in rank order data, it is useful for analyzing what would otherwise be a complex problem. Mutants with phenotypes which differ in response to a single agent, but are similar for other analogues are clustered. Mutants which exhibit similar growth for one analogue, but different growth on the remainder do not necessarily cluster together.

Sequence data for the selected mutants are shown in Table 6.6 and mapped onto

Table 6.4 (1 of 10): Phenotype data for all mutants which exhibited adenine utilization and differential analogue sensitivities. The data are not smoothed, and missing data points are empty. Key to media identification: ADE: adenine; DAP: 2,6-diaminopurine; 8AA: 8-azaadenine; PYP6: 4-aminopyrazolo(3-4,d)pyrimidine; 6MCP: 6-mercaptopurine; CP: 6-chloropurine; HYPO: hypoxanthine.

Table 6.4 (cont. 2 of 10)

CLONE	ADE	ADE DAP	ADE 8AA	ADE PYP	ADE 6MCP	ADE 6CP	HYPO DAP	HYPO 8AA	HYPO PYP	HYPO 6MCP	HYPO 6CP
gh 4	3	0	2	3	0	0	3	3	3	0	3
gh6	3	3	2	3	0	0	3	3	3	0	3
gh8	2	0	0	0	0	0	3	3	3	0	3
gh18	3	0	0	0	0	0	3	3	3	0	3
gh22	3	0	0	3	0	0	3	3	3	0	3
gh24	3	0	0	3	0	0	3	3	3	2	3
gh43	3	0	0	0	0	2	3	3	3	0	3
gh47	3	0	3	3	0	0	3	3	3	2	3
gh49	3	0	0	3	0	0	3	3	3	0	3
gh65	3	0	0	3	0	0	3	3	3	0	3
gh67	3	3	2	3	0	0	3	3	3	0	2
gh69	3	3	0	2	0	0	3	3	3	0	3
gh73	3	3	2	0	0	0	3	3	3	3	3
a14	3	3	0	3	2	1	3	3	3	0	2
a32	3	0	0	3	2	2	3	3	3	2	2
a34	2	0	0	3	2	2	3	3	3	2	3
a36	2	0	0	0	0	2	3	3	3	0	3
a38	3	0	0	3	0	0	3	3	3	0	3
a58	2	0	0	1	1	2	3	3	3	0	3
a60	3	2	2	3	1	2	3	3	3	0	3
a62	2	0	1	0	1	1	3	3	3	2	3
a75	3	2	1	3	0	2	3	3	3	0	3
a77	3	0	3	2	0	2	3	3	3	0	3
a79	3	3	0	2	1	1	3	3	2	0	3
a81	3	3	0	2	1	1	3	3	2	2	0
a83	3	0	2	2	0	0	3	3	3	0	3

Table 6.4 (cont. 3 of 10)

CLONE	ADE	ADE DAP	ADE 8AA	ADE PYP	ADE 6MCP	ADE 6CP	HYPO DAP	HYPO 8AA	HYPO PYP	HYPO 6MCP	HYPO 6CP
c9	3	0	0	0	0	0	2	3	2	0	0
c11	3	3	3	2	3	0	3	3	3	0	2
c13	3	0	3	2	3	0	3	3	3	0	2
c15	3	0	0	0	0	0	3	3	3	0	2
b9	3	3	0	0	0	0	2	2	0	0	0
b11	3	0	0	0	0	0	3	3	2	2	0
b13	3	2	3	2	0	0	3	3	2	0	2
b29	3	0	0	0	0	0	3	2	0	2	0
b31	3	3	0	2	0	0	3	2	3	2	2
b33	3	0	0	0	0	0	3	3	0	2	0
b35	3	0	0	3	0	2	3	3	3	2	3
b37	3	0	0	0	0	0	3	3	3	0	3
b53	3	3	3	2	0	0	3	2	3	0	2
b55	3	0	0	0	0	0	3	3	3	0	2
b57	3	0	0	0	0	0	3	3	3	2	2
b76	3	0	0	0	0	0	3	3	0	0	0
b78	3	3	0	2	0	0	3	2	3	0	0
b80	3	0	0	0	0	0	3	3	0	0	0
b82	3	3	0	2	0	0	3	2	3	2	0
k3	3	2	0	3	1	0	2	3	0	1	2
k5	3	3	0	3	1	0	2	3	3	1	2
k7	3	3	2	3	1	2	3	3	0	1	2
k9	3	3	0	3	1	0	3	3	3	1	2
k11	3	3	0	2	1	0	2	3	0	1	0
k13	3	3	0	2	1	0	0	3	0	1	0
k15	3	2	0	2	1	0	2	3	0	1	0
k17	3	1	0	3	1	0	3	3	3	1	2
k19	3	1	0	3	1	0	3	3	0	1	0

Table 6.4 (cont. 4 of 10)

CLONE	ADE	ADE DAP	ADE 8AA	ADE PYP	ADE 6MCP	ADE 6CP	HYPO DAP	HYPO 8AA	HYPO PYP	HYPO 6MCP	HYPO 6CP
k21	3	2	0	3	1	0	3	3	0	1	0
k23	3	2	0	3	1	0	3	3	0	1	0
k25	3	3	0	3	1	0	2	3	0	1	0
k27	3	1	0	0	1	0	0	3	0	1	0
k29	3	3	0	0	1	0	2	3	0	1	0
k31	3	3	0	2	1	0	3	3	0	1	0
k33	3	3	0	2	1	0	3	3	0	1	0
k35	3	3	0	2	1	0	2	3	0	1	0
k37	3	2	0	2	1	0	2	3	0	1	0
k40	3	3	0	1	1	0	3	3	0	1	0
k42	3	2	0	1	1	0	3	3	0	1	0
k44	3	3	0	0	1	0	3	3	0	1	0
k50	3	1	0	0	1	0	2	3	0	1	0
k51	3	3	0	1	1	0	0	0	0	1	0
k53	3	2	0	1	1	0	3	3	1	1	0
k55	3	3	0	0	1	0	2	3	0	1	0
k57	3	3	0	1	1	0	3	3	0	1	2
k59	3	3	0	1	1	2	3	3	2	1	2
k61	3	2	0	1	1	0	3	3	0	1	0
k76	3	3	0	1	1	0	0	3	2	1	2
k78	3	3	0	1	1	0	0	3	0	1	0
k80	3	3	0	1	1	0	0	3	0	1	0
L3	2	2	0	1	1	1	3	2	2	3	0
L5	1	3	2	0	1	0	0	0	0	0	0
L7	3	0	0	1	2	0	3	2	0	3	1
L9	3	2	2	1	1	0	0	0	0	3	0
L11	0	1	0	2	1	0	0	2	0	3	0
L13	0	2	1	0	1	0	0	0	0	0	0
L15	2	2	1	0	1	0	2	2	1	3	1
L17	3	2	0	1	1	0	2	2	0	3	0
L19	2	2	1	0	1	0	0	2	0	3	0
L21	2	3	0	0	1	0	2	2	0	3	0

Table 6.4 (cont. 5 of 10)

CLONE	ADE	ADE DAP	ADE 8AA	ADE PYP	ADE 6MCP	ADE 6CP	HYPO DAP	HYPO 8AA	HYPO PYP	HYPO 6MCP	HYPO 6CP
L23	0	1	0	0	1	0	0	0	0	3	0
L25	0	2	0	0	1	0	3	2	1	3	1
L29	3	2	0	2	1	0	2	2	1	3	0
L31	0	1	0	1	1	0	3	2	3	3	1
L33	1	1	0	0	1	0	2	0	0	0	0
L35	1	1	0	0	1	0	0	1	0	0	0
L37	3	3	0	0	1	0	2	1	0	2	0
L46	3	3	0	1	1	0	2	1	0	3	1
L48	3	2	0	0	1	0	2	1	0	2	0
L50	3	3	0	0	1	0	0	2	0	0	1
M3	3	0		0	0	0	3	2	0	1	1
M5	3	0		0	0	0	2	2	0	1	0
M7	3	0		0	0	0	3	2	0	1	0
M9	3	0		0	0	0	0	2	0	1	0
M11	2	0		0	1	0	2	2	0	1	0
M13	2	0		0	0	0	1	1	0	1	0
M15	2	0		0	1	0	1	2	0	1	0
M17	3	0		0	0	0	0	2	0	1	0
M19	3	0		0	0	0	2	2	0	1	0
M21	2	0		0	0	0	2	2	0	1	0
M23	2	0		0	0	0	2	2	0	1	0
M25	3	0		0	0	0	2	2	0	1	0
M29	3	0		0	0	1	0	2	0	1	0
M31	3	0		0	0	0	0	2	0	1	0
M33	2	0		0	0	0	0	2	0	1	0
M35	2	0		0	0	0	0	2	0	0	0
M37	3	0		0	1	0	3	2	0	1	0
M40	2	0		0	0	0	3	3	1	1	1
M42	2	0		0	0	0	3	2	0	1	0
M44	2	0		0	0	0	0	2	0	1	0
M46	2	0		0	0	0	3	2	0	1	1
M47	2	0		0	0	0	3	2	0	1	0

Table 6.4 (cont. 6 of 10)

CLONE	ADE	ADE DAP	ADE 8AA	ADE PYP	ADE 6MCP	ADE 6CP	HYPO DAP	HYPO 8AA	HYPO PYP	HYPO 6MCP	HYPO 6CP
M48	3	0		0	0	0	0	2	0	1	1
M50	3	0		0	0	0	2	2	0	0	0
M53	2	0		0	0	0	3	2	0	1	0
M55	3	0		0	0	0	3	2	0	1	2
M57	3	1		0	0	0	3	2	0	1	2
M59	3	0	3	0	2	0	3	2	0	1	0
M61	3	2	3	0	2	0	3	2	0	2	0
N 3	3	2	1	0	0	0	2	2	1	1	1
N5	3	2	1	0	0	1	0	2	0	1	0
N7	3	2	1	1	0	0	2	2	0	1	0
N11	0	2	0	0	0	0	3	2	3	1	1
N13	1	2	0	0	0	0	3	2	0	0	0
N15	3	2	1	1	0	1	3	3	0	1	0
N17	3	2	1	0	0	0	2	2	0	1	0
N19	1	2	0	0	0	0	3	2	0	0	1
N21	1	2	0	0	0	0	3	2	0	0	0
N23	3	2	0	0	0	0	2	2	0	1	0
N25	3	2	0	0	0	0	3	2	0	0	0
N29	3	2	0	0	0	1	2	2	0	1	1
N31	0	2	0	0	0	0	3	2	3	0	0
N33	1	2	0	0	0	0	0	2	0	0	0
N35	3	2	0	0	0	0	0	2	1	1	0
N37	3	2	1	0	0	1	3	3	0	2	2
N40	3	2	0	0	0	1	3	3	3	2	0
N42	3	2	0	0	0	0	0	2	0	1	0
N44	1	2	1	0	0	0	0	2	2	1	0
N46	3	2	0	0	0	0	0	2	0	0	0
N48	3	2	0	0	0	0	3	2	0	0	0
N50	3	2	1	1	0	1	3	2	0	2	0
N53	3	2	1	0	0	0	3	2	0	2	0
N65	3	2	0	0	0	1	3	2	1	0	1
N67	3	2	1	0	0	0	3	2	0	2	1

Table 6.4 (cont. 7 of 10)

CLONE	ADE	ADE DAP	ADE 8AA	ADE PYP	ADE 6MCP	ADE 6CP	HYPO DAP	HYPO 8AA	HYPO PYP	HYPO 6MCP	HYPO 6CP
N69	3	2	1	1	0	1	3	2	0	1	1
N71	3	3	0	0	0	1	0	2	0	1	0
N73	3	3	1	1	0	1	0	2	0	1	0
N75	3	3	1	1	0	0	3	2	1	2	1
N77	3	3	0	0	0	1	2	2	0	1	0
N79	3	3	2	1	0	1	3	2	0	2	1
N81	3	3	1	0	0	1	2	2	1	1	1
N83	3	3	1	0	0	1	3	2	3	1	1
O3	2	0	0	3	0	0	0	0	0	0	0
O5	3	0	0	2	0	0	0	0	0	0	0
O7	3	0	3	2	0	0	0	2	0	0	0
O9	3	3	2	0	0	0	0	0	0	2	0
O11	3	3	0	0	0	0	0	0	0	2	0
O13		3	3	3	2	0	0	0	0	0	0
O15	3	3	2	0	0	0	0	0	0	0	0
O17	3	3	2	2	0	0	3	2	0	2	0
O19	3	3	3	3	3	0	0	2	0	2	2
O21	3	3	0	3	0	0	0	2	2	2	3
O23	3	3	2	3	0	0	0	2	3	2	3
O25	3	3	3	3	2	0	0	2	0	0	0
O29	3	3	3	3	3	0	3	2	3	2	2
O31	3	3	3	3	2	0	3	2	0	2	0
O33	3	3	3	3	0	0	0	2	0	2	0
O35	3	3	3	3	0	0	0	0	0	2	0
O37	3	3	3	3	0	0	0	0	0	2	0
O40	3	3	3	3	0	0	0	0	0	2	0
O42	3	3	2	3	0	0	0	0	0	2	0
O44	3	3	3	3	2	0	0	2	0	2	0
O46	3	3	3	3	0	0	0	0	0	2	0
O48	3	3	3	3	0	0	0	0	0	2	0
O50	3	3	3	3	0	0	0	2	0	2	0
O53	3	3	3	3	2	0	0	2	0	2	0

Table 6.4 (cont. 8 of 10)

CLONE	ADE	ADE DAP	ADE 8AA	ADE PYP	ADE 6MCP	ADE 6CP	HYPO DAP	HYPO 8AA	HYPO PYP	HYPO 6MCP	HYPO 6CP
O55	3	3	3	3	0	0	0	2	0	2	0
O57	3	3	3	3	0	0	3	2	3	2	0
O59	3	3	3	3	0	0	0	2	3	2	3
O66	3	2	3	0	0	3	2	2	2	3	3
O68	3	2	3	0	2	0	2	2	0	3	3
O70	3	2	3	0	2	0	3	2	0	3	3
O72	3	2	3	0	2	0	0	2	0	3	3
O76	3	2	3	0	2	0	0	3	0	3	3
O78	3	2	3	0	2	0	3	3	0	3	3
O80	3	2	3	0	0	0	3	3	3	3	3
O82	3	2	3	0	0	0	2	3	3	3	3
O84	3	2	3	0	0	0	2	3	0	3	3
S17	3	1	0	1	0	0	3	2	2	1	2
S19	3	1	0	1	0	0	3	2	2	1	2
S21	3	0	0	0	0	0	3	2	2	3	0
S23	3	1	0	0	0	0	3	2	2	1	2
S25	3	0	0	1	0	0	3	2	2	3	1
S40	3	1	0	1	0	0	3	2	2	3	1
S42	3	1	1	1	0	0	3	2	2	3	0
S44	3	2	1	1	0	0	3	2	1	3	0
S46	3	2	1	1	0	0	3	2	2	3	0
S48	3	0	0	1	0	0	3	2	2	2	0
S50	3	1	0	1	0	0	3	1	1	1	0
S64	3	2	1	3	0	0	3	2	2	1	2
S66	3	2	1	3	0	0	3	2	2	1	2
S68	3	2	0	2	1	0	3	2	2	1	2
S70	3	2	1	3	1	0	3	2	2	1	2
S72	3	2	0	2	0	0	3	2	2	1	2
S85a	3	0	1	0	0	0	3	2	2	1	2
S86	3	1	1	2	1	0	3	2	2	1	2
S88	3	1	0	2	0	0	3	2	2	1	2
S90	3	2	0	2	0	0	3	2	2	1	2

Table 6.4 (cont. 9 of 10)

CLONE	ADE	ADE DAP	ADE 8AA	ADE PYP	ADE 6MCP	ADE 6CP	HYPO DAP	HYPO 8AA	HYPO PYP	HYPO 6MCP	HYPO 6CP
S92	3	1	0	2	0	0	3	2	2	1	2
S99A	3	0	1	2	0	0	3	2	2	1	2
S99	3	0	0	0	0	0	3	2	2	0	0
100a	3	1	0	0	0	0	3	2	2	1	2
T17	3	3	0	0	3	0	3	2	3	0	2
T19	3	3	0	0	3	0	3	3	3	0	2
T21	3	3	1	1	2	1	3	3	3	1	2
T23	3	3	0	0	2	0	3	3	3	2	2
T25	3	3	1	0	0	0	3	3	3	1	2
T26a	3	3	1	0	0	0	3	3	3	0	1
T40	3	3	0	0	0	1	3	3	3	0	2
T42	3	3	0	0	0	0	3	3	3	0	1
T44	3	3	0	0	0	0	3	3	3	0	1
T46	3	3	0	0	0	0	3	3	3	0	1
T48	3	0	0	0	0	0	3	3	3	0	0
T50	3	3	1	1	2	1	3	3	3	0	1
T64	3	3	1	1	1	1	3	3	3	1	2
T66	3	3	0	0	1	0	3	3	3	0	1
T68	3	3	0	0	3	0	3	3	3	0	1
T70	3	3	0	0	2	0	3	3	3	0	2
T72	3	3	0	0	0	0	3	3	3	0	2
T74	3	3	1	0	0	0	3	3	3	0	1
T85a	3	3	0	0	1	0	3	3	3	0	2
T86	3	3	0	0	1	0	3	3	3	0	2
T88	3	3	0	0	2	0	3	3	3	0	0
T90	3	3	0	0	1	0	3	2	3	0	1
T92	3	3	1	1	1	1	3	2	3	1	1
T99a	3	3	0	0	1	0	3	2	3	1	1
T99	3	3	0	0	1	0	3	2	3	0	1
T100a	3	3	0	1	1	1	3	2	3	0	0
V17	3	0	0	1	0	2	0	1	3	1	0
V19	3	2	0	1	0	1	3	0	3	1	1

Table 6.4 (cont. 10 of 10)

CLONE	ADE	ADE DAP	ADE 8AA	ADE PYP	ADE 6MCP	ADE 6CP	HYPO DAP	HYPO 8AA	HYPO PYP	HYPO 6MCP	HYPO 6CP
V21	3	0	0	3	3	2	0	1	3	0	0
V23	3	0	0	0	0	0	0	1	3	0	0
V25	3	0	1	3	3	2	1	1	3	0	0
V40	3	2	0	3	2	1	3	1	3	0	0
V42	3	0	1	3	2	2	2	1	3	0	1
V44	3	2	0	3	3	2	3	1	3	2	0
V46	3	2	0	1	1	2	3	1	3	3	0
V48	1	0	0	0	0	0	3	0	3	1	0
V64	3	2	0	2	0	1	3	0	3	2	0
V66	3	1	0	2	0	1	2	1	3	0	0

Table 6.5 (1 of 2): Table of mutagenic PCR-generated clones showing the scores for growth on the selective media. Scoring was based on rank ordering of growth with respect to control strains B2325::pKF1.0 (*aprt*), and B2325::pKFA26(*aprt*) as indicated in Table 6.1. Clone labels are arbitrary.

Table 6.2 (cont. 2 of 2) :

Clone	ADE	ADE DAP	ADE 8AA	ADE PYP	ADE 6MCP	ADE 6CP	HYPO DAP	HYPO 8AA	HYPO PYP	HYPO 6MCP	HYPO 6CP
a34	2	0	0	3	2	2	3	3	3	2	3
a62	2	0	1	0	1	1	3	3	3	2	3
b37	3	0	0	0	0	0	0	3	3	0	3
b55	3	0	0	0	0	0	0	3	3	0	2
b82	3	3	0	2	0	0	0	2	3	2	0
gh8	2	0	0	0	0	0	0	3	3	0	3
gh73	3	3	2	0	0	0	0	3	3	3	3
k5	3	3	0	3	1	0	2	3	3	1	2
k7	3	3	2	3	1	2	3	3	0	1	2
k27	3	1	0	0	1	0	0	3	0	1	0
O3	2	0	0	3	0	0	0	0	0	0	0
O9	3	3	2	0	0	0	0	0	0	2	0
O15	3	3	2	0	0	0	3	0	0	0	0
O19	3	3	3	3	3	0	3	2	0	2	2
O29	3	3	3	3	3	0	3	2	3	2	2
O31	3	3	3	3	2	0	0	2	0	2	0
O57	3	3	3	3	0	0	3	2	3	2	0
O59	3	3	3	3	0	0	3	2	3	2	3
O64	3	2	3	2	1	0	2	3	0	1	2
O68	3	2	3	0	2	0	0	2	0	3	3
O 78	3	2	3	0	2	0	0	3	0	3	3
S21	3	0	0	0	0	0	3	2	2	1	0
S40	3	1	0	3	0	0	3	2	2	2	1
S44	3	2	1	2	0	0	3	2	1	2	0
S46	3	2	1	2	0	0	3	2	2	2	0
S48	3	0	0	2	0	0	3	2	2	2	0
S50	3	1	0	3	0	0	3	1	1	2	0
T23	3	3	0	2	2	0	3	3	3	2	2
T99a	3	3	0	3	1	0	3	2	3	1	1
V46	3	2	0	1	1	2	3	1	3	2	0

Figure 6.4 : Cluster analysis of adenine-utilizing *aprt* mutants, by phenotype of analogue resistance.

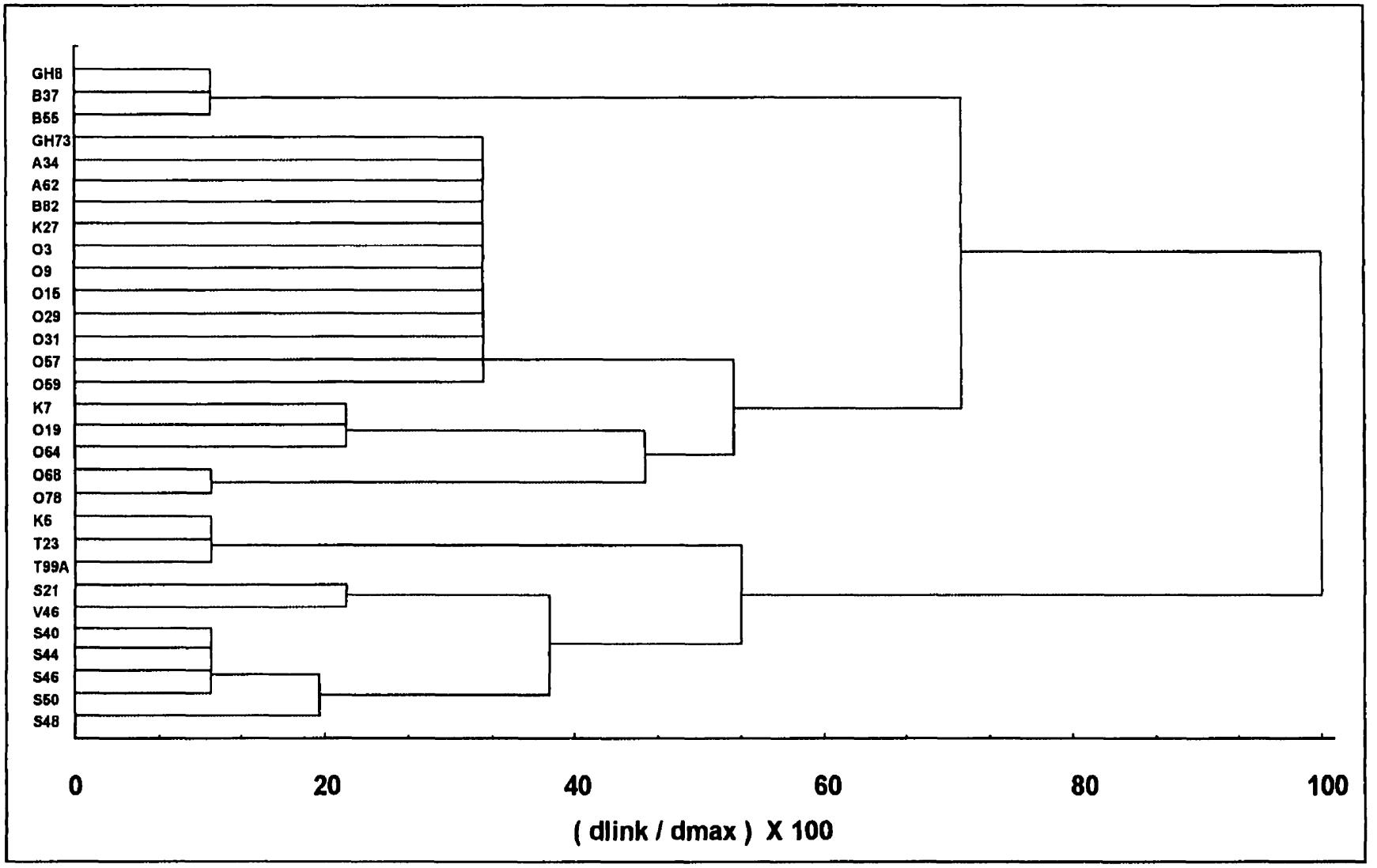


Table 6.6 (1 of 2): Mutants recovered and sequenced from cloned mutagenic PCR products. Clones which were sequenced but which did not exhibit changes in the coding sequence are not shown.

Clone	*Base change* *Amino acid change*						
a34	416 c->t A->V						
a62	151 g->c V->L						
b37	310 g->t E->term						
b55	345 c->t silent						
b82	152 t->c V->A	345 c->t silent					
gh8	101 a->g K->R	116 t->c F->S	208 t->c silent	327 a->g E->D	333 a->g E->D	338 a->c Q->P	409 t->c C->R
gh73	37 a ->c I->L						
k5	443 c->a A->D						
k7	305 c->a A->D						
k27	420 t->g C->W						
O3	418 t->c C->R						
O9	25 c -> g L->V						
O16	25 c -> g L->V						
O19	365 a->g K->R						
O29	298 t-<c S->P	365 a->g K->T					
O31	285 c-<t silent	298 t->a S->T					
O57	25 c -> g L->V						
O59	25 c -> g L->V						

Table 6.3 (cont. 2 of 2):

Clone	*Base change* *Amino acid change*			
O64	365 a->t K->I			
O68	426 g->c silent			
O 78	416 c->g A->G			
S21	154 t->c S->P	306 t->c silent	484 c->a L->I	
S40	99 g->a silent	230 a->g Q->R	452 t->a V->E	492 c->t silent
S44	443 c->a A->D			
S46	46 t ->c F->I	148 c->a silent	351 a->g silent	365 g->a K->R
S48	35 g -> a R->H	362 a->g Q->R		
S50	6 g ->a silent	168 c->a silent	492 c->t silent	
T23	316 g->t G->C			
T99a	79 a->t R->W	175 g->a D->N	292 t->c S->P	357 t->g silent
V46	344 a->t D->V			

the *aprt* sequence in Figure 6.5. There are several mutants, gh8, O29, S21, S40, S46, S48, and T99a, with multiple base changes leading to amino acid changes, and some with base changes which lead to silent mutations. Mutants b82, gh8, O31, S21, S40, S46, and T99a have silent mutations in addition to mutations causing amino acid changes. Most of the mutants have a single mutation in the coding sequence, leading to an amino acid change. Clones b55 and O68, have single silent mutations, while S50 has three silent mutations. One mutant, b37, with a nonsense mutation (termination codon) was recovered.

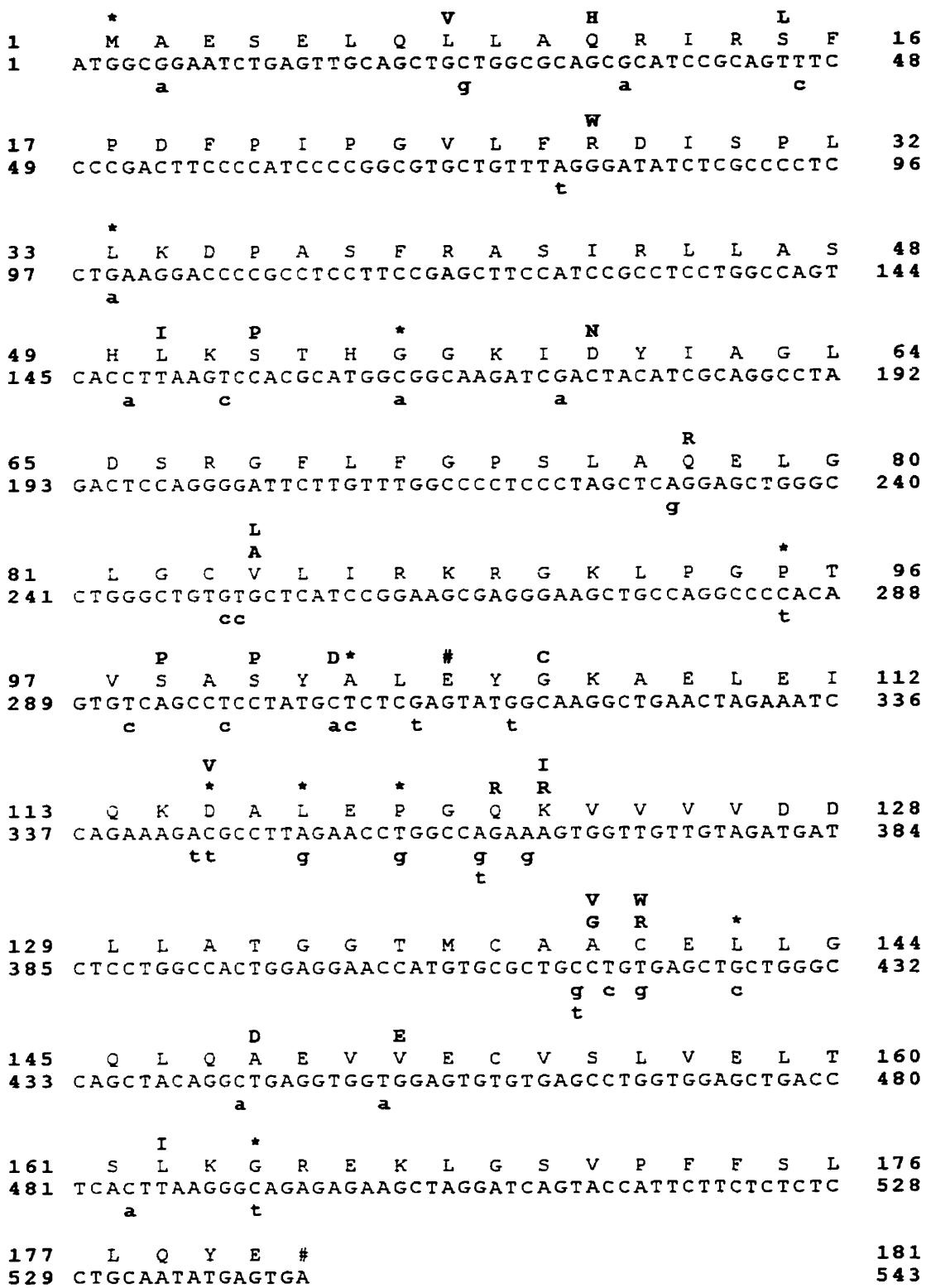
6.4 Discussion

An underlying assumption in cross-experiment comparisons is that a mutant is a mutant. The data here clearly indicate that this assumption is an approximation, but it may not be grossly incorrect. From Table 6.4 it is clear that for any given mutant, with some exceptions, there may be a complex phenotype of mixed resistance to adenine analogs. All of the mutants were selected based upon their ability to utilize adenine, then for their resistance to analogs. Recent work has identified mutants of *aprt* which were initially characterized as deficient in APRT activity, but upon direct assay, found to have activity, although substantially less per unit of protein than the normal or wildtype enzyme (Khattar and Turker, 1997; Khattar *et al.*, 1997). Some of these mutants have also been shown to have differential sensitivity to adenine analogues.

Mutant clones which are resistant to various analogues and also able to utilize adenine are identified. Phenotypic sensitivity to one analogue is in general, a poor predictor of resistance to any other. Mutants with mixed resistance phenotypes are the

Figure 6.5 (1 of 2): Position of mutations on the *aprt* sequence, indicating amino acid changes predicted by translation. (* indicates silent (no amino acid change),# indicates nonsense codon)

Figure 6.5 (2 of 2):



commonest class, but there are few examples of direct correlation. For some mutants, sequence proximity seems to be correlated with phenotypic similarity, as in the case of clones b37 and b55. This correlation is particularly interesting because b55 contains what should be a silent mutation, yet has a phenotype different from the wildtype and vector controls. This may be due to a tRNA suppressor expressed under the growth conditions selective for APRT activity. A clear example of positional correlation with related phenotype is also seen in the cluster including clones O57, O59, O16, O9, and gh73. Notably, this latter cluster also includes clones a34, O29, and O31, which have mutations very distant in the sequence from the remainder of the cluster. Thus closeness of mutation position also appears to be a rather weak predictor of phenotypic similarity, since mutants with amino acid changes close together in the primary structure sequence of the protein, do not necessarily cluster together according to phenotype, and may not have similar phenotypic patterns by examination. Indeed, mutants close together in sequence may have distinctly contradictory patterns, as in the case of clones b37 and k7. Conversely, mutants distant from each other in the primary structure may exhibit quite similar patterns of resistance.

It is clear that mutation data from experiments using different analogues for selection may not be easily comparable, even when the same mutagenesis strategy is used, or when the same sites of mutation might be expected to be recovered. It may be useful to think of resistance to a given analogue as a discrete character, and the combined pattern of resistance to analogues as a set of characters defining a certain mutant. The effects of multiple substrate phenotypes in *aprt* are unlikely to be freely translatable to *hprt*, *lacI*, or

other targets of mutagenesis. But it is clear that, in experiments where selectable mutant phenotypes are scored based on an inferred, but unmeasured, deficiency of gene product activity, this inference may be naive.

These assumptions also permeate analyses of evolutionary history. It is generally assumed that in highly similar homologous target sequences, enzyme mutants which exhibit specific phenotypic properties may be as easily (or not) tolerated in one cellular system as in another. In other words, it is assumed that these altered phenotypic properties will be subject to roughly equivalent selective pressures. In the evolutionary history of a protein such as APRT, there must have been many mutational events which caused loss of function. If these mutations conferred a selective disadvantage to the organisms carrying them, they would rapidly have disappeared from the population. It is likely however that mutations which cause amino acid changes are context sensitive . If the mutations occur in synonymous positions in homologous but different APRT sequences, say between different species A and B, it is possible that the APRT molecule in species A will become nonfunctional, and that in species B might retain its function. If we ignore subtle effects, this mutation would tend to be removed from species A by selective pressures, but not from B. The net effect of such context-sensitive mutations is to generate increasingly divergent sequences.

The evolutionary history of APRT has provided many divergent yet catalytically interchangeable models of the APRT protein. In Chapter 2, sequence analysis of APRT from many species shows that many very different sequences could perform the functions of the CHO enzyme, which in turn can function to replace the native *E. coli* enzyme under

typical growth conditions (Chapter 3). We can simulate the evolutionary process in the laboratory to probe structure and function relationships in our molecule of interest by deliberately increasing mutation frequencies, and by applying selective pressures. Such approaches have already been applied to the selection of novel RNA molecules (Hager *et al.*, 1996; Lohse and Szostak, 1996) and peptides (Roberst and Szostak, 1997) which exhibit catalytic properties . At the protein level such an approach is more technically challenging, due to the need to obtain sufficient protein quantity to test phenotypic properties of novel molecules. At present, bulk synthesis of entire proteins the size of APRT is technically impossible, but the mutagenic PCR approach, combined with *in vivo* expression of cloned sequences and the ability to apply selective pressure to the expressing cell clearly represents a viable alternative approach.

Chapter 7 Structure and function relationships in CHO APRT

7.1 Summary and conclusions

APRT is a member of a family of enzymes widespread in nature, which transfer the phosphoribosyl moiety of phosphoribosyl pyrophosphate, to cosubstrates ranging from nitrogenous bases and amino acids, to nucleotides. APRT catalyzes the addition of phosphoribosyl to free adenine in the presence magnesium, resulting in the formation of adenosine monophosphate (AMP). APRT deficiency in humans has a range of clinical effects, from gouty arthritis to urolithiasis (Kelley *et al.*, 1970; Emmerson *et al.*, 1973; Fox and Kelley, 1973; Fox, 1976).

In addition to adenine, APRT can catalyze phosphoribosyl transfer to analogs of adenine which after reaction with PRPP yield cytotoxic nucleotide analogues. These analogues have been investigated for chemotherapeutic potential in the treatment of neoplasia (Hitchings, 1950; Hitchings and Elion, 1963; Bhalla *et al.*, 1984; Chei *et al.*, 1992), parasitic diseases (Nakamura and James, 1951; Queen *et al.*, 1989), hematological disorders (Parks *et al.*, 1973; Hashimoto *et al.*, 1991, gout, and others (Roblin *et al.*, 1945; Natsumeda *et al.*, 1984).

The ability of an enzyme to discriminate and bind substrates in a highly specific manner is based upon interactions between amino acid side chains in the enzyme and various atoms in the substrate. To date little has been done from the molecular perspective. X-ray crystallographic structures are available only for a few phosphoribosyl transferases, and none for APRT, which has proven difficult to crystallize. These enzymes have predicted structural features similar to other nucleotide binding proteins such as dehydrogenases

(Argos *et al.* 1983), kinases, and oncogene proteins (Fry *et al.*, 1986; Valencia *et al.*, 1991).

In the absence of solved crystal structures, comparison of conserved sequence tracts, motifs and structural predictions between functionally related proteins can be informative about the role of residues and domains in the function or structure of the enzyme (Chapter 2). These conserved regions presumably localize catalytically or structurally indispensable regions under strong selective pressure. The human enzyme for example, differs from the CHO enzyme by 16 conservative amino acid changes. Diverse organisms have undergone divergences of sequence at APRT as great within eukaryotic systems, as between eukaryotes and prokaryotes, as exhibited by the divergence of sequence between *Arabadopsis*, *Homo sapiens*, and yeast.

From sequence comparisons to many other phosphoribosyl transferases clear sequence similarities can be identified, especially in the region of the "B" motif, also found in nucleotide binding proteins. This region has been hypothesized to bind PRPP by interaction with the pyrophosphate at C-1 of the ribose. But this sidegroup has a different shape and relation to the ribose than the pyrophosphate of nucleotides, thus it was hypothesized that the highly conserved "B" sequence, which is likely to serve essentially the same function in most nucleotide binding proteins, actually interacts with the phosphate of the C-5 carbon, distant from the catalytic event around C-1. This has been recently supported by crystal structures of OPRT with bound PRPP.

Analysis of isolated APRT has revealed a burst of AMP synthesis when adenine is added to the purified enzyme (Kenimer *et al.*, 1975; Nagy and Ribet, 1977; Arnold and

Kelley, 1978). In addition, the final product of the reaction exhibits inverted stereochemistry around the C-1 of the ribose (Popjak, 1970; Chelsky and Parsons, 1975). This suggests that there is a multistep interaction of the PRPP with the catalytic site, where the reactive intermediate is moved from one reactive side-group to another. From the reaction schematic, and the model of the surface-walk proposed by Spector (1982), it is apparent that the catalytic site of APRT must contain multiple groups to which the PRPP moiety must bind.

Single residue substitutions of highly conserved aspartates in the "B" sequence have been constructed. Some of these enzymes exhibit decreased affinity for PRPP but unchanged affinity for adenine, while exhibiting substantial catalytic activity (Chapter 4). If these aspartates formed the core of the catalytic event, one would expect a complete loss of catalytic activity, which is not seen. Replacing both aspartates however, causes almost complete loss of activity, consistent with a role in binding PRPP, but not participating directly in catalysis. In addition, substitution of either of the conserved glycines at the carboxyl end of the "B" sequence causes decreased affinity for both substrates, with preservation of much of the enzyme activity. In OPRT, these residues appear to interact with the C-5 phosphate of PRPP, in apposition to the aspartates of the "B" sequence, but remain in proximity to the product, unlike the aspartate residues. Thus even though these mutants exhibit decreased affinity for precursors, the net catalytic rate is probably supported by the decreased affinity for the product. That the catalytic groups are actually distinct from the substrate binding groups, implying that the C-1 end of the ribose is free to react with multiple side-chains, is consistent with our predictions for the role of

the conserved aspartates of the B sequence, and with structural data from OPRT.

Direct physical analysis of purified APRT from several systems, reveals that APRT is a dimeric molecule composed of identical monomers (Dean *et al.*, 1968; Srivastava and Beutler, 1971; Thomas *et al.*, 1973). The observed heterogeneity of quaternary structure among phosphoribosyl transferases suggests that these molecules may have different structural elements for determining or stabilizing the assembled multimer. We hypothesized that the carboxyl terminal region might play a role in dimerization functions, since such a structural role would have relaxed sequence requirements, and could be maintained in the context of a relatively unconserved primary structure. The carboxyl region is the least conserved region of phosphoribosyl transferases, yet mutants which are apparently nonfunctional have been reported in this region (Chapter 5). If dimerization of CHO APRT, which could involve domain swapping or essential stabilization, is an essential component of the overall enzyme function, then most mutations in the region which cause failure of dimerization lead to loss of activity. In the case of the deletion mutant Δ L143, the protein has impaired dimer formation, but nearly full activity relative to the wildtype enzyme. Other deletion mutants in the carboxyl region, Δ S161, and ΔL176 exhibit loss of much of the activity of the wildtype enzyme, but without notable loss of dimerization function. The carboxyl region clearly has an impact upon dimer formation or stability, but the dimerization function is dispensible for catalysis. Thus a role for domain swapping, as has been postulated for the carboxyl region in OPRT, does not seem important in APRT.

In Chapter 6, direct selection of mutants exhibiting resistance to adenine

analogues, different from that of the normal enzyme under the same conditions, revealed that many mutants would be undetected in certain selection conditions. From Table 6.4 it is clear that for any given mutant of APRT, there may be a complex phenotype of mixed resistance to adenine analogs. All of the mutants were selected based upon their ability to utilize adenine, then for their resistance to analogs. Recent work has identified other mutations of *aprt* which were initially characterized as having no APRT activity, but upon direct assay, found to have less activity than the normal enzyme (Khattar and Turker, 1997; Khattar *et al.*, 1997). Some of these mutants have also been shown to have differential sensitivity to adenine analogues. Phenotypic sensitivity to one analogue is in general, a poor predictor of resistance to any other.

An interesting observation of mutagenesis experiments is the occurrence of silent mutations (base changes which should not lead to amino acid changes) with selectable phenotypes. The mutagenic PCR and selection experiments of Chapter 6 also yielded some of these mutants, which exhibited analogue resistance patterns different from the wildtype. It may be useful to think of the combined pattern of analogue resistance as a set of characters defining a certain mutant. It is clear that mutation data from experiments using different analogues for selection may not be easily comparable, even when the same mutagenesis strategy is used, or when the same sites of mutation might be expected to be recovered. Despite the caveat for comparing experimental outcomes, the class and type distributions of mutants with quite different resistance patterns from the mPCR experiments, were not significantly different.

Appendix I Construction of pKF1.0

Introduction

One requirement for expression of mammalian genes in *E. coli* is the use of bacterial expression regulators such as transcription promoters and start sequences, operator sequence for modulation of expression. In addition, the cloning vehicle must contain a suitable bacterial origin of replication, and preferably a selectable antibiotic resistance marker which can be used to maintain the plasmid in the bacterial cell. The plasmid pKK223-3 (Amann *et al.*, 1983; Frost *et al.*, 1984) obtained from Pharmacia (Cat. No. 27-4935-01) satisfied these requirements. pKK223-3 is however, limited by the lack of unique restriction sites in the multiple cloning site (MCS), several of those present being also present in the plasmid sequences. Before pKK223-3 could be used for further cloning experiments, it was necessary to modify the MCS to remove repeated sites and insert novel unique restriction sites.

Materials and Methods

Plasmid pKK223-3 was obtained from Pharmacia as solution in TE (10mM TrisHCl, 1 mM EDTA, pH 8.0). This plasmid construct is represented in Figure 8.1. Note that the multiple cloning site (MCS) contains several restriction sites, including EcoRI, SmaI, BamHI, SalI, PstI, and HindIII. Duplicate sites which are represented within plasmid sequences beyond the MCS are also indicated in the figure. In order to maximize the usefulness of the vector, the MCS was replaced using standard techniques as shown in Figure 8.2. pKK223-3 DNA was digested with EcoRI and HindIII, to remove the entirety of the MCS sequences. After digestion, agarose electrophoresis loading dye

Figure 8.1 : Structural diagram of pKK223-3 (Pharmacia). Note the presence in the MCS of sites for BamHI, AvaI, and SalI which are also present in the remainder of the plasmid. In addition, only the EcoRI site of the MCS is near the optimal distance for expression from the P_{tac} promoter. *rrnB* is a strong transcription terminator form the *E.coli* 5S rRNA gene.

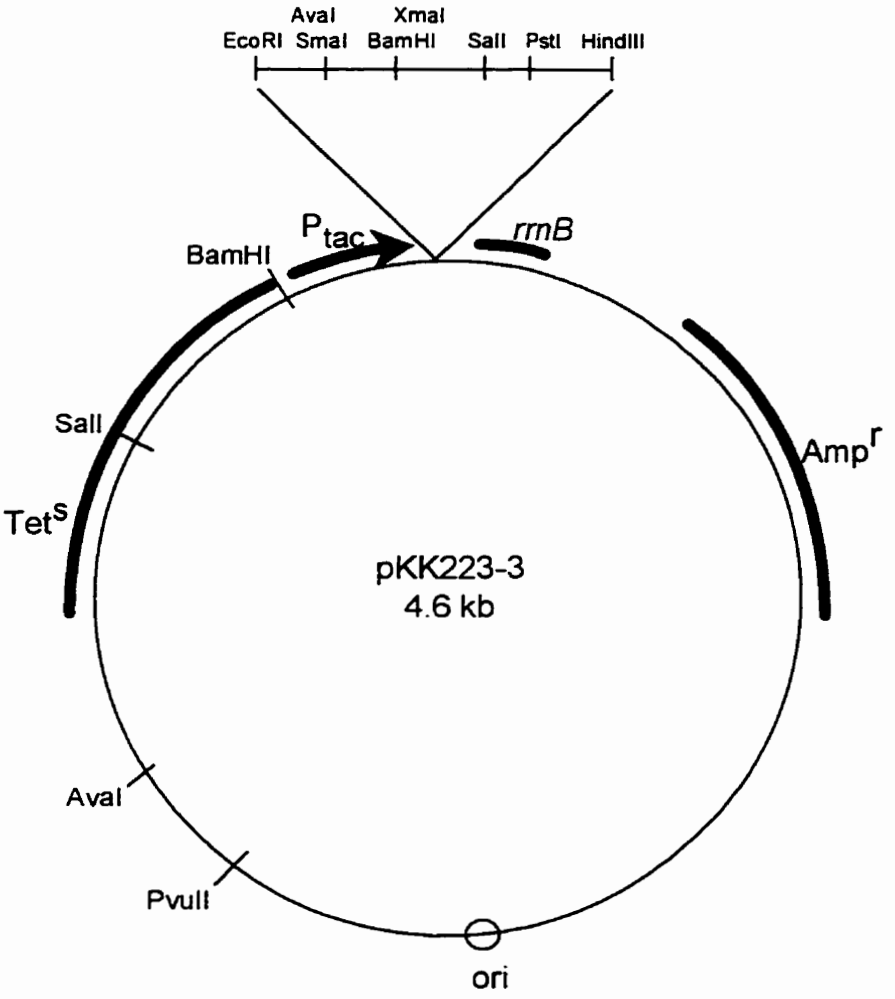
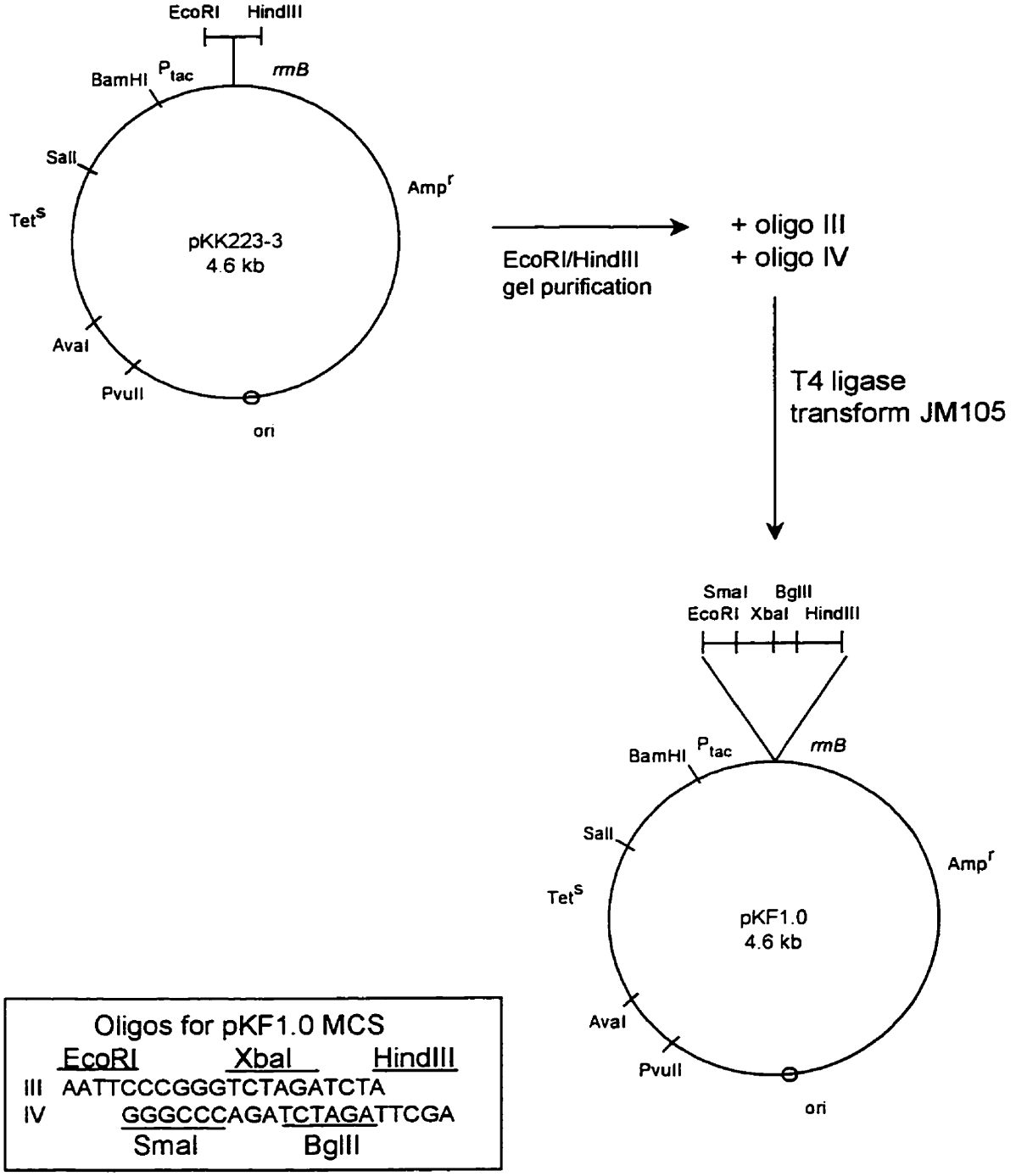


Figure 8.2 : Construction of pKF1.0 as described in the text. The sequences of Oligos III and IV are shown in the box at lower left. The recognition sequences for each enzyme are indicated a solid line next to the labels.



(6X : 60% glycerol, 0.01% bromophenol blue, 0.01% xylene cyanol) was added to a final concentration of 1X. Next the plasmid DNA was electrophoresed on 0.8% low EEO agarose, in 1X TAE, at 10 V/cm of gel width. A prepared DNA sizing ladder (1 kb ladder, BRL)was loaded in a separate lane to allow size determination of fragments. After the dye front marker had migrated approximately 8 cm, the gel was stained with ethidium bromide (1 μg/ml) in 1X TAE for 20 minutes, followed by a brief rinse in distilled water. On a 340 nm transilluminator, the band corresponding to the main plasmid fragment (approximately 4.5 kb) was excised with a scalpel, and the gel containing the band (volume approximately 400 μl), was transferred to a clean microcentrifuge tube. This sample was heated to 70 C in a heat block to melt the gel. 500 μl phenol (equilibrated with 1M TrisHCl pH 8.0) was added to the molten gel, the tube was closed, then vortex mixed for 30 seconds, then immediately centrifuged at 14,000 g for 10 minutes at ambient temperature. The clear upper layer was transferred to a fresh tube, and 2 volumes of cold 95% ethanol were added. After mixing, the tube was centrifuged at 14,000g for 30 minutes. The supernatant was decanted to waste, and the pelleted DNA rinsed with 70% ethanol, followed by a brief drying period in room air. The dried pellet was resuspended in 50 μl Nanopure water, OD_{260} measured, and stored at 4°C until use.

To introduce the novel MCS sequences, oligos III and IV ligated into the digested pKK223-3 plasmid DNA. The sequences of the oligos used are shown in Figure 8.2. 5 μl of each oligonucleotide (in water at OD_{260} approximately 1.0, pMol each) were mixed, and added to 10 μl of the pKK223-3 DNA (0.3 μg). To this was added 5 μl 10X ligation buffer (supplied by the ligase manufacturer), 5 μl 10 mM ATP, 18 μl Nanopure water.

This solution was mixed, and 2 µl T4 DNA ligase (Pharmacia) was added. The reaction was incubated at 14 °C for 18 hours. A parallel control reaction, without oligonucleotide added was also prepared and incubated. After incubation the ligation reactions were stored frozen at -20 °C until use.

Competent cells were prepared following an adaptation of the method of Hanahan (1983). *E. coli* JM105 (Pharmacia) from a fresh overnight culture in SOB (2.0% tryptone (Bacto), 0.5% yeast extract (Difco), 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$), were inoculated into fresh 250 ml SOB in a 1000 ml Erlenmeyer flask, and incubated with vigorous shaking at 37 °C, until the culture reached an OD_{600} of approximately 0.6. The cells were cooled on ice for 10 minutes, then collected by centrifugation in 250 ml polyethylene bottle at 1000 x g. The cell pellet was resuspended with 100 ml of TFB (10 mM potassium acetate, 10 mM PIPES pH 6.7, 100 mM RbCl, 45 mM $MnCl_2$, 10 mM $CaCl_2$, 3 mM hexamine cobalt chloride), incubated on ice for 10 minutes, then repelleted by centrifugation as before. The pellet was resuspended with 12.5 ml of TFB, 200 µl DMSO was added, then the suspension left on ice for 10 minutes. An additional 200 µl DMSO was added, then the cell suspension was dispensed in 1 ml aliquots into 1.25 ml Nunc cryovials, followed by storage in liquid nitrogen.

To thawed 200 µl aliquots of the prepared competent cells on ice, in Falcon 2059 tubes, 10 µl of each of the experimental and control reactions were added with gentle mixing. After 20 minutes on ice, the tubes were transferred to a 42 °C water bath for 90 seconds, then returned to the ice. 800 µl SOC (SOB with 0.4% glucose) were then, added, and the tubes incubated at 37 °C for 30 minutes with shaking. Aliquots of 50, 200,

and 500 µl were plated to LB agar (10% tryptone (Bacto), 5% yeast extract (Difco), 10% NaCl, 15 g/l agar (Difco)) containing 100 µg/ml ampicillin (ICN), and incubated overnight at 37 °C. The following day, individual colonies were selected and picked with a sterile loop into 5 ml of fresh LB, with 100 µg/ml ampicillin for subsequent overnight culturing.

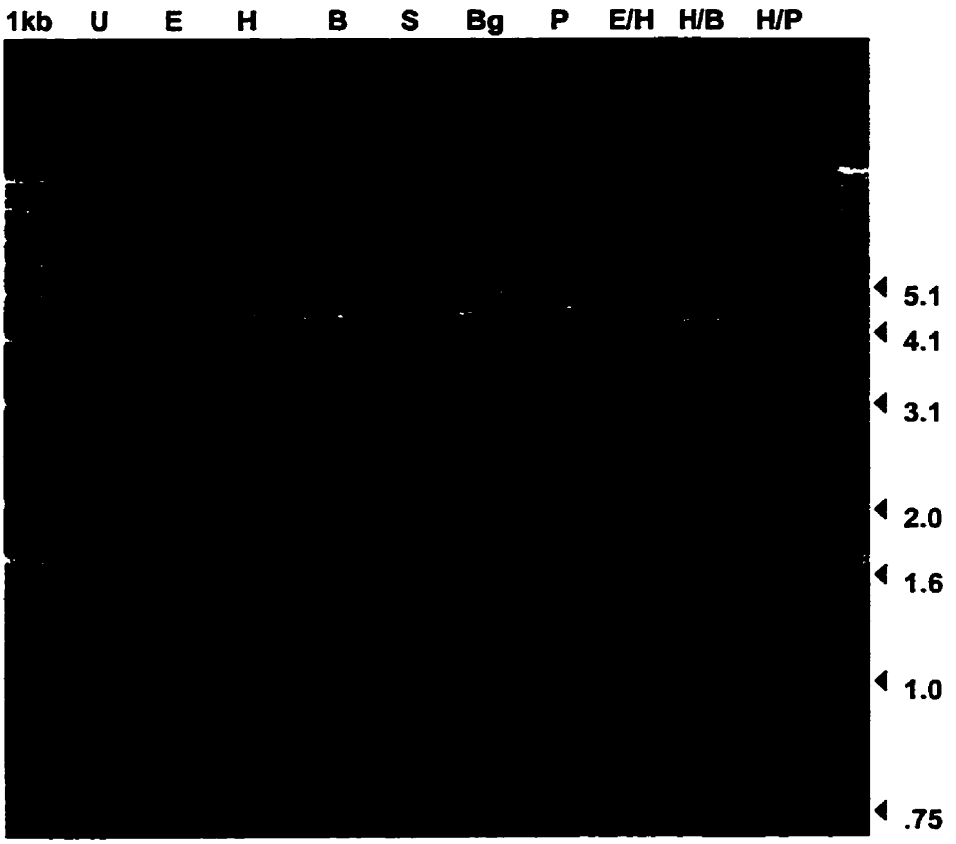
From overnight cultures, DNA was prepared by alkaline lysis as follows. 1.5 ml of the overnight culture was transferred to a fresh tube. Cells were pelleted by centrifugation at 14,000 x g, the supernatant discarded, and the cells resuspended with 200 µl of freshly prepared Solution 1 (50 mM glucose, 25 mM EDTA pH 8.0, 4 mg/ml lysozyme (Boehringer Mannheim)). After 5 minutes at room temperature, 400 µl Solution 2 (0.8% SDS, 0.15 M NaOH) were added, followed by mixing by inversion. When the solution had cleared, 300 µl Solution 3 (60 ml 5 M potassium acetate, 11.5 ml glacial acetic acid) were added, followed by gentle mixing. The suspension was centrifuged at 14,000 x g for 10 minutes, and the supernatant was transferred to a fresh tube. The supernatant was extracted twice with phenol:choloroform (1:1), followed by ethanol precipitation as described above. Dried DNA pellets were resuspended in 50 µl Nanopure water, and stored at 4 °C until analysis.

Plasmid DNA was assessed for the insertion of the new MCS by restriction digestion with BamHI, EcoRI, BglII, SmaI, HindIII, XbaI, and PvuII as described above. Electrophoresis in 0.8% agarose in 1X TAE was performed, with a DNA sizing standard (BRL 1 kb ladder)in flanking lanes. The gel was stained with 1 µg/ml ethidium bromide in 1X TAE, rinsed, and photographed on a transilluminator.

Results and Discussion

Figure 8.3 is a photograph of a diagnostic gel of pKF1.0. The direction of migration during electrophoresis is indicated, as are the sizes of the bands in the DNA sizing standard ladder. In pKF1.0, the MCS contains novel single sites for BglII, PstI, and XbaII which are not contained in the MCS of pKK223-3. Note that there are single restriction sites for BamHI, HindIII, EcoRI, SmaI, XbaI, and PvuII in pKF1.0. The BamHI and PvuII sites are not contained in the MCS, as indicated by the double digests of PvuII/HindIII and BamHI/HindIII which both yield fragment sizes in excess of the size of the MCS (approximately 600 and 350 bp respectively). pKK223-3 exhibits multiple cut sites for BamHI, and PvuII, including those found in the MCS. The promoter to EcoRI-site distance is the same in both pKK223-3 and pKF1.0, and no other markers or sequences were changed.

Figure 8.3: Diagnostic restriction digests of pKF1.0. Sites illustrated are: U, uncut pKF1.0. E; EcoRI. H; HindIII. S; SmaI. Bg; BglII. P; PvuII. E/H; EcoRI-HindIII double digest. H/B; HindIII-BamHI double digest. H/P; HindIII-PvuII double digest. At the right side, the sizes of the DNA sizing standard bands are indicated in kb. Note that the SmaI digestion has not gone to completion.



Appendix II Construction of pMAFA26

Introduction

In order to express certain mutants under conditions which do not necessarily require enzyme activity, such that APRT can be quickly purified without extensive intermediate steps, *aprt* from pKFA26 was subcloned by J. Holcroft of our laboratory, into a fusion protein expression vector, pIH902 (New England Biolabs). This vector contains multiple cloning sequences downstream from and in-frame relative to maltose-binding protein (MBP) from *E. coli*. The entire fusion protein chimera is regulated by the Ptac promoter (*lac*UV5-*trp* hybrid promoter construct; Amann *et al.*, 1983), such that in a *lacI*^{q} host cells, construct expression can be regulated by IPTG induction. Fusion product is extracted using alumina grinding of frozen cells (McIlwein *et al.*, 1948), and is quickly purified on a maltose-Sephadex resin as described below. The protein of interest is then separated from the fusion maltose binding protein by a proteolytic cleavage, and can then be used as is, or separated away from the MBP by a second column step. In general this second step is unnecessary, since MBP has a very specific binding functionality, which is readily saturated by the maltose present in the buffer. After proteolytic cleavage then, the expressed protein, MBP, and maltose are present in the final extract.

Materials and Methods

From pKF15 (see Chapter 2), a fragment containing the entire *aprt* coding sequence was amplified by PCR using primers F1 (ccagg!cctatggcgg) and F2 (gcaatatgagtgaagg!cctcc). These primers amplify the entire coding sequence, producing a

fragment of approximately 560 bp, with StuI restriction sites (marked as ! in the primer sequences) at each end. This fragment was purified away from primers using a minicolumn purification system (Promega Wizard PCR Prep). Vector pMAL-c (new England Biolabs) was digested with StuI, and the large vector fragment was gel-purified (see Chapter 2). The *aprt* PCR fragment was mixed with undigested vector in an approximately 2:1 ratio, in a total volume of 20 µl. Prepared competent cells, *E. coli* DH5αF1 (BRL), were transformed with 5 µl of the experimental and control reaction following the method of Hanahan (1983), as directed by the manufacturer. Transformed cells were plated in serial dilution onto LB agar with 100 µg/ml ampicillin, and 10 µg/ml X-Gal, and incubated overnight at 37 °C.

From ampicillin-resistant transformants exhibiting a white color, several individual colonies were picked and inoculated into liquid LB with 100 µg/ml ampicillin. These cultures were grown overnight with shaking at 37 °C. From overnight cultures, DNA was prepared by alkaline lysis as described previously. Dried DNA pellets were resuspended in 50 µl Nanopure water, and stored at 4 °C until confirmatory restriction digests were performed. From a pMAL-c clone which contained the *aprt* sequence, as determined by restriction digest analysis, attempts were made to extract and purify fusion protein according to the methods described by New England Biolabs, but were unsuccessful due to insufficient spacing between the maltose-binding protein sequences and the APRT protein sequences in the final fusion product. In order to resolve this issue, the *aprt* coding sequence was removed by a BamHI digestion, and gel purified after electrophoresis in 0.8% agarose. The gel-purified fragment was then ligated into BamHI-

digested pIH902 (a modification by NEB of pMAL-c, with an extended polyasparagine spacer peptide). Digested vector and gel-purified *aprt* fragment were mixed in an approximately 1:2 molar ratio, with 5 µl 10X T4 ligase buffer, 5 µl 10 mM ATP, and Nanopure water to a volume of 48 µl. 2 µl T4 ligase (Pharmacia), approximately 14 units, were added, and the reaction incubated at 14 °C overnight. Ligation products were subsequently used to transform DH5αF1, and putative clones selected and grown overnight in LB with 100 µg/ml ampicillin.

Putative subclones were digested with EcoRV, BamHI and XhoI, and SacI-HindIII in a double digest, then electrophoresed on 0.8% agarose in 1X TAE, to verify the size and orientation of the inserted fragment. A DNA sizing standard (BRL 1 kb Ladder) was run in adjacent lanes, so that the size of fragments could be reliably estimated. Once pMAF15 was verified to have the correct size and orientation of insert, it was retransformed into DH5αF1, and frozen stocks prepared.

In order to facilitate subcloning procedures, pMAF15 was subsequently modified to remove a PstI site near the 3' end of the inserted sequence, but outside the coding sequence. pMAF15 was digested with PstI, and the cut end was filled in by treatment with T7 DNA polymerase (USB) in the manufacturer's supplied buffer, and 500 µM of each dNTP. The treated DNA was then ligated with T4 ligase at 15 °C overnight, and transformed into DH5αF1 competent cells. Several clones were picked and the removal of the PstI site, versus its presence in pMAF15, was verified by restriction digestion, followed by 1%agarose gel electrophoresis. The clone selected and verified was subsequently named pMAFA26.

For isolation and purification of APRT, DH5αF1[pMAF15] was grown from frozen stock at 37 °C with shaking, in 1000 ml LB with 100 μg/ml ampicillin until OD_{600} approximately 1.0 was reached. Fusion construct expression was then induced with IPTG at 4 mM for 4 hours. The culture was then distributed to 650 ml centrifuge bottles, and centrifuged at 2500 x g for 20 minutes. The resulting cell pellet was resuspended in a minimum volume of 0.5X TEGN (10X: 100 mM TrisHCl, pH 8.0, 5 mM EDTA, 10 mM MgCl2, 500 mM NaCl, 50% glycerol). The suspension was transferred to fresh 50 ml centrifuge tubes and frozen at -80 °C until use. For crude extract preparation, a cell pellet was weighed, thawed briefly, and transferred to a prechilled mortar on ice. A weight of activated alumina (Sigma Chemical Company) equal to that of the cell pellet was then added. The powder and cell pellet were macerated with the pestle until a thick amber-colored paste was formed. This was transferred into a fresh 50 ml tube, which was then centrifuged at 3000 x g for 10 minutes at 4°C. The supernatant was transferred to a 35 ml Oakridge tube, and centrifuged at 20,000 G for 60 minutes at 4°C. 10 ml of the resulting clarified extract was applied dropwise to an amylose resin column (total column volume 15 ml), which was then washed with 3 column volumes of 0.5X TEGN. 1 ml of 1% maltose in 0.5X TEGN was then added, followed by 0.5X TEGN, and aliquots of 0.5 ml collected for a total of 1 column volume. Aliquots with peak OD_{280} values were pooled. To a 1 ml aliquot of the pooled column eluates, 20 μl of Factor Xa (20 μg) were added, and the mixture incubated at 4°C for 24-26 hours. APRT activity assays were performed on Factor Xa-treated aliquots as described by Okada *et al* (1988). Protein quantitation was performed using a modified Coomassie blue binding assay as described by Bradford *et*

al., 1976).

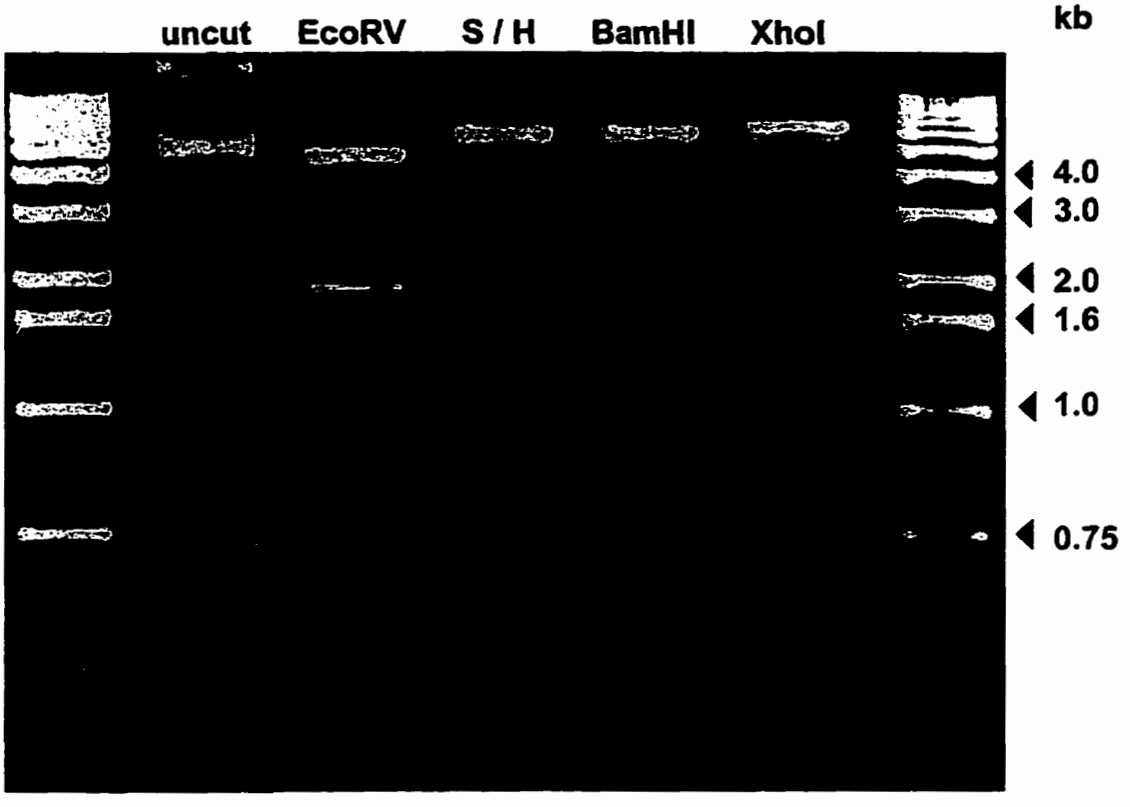
Results and Discussion

In Figure 9.1, the predicted structure of an appropriate pIH902-*aprt* construct is illustrated. In Figure 9.2, the results of restriction digest analysis of one of the putative subclones in the pIH902 vector, pMAF15, is shown. The fragment released by BamHI digestion is of appropriate size as predict from Figure 9.1. The structure of the MBP gene-*aprt* connecting sequence was also verified by direct DNA sequencing.

MBP-APRT fusion protein product was extracted and purified from DH5α expressing *aprt* from pMAF15 and pMAFA26, as described in Methods. Assays of activity in the Factor-XA cleavage reaction, revealed APRT activity in the range of 9 nMol AMP per minute per μg protein in the AMP column eluate. No activity was detectable in control cell extracts (from cells containing the original vector pIH902).

insert figure 9.1

Figure 9.2 : Agarose gel electrophoresis of restriction digests of pMAF15. Sizes of fragments excised with SacI/HindIII (S/H) and BamHI are 680 and 560 bp respectively. From the predicted structure of the construct, there should be a single XhoI site in a derivative containing *aprt*, as shown. There is an EcoRV site in the *aprt* sequence, such that the final construct should have two sites for this enzyme. Refer to Figure 9.1 for locations of restriction sites in pIH902 and in the predicted *aprt*-containing plasmid.



Appendix III

Published in : Mutation Research; DNA Repair. 1994; 314(1): 21-26

Strand bias in mutation involving 5-methylcytosine deamination in the human *hprt* gene.

Adonis Skandalis[a]
Barry N. Ford
Barry W. Glickman

University of Victoria, Department of Biology, Centre for Environmental Health, P.O. Box 1700, Victoria, British Columbia, Canada, V8W 2Y2.
[a] current address: University of Utah, Department of Radiology, 50 North Medical Drive, Salt Lake City, Utah, U.S.A., 84132

Correspondence: Barry N. Ford, University of Victoria, Department of Biology, P.O Box 1700, Victoria, B.C., Canada, V8W 2Y2. Tel.:604-656-8146; Fax:604-656-8148. Reprint requests: BWG

Summary

Despite being generally under-represented in the genome, CpG sequences represent a disproportionately high fraction of sites involved in mutational events leading to human genetic disease. Cytosine within CpG dinucleotides is often modified to 5-methylcytosine. Deamination of 5-methylcytosine *in situ* yields a thymine, which being mispaired with guanine, is potentially mutagenic. Previous reports have indicated that most mutations recovered at these sites appear to originate on the non-transcribed strand as C-->T transitions. This trend may however, reflect the lack of detectable mutant phenotypes resulting from this transition at the complementary positions on the transcribed strand. To date, there has not been a good model system in which mutations can be recovered on both strands at the same CpG site. The human *hprt* gene has MeCpG sites contained within arginine codons for which mutations have been recovered on both strands. From an analysis of a database of *hprt* mutations, a statistically significant strand bias is observed in mutations recovered at CpG sites. We describe some models for the bias of mutation distribution observed at MeCpG sites in light of this and previous work.

Introduction

CpG dinucleotides represent potential sites of cytosine methylation in mammalian cells. They are under-represented in the mammalian genome, but are disproportionately involved in mutational events leading to human genetic disease (Rideout *et al.*, 1990; Schorderet and Gartler, 1992). In fact over 35% of the transition mutations involved in haemophilias and *p53* or retinoblastoma gene mediated cancers involve CpG sites (Cooper and Youssoufian, 1988; Green *et al.*, 1990; Koeberl *et al.*, 1990; Hollstein *et al.*, 1991; Rady *et al.*, 1992).

Over-representation of CpG sites in mutational events in mammalian cells is thought to be due to deamination of 5-methylcytosine (MeC) (Ehrlich *et al.*, 1990). The spontaneous deamination of cytosine to uracil is a frequent event even at 37 °C (Lindahl and Nyberg, 1974). Under similar conditions 5-methylcytosine deaminates at least 4 times faster, producing thymine, which is not a foreign base in DNA. The high rate of 5-methylcytosine deamination and the refractory nature of the resulting thymine to uracil-N-glycosylase has been forwarded as the cause of spontaneous mutational hotspots in *E. coli* (Duncan and Miller, 1980).

Repair pathways for U:G and T:G mismatches have been identified. In the case of uracil residues in DNA, repair pathways include the action of uracil-N-glycosylase to remove the uracil, producing an apurinic site (Loeb, 1984). This is followed by cleavage of the phosphodiester backbone by an endonuclease specific for apurinic sites. Resynthesis across the gap occurs with high fidelity by DNA polymerase and ligase. In the case of G:T mismatches, extracts from mammalian cells can specifically recognise T:G mismatches *in vitro*, and restore them efficiently to C-G basepairs, regardless of the origin, or whether they are methylated or not (Brown and Jiricny, 1987; Brown and Jiricny, 1988; Wiebauer and Jiricny, 1990). Though not yet precisely quantified, a large fraction of MeCpG to TpG events are recognised and properly restored. Despite the availability of repair mechanisms, mutations at these sites are still frequently recovered. Indeed, they represent a disproportionate fraction of the total mutational burden. For example, research on the gene encoding protein kinase 1 has demonstrated a high frequency of transition events involving MeCpG sites. These changes occur mainly within arginine codons (CpGpN) and appear to have arisen on the non-transcribed strand (Steinberg and Gorman, 1992). Mutations involving deamination on the complementary (transcribed) strand were not recovered, making an assessment of their relative occurrence impossible. In the absence of phenotypically identifiable mutations on both strands, there is no way to distinguish whether this reflects a lack of mutations on the transcribed strand, or simply that such changes fail to produce a selectable phenotypes. The situation is complicated by the fact that mutation towards the non-transcribed strand is not a universal phenomenon. Thus there is no reason to assume *a priori* that the observation of strand bias of mutation correlated with transcription represents an actual bias in mutation distribution.

We note that MeCpG doublets in codons encoding amino acids other than arginine will, upon undergoing MeC to T events, often yield relatively conservative missense substitutions as well, regardless of the strand on which the change occurs (Table 1). As a result of the preponderance of conservative substitutions which may occur at MeCpG sites, systems to study strand bias of mutational events at CpG dinucleotides, particularly those with a selectable phenotype for events on both strands at the same sites, have been lacking. Arginine residues encoded by the triplet CpGpN should, if methylated, exhibit elevated

mutation frequencies. This codon may provide a good model target for the study of strand bias of MeCpG-related mutation. Deamination on the non-transcribed strand in this codon would convert MeCpGpA to TpGpA, yielding a nonsense codon, whereas on the transcribed strand, the same transition produces relatively conservative (Arg to His or Gln) substitutions, which depending on their location, may or may not yield a mutant phenotype.

In the human hypoxanthine-guanine phosphoribosyl transferase gene (*hprt*), there are four MeCpG sites within arginine codons. Evidence indicates that mutations at two of these MeCpG doublets can be recovered on either the transcribed strand or the non-transcribed strand (see Materials and Methods). Thus these two sites present an opportunity to discern strand bias without the complication of unknown phenotypic expression. Fortunately, an extensive database of spontaneous and induced mutation data in the *hprt* gene has been developed, which now represents a valuable tool for the analysis of mutational phenomena.

We have examined the question of strand bias at arginine codons in this accumulate database, and show a bias in favour of mutation at CpG sites in the non-transcribed strand. This is consistent with earlier observations and more recent evidence that repair can in some cases, occur preferentially on the transcribed strand (Mellon *et al.*, 1987; Leadon and Lawrence, 1991; Carreau and Hunting, 1992; for a review of strand-specific repair, see Hanawalt and Mellon, 1993). A possible model for the preferential repair of 5-methylcytosine deamination on the transcribed strand is discussed.

Materials and Methods

CpG is a symmetrical dinucleotide, in which cytosine is often methylated in mammalian cells to produce 5-methylcytosine (MeC) residues (Bird, 1980; Bird, 1986). MeC can undergo deamination to T in either strand. Deamination in the non-transcribed (NT) strand leads to a C->T transition at the first position of 5'CpG3', whereas deamination of MeC in the transcribed strand leads to a G->A transition at the second position of 5'CpG3'. To investigate the occurrence of strand bias, we examined CpG sites within arginine codons in the human *hprt* gene.

Our primary resource of mutant *hprt* sequences was the HPRT database which is the result of a collaborative effort among researchers involved in the characterisation and analysis

of mutations in the human *hprt* gene. The database is managed by Drs. T. Skopek and N. Cariello, University of North Carolina, Chapel Hill, N.C.

Statistical analysis of the recovery of mutation on the two strands was done using the X^2 test, with one degree of freedom.

Results

The frequency of C->T transitions in the transcribed and non-transcribed strand is summarised in Table 2. Four Arg codons are present in the human *hprt* gene, each containing the CpG dinucleotide. No mutations have been recovered at the CpG dinucleotide at position 10-11 (part of a CGC codon). The CpG at position 142-143, part of a CGT codon, is also relatively "cold", with only 2 GC->AT events reported, both at position 142. The third CpG, part of a CGA codon at position 151-152, has been the site of 13 mutations. At position 151, 9/10 mutations were C->T transitions. At position 152, 1/10 mutations was a G->A which of course reflects a C->T transition in the transcribed strand. The strand bias at this CpG position is statistically highly significant (X^2=6.4, d.f.=1, $p<0.05$). The fourth CpG is at position 508-509, part of a CGA codon. At position 508, 8/9 mutations were C->T whereas at position 509 the only mutation recovered was a G->A. The strand bias at this CpG site is also statistically significant (X^2=5.4, d.f.=1, $p<0.05$). Between the two available sites, 17 of a total of 19 mutants occurred in the NT strand.

Discussion

The data indicate that there is a significant strand bias for mutations arising from the deamination of methylated cytosine in the human *hprt* gene, most mutations (17/19) originating on the non-transcribed strand. This conclusion is based on the collective database of accumulated mutants at the *hprt* locus. These observations raise questions regarding the possible mechanisms responsible for the observed strand bias. Several different, but not necessarily mutually exclusive models, can account for the observation.

One possibility is that the two strands deaminate with different rates. Indeed, it has been reported that in single-stranded DNA, C and MeC are as much as 100-fold more labile than in double-stranded DNA (Lindahl and Nyberg, 1974; Wang *et al.*, 1982; Green *et al.*, 1990). During transcription the transcribed strand is temporarily associated with mRNA and

the transcription complex, whereas the non-transcribed strand may be considered to be essentially single stranded. We have previously suggested a similar model for the strand specificity observed in mutations resulting from the deamination of C in *E. coli* (Fix and Glickman, 1986; Fix and Glickman, 1987). It should be noted however that the existence of a single-stranded form during transcription is probably very short. Thus differential deamination due to transcription seems unlikely to account for the observed level of strand bias.

Alternatively, the observed bias may be accounted for by the preferential repair of G-T mismatches in the transcribed strand. Several reports indicate a preferential repair of lesions located in the transcribed strand, for example following treatment by UV light or other agents (Mellon *et al.*, 1987; Leadon and Lawrence, 1991). A similar preference may exist for G:T mismatches which appear to be repaired by a mechanism analogous to the very short patch (VSP) excision repair system identified in *E. coli* (Lieb, 1987; Zell and Fritz, 1987; Wiebauer and Jiricny, 1990). G:T mispairs *in vitro* are restored to G:C pairs with the release of a single, unmodified T.

The mechanism of activation of the glycosylase involved remains unknown, but for the glycosylase to identify T as a potential error, requires the base to be "tagged" by some factor that differentiates T in a T:G mismatch. A potential candidate protein factor which binds to G:T mismatches and that may serve to "tag" the mispaired T, has been identified, but its full role remains to be elucidated (Jiricny *et al.*, 1988). Following this model, the "tagged" T could block transcription in the transcribed strand and possibly initiate transcription-mediated repair systems. In the non-transcribed strand the "tagged" T would be unlikely to interfere with the transcription complex, and may escape efficient detection. T-tagging by a specific factor has a similar predicted effect to that observed for BPDE adducts on guanine residues, which are also thought to affect transcription, and thus initiate strand-specific repair (Andersson *et al.*, 1992). BPDE mutagenesis of *hprt* also shows a marked strand bias, favouring mutation in the non-transcribed strand.

While tagging T in G:T mismatches could work well during transcription, it would nonetheless create problems during replication. Misincorporation of G opposite T is a frequent polymerase error (Grosse *et al.*, 1983). Consequently, preferential repair of G:T mismatches to

G:C pairs would be mutagenic. It appears likely that the G:T to G:C-specific repair mechanism would be inoperable during and immediately following replication in order to avoid the misrepair of errors from G misincorporation. This would require an alternate repair process to determine the correct strand to repair in the case of G:T mispairing events during replication (Holmes *et al.*, 1990).

Several implications follow from the above model. If repair of G:T mismatches in mammalian cells is indeed linked to transcription, no strand bias would be expected for MeC -> T mutation (*i.*e. at CpG sites) in inactive genes. Cell lines from patients with Cockayne's syndrome for example, already exist. These cell lines have impaired repair of transcribed genes (Venema *et al.*, 1991), but apparently otherwise normal excision repair systems. It is intriguing to speculate that the defect in Cockayne's syndrome may not be a direct repair function at all, but a transcription-specific signalling mechanism. This is parallels our prediction that a mutant in which the G-T mismatch binding protein, which may signal G-T repair during transcription, could be isolated. Such a mutant would exhibit an elevated rate of MeC ->T events different and an altered bias towards events in the non-transcribed strand. In the absence of extensive data at sites for which mutations can be recovered from both strands based on phenotypic selection, any model of the events controlling strand preference in damage or repair must remain speculative. These observations however, indicate the need to understand DNA damage and repair in their proper *in vivo* context.

Acknowledgements

The authors wish to thank Drs. Skopek and Carriello for access to the *hprt* database. This work was supported by an IRC award to BWG from the Natural Science and Engineering Research Council.

References

Andersson, B., Falt, S., and Lambert, B. (1992) Strand specificity for mutations induced by (+)-anti BPDE in the *hprt* gene in human T-lymphocytes, Mut. Res. 269, 129-140.

Bird, A. (1980) DNA methylation and the frequency of CpG in animal DNA, Nuc. Acids. Res.,

8, 1499-1504.

Bird, A. (1986) CpG-rich islands and the function of DNA methylation, Nature, 321, 209-213.

Brown, T, And Jiricny, J. (1987) A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine, Cell, 50, 945-950.

Brown, T., and Jiricny, J. (1988) Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells, Cell, 54, 705-711.

Carreau, M., and Hunting, D. (1992) Transcription-dependent and independent DNA excision repair pathways in human cells, Mut. Res., 274, 57-64.

Cooper, D., and Youssoufian, H. (1988) The CpG dinucleotide and human genetic disease, Hum. Genet., 78, 151-155.

Duncan, B. and Miller, J. (1980) Mutagenic deamination of cytosine residues in DNA, Nature, 287, 560-561.

Ehrlich, M., Zhang, X., and Inamdar, N. (1990) Spontaneous deamination of cytosine and 5-methylcytosine residues in DNA and replacement of 5-methylcytosine with cytosine residues, Mut. Res., 238, 277-286.

Fix, D. and Glickman, B. (1986) Differential enhancement of spontaneous transition mutations in the *lacI* gene of an *ung* strain of *Escherichia coli*, Mut. Res. 175, 41-45.

Fix, D., and Glickman, B. (1987) Asymmetric cytosine deamination revealed by spontaneous mutational specificity in an *ung*⁻ strain of *E. coli*, Mol. Gen. Genet., 209, 78-82.

Green, P., Montandon, J., Bentley, D., Ljung, R., Nilsson, I., and Giannelli, F. (1990) The incidence and distribution of CpG -> TpG transitions in the coagulation factor IX gene. A fresh look at CpG mutational hotspots, Nuc. Acids Res., 18, 3227-3231.

Grosse, F., Krauss, G., Knill-Jones, J., and Fersht, A. (1983) Accuracy of DNA polymerase a in copying natural DNA, EMBO J., 2, 1515-1519.

Hanawalt,P., and Mellon, I. (1993) Stranded in an active gene. Current Biol., 3, 67-69.

Hollstein, M., Sidransky, D., Vogelstein, B., and Harris, C. (1991) *p53* mutations in human cancers, Science, 253,49-53.

Holmes, J., Clark,S., and Modrich, P. (1990) Strand-specific mismatch correction in nuclear extracts of human and *Drosophila melanogaster* cell lines, Proc. Nat. Acad. Sci. 87,

5837-5841.

Jiricny, J., Hughes, M., Corman, N., and Rudkin, B. (1988) A human 200-kDa protein binds selectively to DNA fragments containing G:T mismatches, Proc. Nat. Acad. Sci. 85, 8860-8864.

Koeberl, D., Bottema, C., Sarkar, G., Ketterling, R., Chen, S., and Sommer, S. (1990) Recurrent nonsense mutations at arginine residues cause severe haemophilia in unrelated haemophiliacs, Hum. Genet., 84, 387-390.

Leadon, S., and Lawrence, D. (1991) Preferential repair of DNA damage on the transcribed strand of the human metallothionen genes requires RNA polymerase II, Mut. Res., 255, 67-78.

Lieb, M. (1987) Bacterial genes *mutL*, *mutS*, and *dcm* participate in repair of mismatches at methylcytosine sites, J. Bact., 169, 5241-5246.5

Lindahl, T., and Nyberg, B. (1974) Heat-induced deamination of cytosine residues in deoxy-ribonucleic acid, Biochem., 13, 3405-3410.

Loeb, L. (1985) Apurinic sites as mutagenic intermediates, Cell, 40, 483-484.

Mellon, I., Pivak, G., and Hanawalt, P. (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene, Cell, 51, 241-249

Rady, P., Scinicariello, F., Wagner, F., and King, S. (1992) *p53* mutations in basal cell carcinomas, Cancer Res., 52, 3804-3806.

Rideout, W., Coetzee, G., Olumi, A., and Jones, P. (1990) 5-methylcytosine as an endogenous mutagen in the human LDL receptor and *p53* genes, Science, 249, 1288-1290.

Schorderet, D., Gartler, S. (1992) Analysis of CpG suppression in methylated and nonmethylated species, Proc. Nat. Acad. Sci., 89, 957-961.

Steinberg, R., and Gorman, K. (1992) Linked spontaneous CG -> TA mutations at CpG sites in the gene for protein kinase regulatory subunit, Mol. Cell. Biol., 12, 767-772.

Venema, J., Mullenders, H., Natarajan, A., van Zeeland, A., and Mayne, L. (1991) The genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA damage in trancriptionally active DNA, Proc. Nat. Acad. Sci., 87, 4707-4711.

Wang, R., Kuo, K., Gerhke, C., and Huang, L.-H., and Erhlich, M. (1982) Heat and alkali-induced deamination of 5-methylcytosine and cytosine residues in DNA, Biochim. Biophys. Acta, 697, 371-377.

Wiebauer, K., and Jiricny, J. (1990) Mismatch-specific thymine DNA glycosylase and DNA polymerase b mediate the correction of G-T mispairs in nuclear extracts from human cells, Proc. Nat. Acad. Sci., 87, 5842-5845.

Zell, R., and Fritz, H. (1987) DNA-mismatch repair in *Escherichia coli* counteracting the hydrolytic deamination of 5-methyl-cytosine residues, EMBO J., 6, 1809-1815.

TABLE 1 : C->T TRANSITIONS AT CpG DINUCLEOTIDES

Nontranscribed Strand Mutations (MeCpG -> TpG) :

Codons containing MeCpG doublets

Codon	**Residue**	**Transition**	**Products**
CGN	Arg	TGN N=A	stop
		N=C,T	Cys
		N=G	Trp
ACG	Thr	ATG	Met
CCG	Pro	CTG	Leu
GCG	Gly	GTG	Val
TCG	Ser	TTG	Leu

Codons Spanning MeCpG doublets NpNpCpGpNpN

Codon	**Residue**	**Transition**	**Products**
AAC	Asn	AAT	Asn
ACC	Thr	ACT	Thr
AGC	Ser	AGT	Ser
ATC	Gly	ATT	Ile
CAC	His	CAT	His
CCC	Pro	CCT	Pro
CGC	Arg	CGT	Arg
CTC	Leu	CTT	Leu
GAC	Asp	GAT	Asp
GCC	Ala	GCT	Ala
GGC	Gly	GGT	Gly
GTC	Val	GTT	Val
TAC	Tyr	TAT	Tyr
TCC	Ser	TCT	Ser
TGC	Cys	TGT	Cys
TTC	Phe	TTT	Phe

TABLE 1 cont.

Transcribed Strand Mutations :

(MeCpG ->MeCpA) Codons Containing MeCpG doublets

Codon	**Residue**	**Transition**	**Products**
CGN	Arg	CAN N=A,G	Gln
		N=C,T	His
ACG	Thr	ACA	Thr
CCG	Pro	CCA	Pro
GCG	Gly	GCA	Ala
TCG	Ser	TCA	Leu

Codons Spanning MeCpG doublets NpNpCpGpNpN

Codon	**Residue**		**Transition**	**Products**
GAN	N=A,G	Glu	AAN	Lys
	N=C,T	Asp		Asn
GCN		Ala	ACN	Thr
GGN	N=A,G	Gly	AGN	Arg
	N=C,T	Gly	AGN	Ser
GTN	N=A,C,T	Val	ATN	Ile
	N=G	Val	ATG	Met

TABLE 2 : FREQUENCY OF C->T TRANSITIONS IN Arg CODONS CONTAINING CpG DINUCLEOTIDES

POSITION	NON-TRANSCRIBED STRAND	TRANSCRIBED STRAND
10-11	0	0
142-143	2	0
151-152	9	1
508-509	8	1

References

Adair, G., J. Carver and D. Wandres (1980) Mutagenicity testing in mammalian cells. I. Derivation of a Chinese hamster ovary cell line hetrozygous for the adenine phosphoribosyltransferase and thymidine kinase loci, Mutat. Res. 72, 187-205.

Aghajari, N., K. Jensen and M. Gajhede (1994) Crystallization and preliminary X-ray diffraction studies on the apo form of orotate phosphoribosyl transferase from *Escherichia coli*, J. Mol. Biol. 241, 292-294.

Ali,L., and Sloan, D. (1982) Studies of the kinetic mechanism of hypoxanthine-guanine phosphoribosyl transferase from yeast. J. Biol. Chem. 257,1149-1155.

Ali, L.Z. and D.L. Sloan (1983) Studies of the mechanism of activation of the yeast hypoxanthine-guanine phosphoribosyl transferase catalyzed reactions by divalent metal ions, Biochemistry 22, 3419-3424.

Al-Safi, S.A. and J.L. Maddocks (1984) Azathioprine and 6-mercaptopurine (6-MP) suppress the human mixed lymphocyte reaction (MLR) by different mechanisms, Br. J. Clin. Pharmacol. 17, 417-422.

Al-Shawi, D. Parsonage and A. Senior (1988) Directed mutagenesis of the strongly conserved aspartate-242 in the β-subunit of *Escherichia coli* proton-ATPase, J. Biol. Chem. 260, 19633-19639.

Alexiou, M. and H.J. Leese (1994) Enzymes of purine salvage and catabolism in the mouse preimplantation, J. Reprod. Fert. 101, 151-158.

Amann, A., Brosius,J., Ptashne, M. (1983) Vectors bearing a hybrid *trp-lac* promoter useful for regulated expression of cloned genes in *Escherichia coli* Gene, 25,167-178.

Antosiewicz, J., S.T. Wlodek and J.A. McCammon (1996) Acetylcholinesterase: role of the enzyme's charge distribution in steering charged ligands toward the active site, Biopolymers, 39, 85-94.

Arnold, W. and W. Kelley (1978) Adenine phosphoribosyltransferase, Meth. Enz. 51, 568-574.

Argos, P. (1988) An investigation of protein subunit and domain interfaces, Protein Eng. 2, 101-113.

Argos, P., M. Hanei, J.M. Wilson and W.N. Kelley (1983) A possible nucleotide-binding domain in the tertiary fold of phosphoribosyl transferases, J. Biol. Chem. 258, 6450-6457.

Arnold, W. and W. Kelley (1978) Adenine phosphoribosyl transferase, Meth. Enz. 51, 568-574.

Arnold, W., R. Lamb and W. Kelley (1974) Human hypoxanthine-guanine phosphoribosyl transferase (HGPRT): purification and properties, Adv. Exp. Med. Biol. 441A, 5-14.

Asahi H., T. Kanazawa, Y. Kajihara, K. Takahashi, and T. Takahashi (1996) Hypoxanthine: a low molecular weight factor essential for growth of erythrocytic *Plasmodium falciparum* in a serum-free medium, Parasitology 113, 19-23

Ashton, R.W., R.S. Strauss, S.H. Chung and D.L. Sloan (1989) Orotate phosphoribosyl transferase from yeast: Studies of the structure of the pyrimidine substrate binding site, Arch. Biochem. Biophys, 272, 421-432.

Attwood, T., E. Eliopoulos and J. Findlay (1991) Multiple sequence alignment of protein families showing low sequence homology: a methodological approach using database pattern-matching discriminators for G-protein-linked receptors, Gene, 98, 153-159.

Avila J., and M. Casanova (1982) Comparative effects of 4-aminopyrazolopyrimidine, its 2'-deoxyriboside derivative, and allopurinol on in vitro growth of American Leishmania species, Antimicrob. Agents Chemother. 1982 22, 380-385.

Barlow, D.J. and J.M. Thornton (1988) Helix geometry in proteins, J. Mol. Biol. 201, 601-619.

Bartsch, H. and E. Hietanen (1996) The role of individual susceptibility in cancer burden related to environmental exposure, Environ. Health Perspect. 104 Suppl 3:569-77, 569-577.

Batt, C., P. Oren, J. Webb, and P. Flicke (1990)An improved method for oligonucleotide-mediated site-directed mutagenesis, Biotechniques, 9,554

Bell, R.M. and D.E.J. Koshland (1970) A phosphoribosyl-enzyme covalent intermediate in the first step of histidine biosynthesis, Bioch. Biophys. Res. Comm. 38, 539-545.

Bennet, L., M. Vail, S. Chumley and J. Montgomery (1966) Activity of adenosine analogs against a cell cultur line resistant to 2-fluoroadenine, Biochem. Pharmacol. 15, 1719-1728.

Bennett, M.J., S. Choe and D. Eisenberg (1994) Domain swapping: entangling alliances between proteins, Proc. Natl. Acad. Sci. U. S. A. 91, 3127-3131.

Berger, S.A. and P.R. Evans (1990) Active-site mutants altering the cooperativity of *E. coliW* phosphofructokinase, Nature, 343, 575-576.

Berlin, R. (1969) Adenylate pyrophosphorylase: purification, reaction sequence, and inhibition by sodium ion, Arch. Biochem. Biophys. 134, 120-129.

Bhalla, K., R. Nayak, A. Deitch and S. Grant (1984) Synergistic interaction between acividin (AT-125) and 6-thioguanine in the murine leukemia L1210, Biochem. Pharmacol. 33, 247-240.

Bhatia, M.B. and C. Grubmeyer (1993) The role of divalent magnesium in activating the reaction catalyzed by orotate phosphoribosyl transferase, Arch.Biochem. Biophys. 303, 321-325.

Bockamp, O., R. Blasco and E. Vinuela (1990) *Escherichia coli* thymidine kinase: nucleotide sequence of the gene and relationships to other thymidine kinases, Gene, 101, 9-14.

Bohm, G. (1997) New approaches in molecular structure prediction, Biophys. Chem. 1996. Mar. 7. 59, 1-32.

Botstein, D. and D. Shortle (1985) Strategies and applications of *in vitro* mutagenesis, Science, 229, 1193-1201.

Bowie, J. and R. Sauer (1989) Identifying determinants of folding and activity for a protein of unknown structure, Proc. Natl. Acad. Sci.U.S.A. 86, 2152-2156.

Boyer, J., A. Umar, J. Risinger, J. Lipford, M. Kane, S. Yin, J. Barret, R. Kolodner, and T. Kunkel (1995) Microsatellite instability, mismatch repair deficiency, and genetic defects in human cancer cell lines, Cancer Res. 55, 6063-2070.

Bradford, M.M. (1976) A rapid a sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye binding Anal. Biochem. 72, 248-254.

Branch, P., R. Hampson and P. Karran (1995) DNA mismatch binding defects, DNA damage tolerance, and mutator phenotypes in human colorectal carcinoma cell lines, Cancer Res. 55, 2304-2309.

Busetta, B. (1988) The use of folding patterns in the search of protein structural similarities; a three-dimensional model of phosphoribosyl transferases, Biochim. Biophys. Acta, 957, 21-33.

Carter, P. and J. Wells (1987) Engineering enzyme specificity by "substrate-assisted catalysis", Science, 237, 394-397.

Carver, J., G. Adair and D. Wandres (1980) Mutagenicity testing in mammalian cells. II.

Validation of multiple drug-resistance markers having practical applications for screening potential mutagens, Mutat. Res. 72, 207-230.

Chappell, D.J. and M. Slaytor (1992) Purine salvage pathways in the Australian termite, *Nasutitermes walkeri*, Insect Biochem. Mol. Biol., 22, 175-179.

Chelsky, D. and S.M. Parsons (1975) Stereochemical course of the adenine triphosphate phosphoribosyl transferase reaction in histidine biosynthesis, J. Biol. Chem. 250, 5669-5673.

Chen, J., A. Sahota, T. Laxdal, M. Scrine, S. Bowman, C. Cui and P. Stambrook,J; (1991) Identification of a single missense mutation in the adenine phosphoribosyl transferase (APRT) gene from five Icelandic patients and a single British patient, American Journal Of Human Genetics, 49, 1306-1311.

Cheng, Y.E., D.Y. Kwoh, T.J. Kwoh, B.C. Soltveldt and D. Zipser (1981) Stabilization of a degradable protein by its overexpression in *Escherichia coli*, Gene, 14, 121-130.

Cheung, D.T. and M.E. Nimni (1982) Mechanism of crosslinking of proteins by glutaraldehyde II. Reaction with monomeric and polymeric collagen, Connect. Tissue Res. 10, 201-216.

Choi, H, and J. Stocker (1992) The effect of MTA analogs on purine metabolism, Korean Biochem. J., 25,39-47.

Chothia, C., M. Levitt and D. Richardson (1977) Structure of proteins: packing of α-helices and pleated sheets, Proc. Natl. Acad. Sci. U.S.A. 74, 4130-4134.

Chou, P.Y. and G.D. Fasman (1978) Prediction of the secondary structure of proteins from their amino acid sequence, Adv. Enzymol. Relat. Areas. Mol. Biol. 47, 45-148.

Chou, P.Y. and G.D. Fasman (1978) Empirical predictions of protein conformation, Annu. Rev. Biochem. 47, 251-276.

Cleton Jansen, A.M., H.W. Moerland, D.F. Callen, N.A. Doggett, P. Devilee and C.J. Cornelisse (1995) Mapping of the breast basic conserved gene (D16S444E) to human chromosome band 16q24.3, Cytogenet. Cell Genet. 68, 49-51.

Cohen, C. and D. Perry (1990) α-helical coiled coils and bundles; how to design an α-helical protein, Prot. Struct. Func. Genet. 7, 1-15.

Cooper, G.E., D.L. Dimartino and M.S. Turker (1991) Molecular analysis of APRT deficiency in mouse P19 teratocarcinoma stem, Somat. Cell Mol. Genet., 17, 105-116.

Cooper, G.E., N.H. Khattar, P.L. Bishop and M.S. Turker (1992) At least two distinct epigenetic mechanisms are correlated with high-frequency "switching" for APRT phenotypic expression in mouse embryonal carcinoma stem cells, Somat. Cell Mol. Genet. 18, 215-225.

Cooper D., and H. Youssoufian (1988) The CpG dinucleotide and human genetic disease, Hum. Genet. 78, 151-155.

Crick, F.H.C. (1953) The packing of α-helices: simple coiled-coils, Acta. Cryst. 6, 689-697.

Dahnke, T., R.T. Jiang and M.D. Tsai (1991) Mechanism of adenylate kinase .12. prediction and demonstration of enhancement of phosphorus stereospecificity by site-directed mutagenesis, J. Am. Chem. Soc. 113, 9388-9389.

Dare, E., M. Roelofs, S. Simi, S. Pavanello, F. Pilli and V. Bianchi (1992) Metabolic conssequences of adenine-phosphoribosyl transferase deficiency in V79 hamster fibroblasts, Experimental Cell Research, 203, 336-343.

Davidson B., M. Pashmforoush, W. Kelley, and T. Palella (1988) Genetic basis of hypoxanthine guanine phosphoribosyltransferase deficiency in a patient with the Lesch-Nyhan syndrome (HPRTFlint), Gene 63, 331-336

Dayhoff, M.O., R.M. Schwarz and B.C. Orcutt (1978) A model of evolutionary change in proteins, Atlas of Protein Sequence and Structure, 5; Supp. 3, 345-352.

Dayton, J.S., L.A. Turka, C.B. Thompson and B.S. Mitchell (1992) Comparison of the effects of mizoribine with those of azathioprine, 6- mercaptopurine, and mycophenolic acid on T lymphocyte proliferation and purine ribonucleotide metabolism, Mol. Pharmacol. 41, 671-676.

de Boer, H.A. and H.M. Shepard (1983) Strategies for optimizing foreign gene expression in *Escherichia coli*, in: A.M. Kimon (Ed.), Genes: Structure and Function, John Wiley and Sons, New York, pp. 205-248.

de Boer, J.G. and B.W. Glickman (1989) Sequence specificity of mutation induced by the anti-tumor drug cisplatin in the CHO *aprt* gene, Carcinogenesis, 10, 1363-1367.

de Boer, J.G. and B.W. Glickman (1991) Mutational analysis of the structure and function of the adenine phosphoribosyl transferase enzyme, J. Mol. Biol. 221, 163-174.

Dean, B.M., R.W.E. Watts and W.J. Westwick (1968) Human erythrocyte AMP:pyrophosphate phosphoribosyl transferase (EC 2.4.2.7), FEBS Lett. 1, 179-182.

Desideri, A., M. Falconi, F. Polticelli, M. Bolognesi, K. Djinovic and G. Rotilio (1992) Evolutionary conservativeness of electric field in the Cu,Zn superoxide dismutase active site. Evidence for co-ordinated mutation of charged amino acid residues, J. Mol. Biol. 223, 337-342.

Dodin, G., D. Lalart and J. Dubois (1982) Role of magnesium cations in the yeast orotate phosphoribosyl transferase catalyzed reaction. Mecahnism of the inhibition by Cu^{++} and Ni^{++} ions, J. Inorg. Biochem. 16, 210-213.

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

Drobetsky, E. A. Grosovsky, and B. Glickman (1987) The specificity of UV-induced mutations at an endogenous locus in mammalian cells, Proc. Natl. Acad. Sci. U.S.A. 84-9103-9107.

Drobetsky, E. A., Turcotte J., and A. Chateauneuf (1995) A role for ultraviolet A in solar mutagenesis, Proc. Nat. Acad. Sci. 92:2350-2354.

Edwards, S.L., J. Kraut, N. Xuong, V. Ashford, T.P. Halloran and S.E. Mills (1988) Crystallization and purification of the enzyme anthranilate phosphoribosyl transferase, J. Mol. Biol. 203, 523-524.

Efimov, A. (1993) Standard structures in proteins, Prog. Biophys. Molec. Biol. 60, 201-239.

Emmerson B, R. Gordon, and L. Thompson (1975) Adenine phosphoribosyltransferase deficiency: its inheritance and occurrence in a female with gout and renal disease. Aust N Z J Med, 5, 440-446

Engle, S.J., M.G. Stockelman, J. Chen, G. Boivin, M.N. Yum, P.M. Davies, M.Y. Ying, A. Sahota, H.A. Simmonds, P.J. Stambrook and J.A. Tischfield (1996) Adenine phosphoribosyl transferase-deficient mice develop 2,8-dihydroxyadenine nephrolithiasis, Proc. Natl. Acad. Sci. U. S. A. 93, 5307-5312.

Fersht, A. (1985) Enzyme Structure and Function, W.H. Freeman and Co. New York, pp. 25-16.

Fieldhouse, D., F. Yazdani and G.B. Golding (1997) Substitution rate variation in closely related rodent species, Heredity, 78, 21-31.

Fox I., J. Meade, and W. Kelley (1973) Adenine phosphoribosyl transferase deficiency in man. Report of a second family, Am. J. Med. 55, 614-620

Fox, I.H. (1976) Purine enzyme abnormalities: A four year experience, Adv. Exp. Med. Biol. 76A, 265-269.

Fox, M., J. Boyle and A. Kinsella (1991) Nucleoside salvage and resistance to antimetabolite anticancer agents, Br. J. Cancer, 64, 428-436.

Fratini A, R. Simmers, D. Callen, V. Hyland, J. Tischfield, P. Stambrook, and G. Sutherland (1986) A new location for the human adenine phosphoribosyl transferase gene (APRT) distal to the haptoglobin (HP) and fra(16)(q23)(FRA16D) loci, Cytogenet. Cell. Genet. 43, 10-13.

Frishman, D. and P. Argos (1997) The future of protein secondary structure prediction accuracy, Fold. Des. 1997. 2, 159-162.

Frost J., J. Bender, J. Kadonaga, and J. Knowles (1984) Dehydroquinate synthase from *Escherichia coli*: purification, cloning, and construction of overproducers of the enzyme, Biochem. 23, 4470-4475

Fry, D.C., S.A. Kuby and A.S. Mildvan (1986) ATP-binding site of adenylate kinase: Mechanistic implications of its homology with *ras*-encoded p21, F_1ATPae, and other nucleotide-binding proteins, Proc. Natl. Acad. Sci. U.S.A. 83, 907-911.

Fujimori, S., I. Akaoka, K. Sakamoto, H. Yamanaka, K. Nishioka and N. Kamatani (1985) Common characteristics of mutant adenine phosphoribisyltransferases from four separate Japanese families with 2,8-dihydroxyadenine urolithiasis associated with partial enzyme deficiencies, Hum. Genet. 71, 171-176.

Gadd, R.E.A. and J.F. Henderson (1969) Inhibition of adenine phosphoribosyl transferase by pyrophosphate, Biochim. Biophys. Acta, 191, 735-737.

Gadd, R. and J. Henderson (1970) Studiews of the binding of phosphoribosyl pyrophosphate to adenine phosphoribosyl transferase, J. Biol. Chem. 245, 2979-2984.

Geourjon, C. and G. Deleage (1997) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments, Comput. Appl. Biosci. 11, 681-684.

Gorbalenya, A.E. and E.V. Koonin (1989) Viral proteins containg the purine NTP-binding sequence pattern, Nuc. Acids Res. 17, 8413-8417.

Graf, L.H., Jr., J.A. McRoberts, T.M. Harrison and D.W. Martin,Jr. (1976) Increased PRPP synthetase activity in cultured rat hepatoma cells containing mutations in the hypoxanthine-guanine phosphoribosyl transferase gene, J. Cell Physiol. 88, 331-342.

Green P., A. Montandon, D. Bentley, R. Ljung, I. Nilsson, and F. Giannelli (1990) The incidence and distribution of CpG→TpG transitions in the coagulation factor IX gene. A fresh look at CpG mutational hotspots, Nuc. Acids Res. 18, 3227-3231

Green, S.M., A.G. Gittis, A.K. Meeker and E.E. Lattman (1995) One-step evolution of a dimer from a monomeric protein, Nat. Struct. Biol. 2, 746-751.

Grosovsky, A.J., E.A. Drobetsky and B.W. Glickman (1986) Southern analysis of genomic alterations in gamma-ray induced APRT- hamster cell mutants, Genetics, 113, 405-415.

Groth, D. and L. Young (1971) On the formation of an intermediate in the adenine phosphoribosyl transferase reaction, Bioch. Biophys. Res. Comm. 43, 82-87.

Gupta, P.K., A. Sahota, S.A. Boyadjiev, S. Bye, C. Shao, J.P. O'Neill, T.C. Hunter, R.J. Albertini, P.J. Stambrook and J.A. Tischfield (1997) High frequency in vivo loss of heterozygosity is primarily a consequence of mitotic recombination, Cancer Res. 57, 1188-1193.

Hager, A.J., J.D. Pollard and J.W. Szostak (1996) Ribozymes: aiming at RNA replication and protein synthesis, Chem. Biol. 3, 717-725.

Hakoda, M., H. Yamanaka and N. Kamatani (1991) Diagnosis of heterozygous states for adenine phosphoribosyltransferase, American Journal Of Human Genetics, 48, 552-562.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids, J. Mol. Biol. 166, 557-580.

Harwood, J., A. Tachibana and M. Meuth (1991) Multiple dispersed spontaneous mutations: A novel pathway of mutation in a malignant human cell line, Mol. Cell. Biol., 11, 3163-3170.

Hashimoto, H., M. Kubota, T. Shimizu, T. Takimoto, T. Kitoh, Y. Akiyama and H. Mikawa (1990) Biochemical basis of the prevention of 6-thiopurine toxicity, Leukemia Res. 14, 1061-1066.

Hashimoto, T., Y. Nakamura, F. Nakamura, T. Shirakura, J. Adachi, N. Goto, K. Okamoto and M. Hasegawa (1997) Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*, Mol. Biol. Evol. 1994. Jan. 11, 65-71.

Henderson, J.F., H.R. Miller, W.N. Kelley, F.M. Rosenbloom and J.E. Seegmiller (1968) Kinetic studies of mutant human erythrocyte adenine phosphoribosyl transferases, Can. J.

Biochem. 46, 703-706.

Henriksen, A., N. Aghajari, K.F. Jensen and M. Gajhede (1996) A flexible loop at the dimer interface is a part of the active site of the, Biochem. 35, 3803-3809.
Heringa, J. and W.R. Taylor (1997) Three-dimensional domain duplication, swapping and stealing, Curr. Opin. Struct. Biol. 7, 416-421.

Hicks, M.R., D.V. Holberton, C. Kowalczyk and D.N. Woolfson (1997) Coiled-coil assembly by peptides with non-heptad sequence motifs, Fold. Des. 2, 149-158.

Hidaka, Y., T.D. Pallela, T. O'Toole, S.A. Tarle and W.N. Kelley (1987) Human adenine phosphoribosyltransferase. Identification of allelic mutations at the nucleotide level as a cause of complete deficiency of the enzyme, J. Clin. Invest. 80, 1409-1415.

Hidaka, Y., S.A. Tarle, S. Fujimori, N. Kamatani, W.N. Kelley and T.D. Palella (1988) Human adenine phosphoribosyltransferase deficiency Demonstration of a single mutant allele common to the Japanese, J. Clin. Invest. 81, 945-950.

Higgins, D.G. (1997) CLUSTAL V: multiple alignment of DNA and protein sequences, Methods Mol. Biol. 25, 307-318.

Hitchings, G., G. Elion, E. Falco, P. Russel, M. Sherwood and H. Vanderwerff (1950) Antagonists of nucleic acid derivatives. I. The *Lactobacillus casei* model, J. Biol. Chem. 183, 1-9.

Hitchings, G. and G. Elion (1961) Purine analogues, in: R. Hochster and J. Quastel (Eds.), Metabolic Inhibitors, Academic Press, New York, pp. 215-237.

Hochstadt-Ozer, J. and E. Stadtman (1971) The regulation of purine utilization in bacteria: I. Purification of adenine phosphoribosyltransferase from *Escherichia coli* K12 and control of activity by nucleotides, J. Biol. Chem. 246, 5294-5303.

Holden, J. and W. Kelley (1978) Human hypoxanthine-guanine phosphoribosyl transferase: Evidence for a tetrameric structure, J. Biol. Chem. 253, 4459-4463.

Holden, J., Meredith, G., and W. Kelley (1979) Human adenine phosphoribosyl transferase: Affinity purification, subunit structure, amino acid composition, and peptide mapping, J. Biol. Chem. 254, 6951-6955.

Hori, M. and J.F. Henderson (1966a) Purification and properties of adenylate pyrophosphorylase from Ehrlich ascites tumor cells, J. Biol. Chem. 241, 1406-1411.

Hori, M. and J.F. Henderson (1966b) Kinetic studies of adenine phosphoribosyl

transferase, J. Biol. Chem. 241, 3404-3408.

Hori H.,and S. Osawa (1979) Evolutionary change in 5S RNA secondary structure and a phylogenic tree of 54 5S RNA species, Proc. Natl. Acad. Sci. U.S.A. 76, 381-385
Huang, X. (1997) On global sequence alignment, Comput. Appl. Biosci. 10, 227-235.

Hughes, S.H., G.M. Wahl and M.R. Cappechi (1975) Purification and characterization of mouse hypoxanthine-guanine phosphoribosyl transferase, J. Biol. Chem. 250, 120-126.

Inoue, H., H. Nojima and H. Okayama (1990) High efficiency transformation of *Escherichia coli* with plasmids, Gene, 96, 23-28.

Johnson, L.A., R.B. Gordon and B.T. Emmerson (1977) Adenine phosphoribosyl transferase: A simple spectrophotometric assay and the incidence of mutation in the normal population, Bioch. Genet. 15, 265-272.

Johnson, G., L. Eisenberg and B. Migeon (1979) Human and mouse hypoxanthine guanine phosphoribosyl transferases: dimers and tetramers, Science, 203, 174-176.

Kalle, G. and J. Gots (1963) Genetic alteration of adenylic pyrophosphorylase in *Salmonella*, Science, 14268, 680-681.

Kamatani, N., S. Kuroshima, C. Terai, Y. Hidaka, T.D. Palella and K. Nishioka (1989) Detection of an amino acid substitution in the mutant enzyme for a special type of adenine phosphoribosyltransferase (APRT) deficiencey by sequence-specific protein cleavage, Am. J. Hum. Genet. 45, 325-331.

Kambayashi, T. Nakanashi, K, Suzuki, K. Fujita, A. Tajima, and k. Kawabe (1994) Two siblings with 2,8-dihydroxyadenine urolithiasis, Acta Urol. Jpn. 40, 1097-1101.

Kammerer, R.A. (1997) Alpha-helical coiled-coil oligomerization domains in extracellular proteins, Matrix Biol. 15, 555-68.

Katz, A. and M. Simon (1995) A segment of the C-terminal half of the G-protein beta 1 subunit specifies its interaction with the gamma 1 subunit, Proc. Natl. Acad. Sci. U.S.A. 92,1998-2002

Kelley W., R. Levy, F. Rosenbloom, J. Henderson, and J. Seegmiller (1968) Adenine phosphoribosyl transferase deficiency: a previously undescribed genetic defect in man, J, Clin, Invest. 47, 2281-2289

Kenimer, J.G., L.G. Young and D.P. Groth (1975) Purification and properties of rat liver adenine phosphoribosyltransferase, Biochim. Biophys. Acta, 384, 87-101.

Khattar, N.H., G.E. Cooper, D.L. Dimartino, P.L. Bishop and M.S. Turker (1992) Molecular and biochemical elucidation of a cellular phenotype characterized by adenine analogue resistance in the presence of high levels of adenine phosphoribosyl transferase activity, Biochem. Genet. 30, 635-648.

Khattar, N.H., C.D. Jennings, K.A. Walker and M.S. Turker (1997) Isolation and characterization of mutations in the mouse APRT gene that encode functional enzymes with resistance to toxic adenine analogs, Adv. Exp. Med. Biol. 370, 665-670.

Khattar, N.H., and M.S. Turker (1997) A role for certain mouse Aprt sequences in resistance to toxic adenine analogs, Somat. Cell Mol. Genet. 23, 51-61.

King, A. and D.W. Melton (1987) Characterization of cDNA clones for hypoxanthine-guanine phosphoribosyl transferase from the human malarial parasite, *Plasmodium falciparum*: comparison to the mammalian gene and protein, Nucl. Acid. Res. 15, 10469-10481.

King, A., E. Gillespie, P. McKenna, and Y. Barnett (1994) An investigation of mutation as a function of age in humans, Mut. Res. 316, 79-90.

Kleeman, J.E. and S.M. Parsons (1976) Reverse direction substrate kinetics and inhibition studies on the first enzyme of histidine biosynthesis, adenosine triphosphate phosphoribosyl transferase, Arch. Biochem. Biophys, 175, 687-693.

Kocharayan, S., V. Livshits and V. Sokhodolets (1975) Genetic study of *Escherichia coli* K-12 mutants resistant to 2, 6-diaminopurine, Sov. Genet. 11, 1417-1425.

Kocharyan, S., T. Chukanova and V. Sukhodolets (1978) Mutations of resistance to 2,6-diaminopurine and 6-methylpurine affecting adenine phosphoribosyl transferase in *Escherichia coli* K12, Genetica, 13, 1226-1310.

Kosinski, M.J. and J.E. Bailey (1992) Structural characteristics of an abnormal protein influencing its proteolytic susceptibility, J. Biotechnol. 23, 211-223.

Koyama, H. and H. Kodama (1982) Adenine phosphoribosyltransferase deficiency in cultured mouse mammary tumor FM3A cells resistant to 4-carbamoylimidazolium 5-olate, Canc. Res. 42, 4210-4214.

Krenitsky, T.A., S.M. Neil, G.B. Elion and G.H. Hitchings (1969a) Adenine phosphoribosyl transferase from monkey liver. Specificity and properties, J. Biol. Chem. 244, 4779-4784.

Krenitsky, T.A., R. Papaioannou and G.B. Elion (1969b) Human hypoxanthine

phosphoribosyl transferase. I. Purification, properties, and specificity, J. Biol. Chem. 244, 1263-1270.

Kuma, K. and T. Miyata (1997) Mammalian phylogeny inferred from multiple protein data, Jpn. J. Genet. 69, 555-566.

Kunkel, T.A. and L.A. Loeb (1979) On the fidelity of DNA replication. Effect of divalent metal ion activators and deoxyrionucleoside triphosphate pools on in vitro mutagenesis, J. Biol. Chem. 254, 5718-5725.

Laemmli, U. and M. Favre (1973) Maturation of the head of bacteriophage T4. I. DNA packaging events, J. Mol. Biol. 80, 575-599.

Laemmli, U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature, 227, 680-685.

Le Page, G.A. and J.L. Greenlees (1995) Investigation of disease system for cancer chemotherapy screening: Incorporation of glycine-2-C^{14} into ascites tumor cell purines as a biological test system, Canc. Res. 15, 102-100.

Leon, D.A., F.W. Herberg, P. Banky and S.S. Taylor (1997) A stable alpha-helical domain at the N terminus of the RI-alpha subunits of cAMP-dependent protein kinase is a novel dimerization/docking motif, J. Biol. Chem. 272, 28431-28437.

Lesch, M. and W. Nyhan (1964) A familial disorder of uric acid metabolism and central nervous system function, Am. J. Med. 36, 561-570.

Leung, D., E. Chen and D. Goeddel (1989) A method for random mutagenesis fo a defined DNA segment using a modified polymerase chain reaction, Technique, 1, 11-15.

Levine, R. and M. Taylor (1982) Mecahnism of adenine toxicity in *Escherichia coli*, J. Bact. 149, 923-930.

Liang, P., G. Phillips and M. Glaser (1991) Assignment of the nucleotide binding sites and the mechanism of substrate inhibition of *Escherichia coli* adenylate kinase, Prot. Struct. Func. Genet. 9, 28-36.

Lieberman, I. and P. Ove (1960) Enzyme studies with mutant mammalian cells, J. Biol. Chem. 235, 1765-1768.

Lin-Goerke J., D. Robbins, and J. Burczak (1997) PCR-based random mutagenesis using manganese and reduced dNTP concentration, Biotechniques 23, 409-412

Loetscher, P., G. Pratt and M. Rechsteiner (1991) The C terminus of mouse ornithine

decarboxylase confers rapid degradation on dihydrofolate reductase. Support for the pest hypothesis, J. Biol. Chem. 266, 11213-11220.

Lohse, P.A. and J.W. Szostak (1996) Ribozyme-catalysed amino-acid transfer reactions, Nature, 381, 442-444.
Makrides, S.C. (1996) Strategies for achieving high-level expression of genes in *Escherichia coli*, Microbiol. Rev. 60, 512-538.

Mattano, S., T. Pallela and B. Mitchell (1990) mutations induced at the hypoxanthine-guanine phosphoribosyltransferase locus of human T-lymphocytes by perturbations of purine deoxynucleoside triphosphate pools, Cancer Res. 50, 4566-4571.

McCoy, M., E.S. Stavridi, J.L.F. Waterman, A.M. Wieczorek, S.J. Opella and T.D. Halazonetis (1997) Hydrophobic side-chain size is a determinant of the three-dimensional structure of the p53 oligomerization domain, EMBO J. 16, 6230-6236.

McIlwain, H. (1948) Preparation of cell-free bacterial extracts with powdered alumina, J. Gen. Microbiol. 2, 288-291.

McLachlin, D.T. and S.D. Dunn (1997) Dimerization interactions of the b subunit of the *Escherichia coli* F1F0- ATPase, J. Biol. Chem. 272, 21233-21239.

Montero, C. and P. Llorente (1991) Further studies on the purine phosphoribosyl transferase "burst" velocity, Biochim. Biophys. Acta, 1084, 149-154.

Montgomery, J.A. (1982) Studies on the biologic activities of purine and pyrimidine analogues, Medicin. Res. Rev. 2, 271-270.

Moore, T.F. and D.G. Whittingham (1992) Imprinting of phosphoribosyltransferases during preimplantation development, Development, 115, 1011-1016.

Mullis, K., F. Faloona, S. Scharf, R. Saiki, G. Horn and H. Erlich (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction, Cold Spring Harb. Symp. Quant. Biol. 51 Pt 1:263-73, 263-273.

Murphy, J.M. and E.R. Tully (1984) Adenine metabolism in rat skeletal muscle, Adv. Exp. Med. Biol. 165, 419-424.

Murray, A. (1971) The biological significance of purine salvage, Ann. Rev. Biochem. 40, 811-826.

Musick, W.D. (1977) Preliminary crystallographic studies on quinolinate phosphoribosyl transferase, J. Mol. Biol. 117, 1101-1107.

Musick, W. (1981) Structural features of the phosphoribosyl transferases and their relationship to the human deficiency disorders of purine and pyrimidine metabolism, CRC Crit. Rev. Biochem., 11,1-34.

Nagao, M., K. Wakabayashi, T. Ushijima, M. Toyota, Y. Totsuka and T. Sugimura (1996) Human exposure to carcinogenic heterocyclic amines and their mutational fingerprints in experimental animals, Environ. Health Perspect. 104 Suppl 3:497-501, 497-501.

Nagy, M. and A. Ribet (1977) Purification and comparative study of adenine and guanine phosphoribosyltransferases from *Schizosaccharomyces pombe*, Eur. J. Biochem. 77, 77-85.

Nakamura, M. and M. James (1951) Effect of purine derivatives on *Trypanosoma cruzi in vitro*, Exp. Parasit. 1, 19-26.

Nalbantoglu, J., G. Phear and M. Meuth (1987) DNA sequence analysis of spontaneous mutations at the aprt locus of hamster cells, Molecular. And. Cellular. Biology. 7, 1445-1449.

Natsumeda, Y., N. Prajda, J.P. Donohue, J.L. Glover and G. Weber (1984) Enzymic capacities of purine *de novo* and salvage pathways for nucleotide synthesis in mormal and neoplastic tissues, Cancer Res. 44, 2475-2479.

Needleman, S. and C. Wunsch (1969) A general method applicable to the search for similarities in the amino acid sequence of two proteins, J Mol Biol, 48, 443-453.

Neuhoff, V., N. Arold, D. Taube and W. Ehrhardt (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250, Electrophoresis, 9, 255-262.

Nguyen, B.T., Y.M. El Sayed and W. Sadee (1984) Interaction among the distinct effects of adenine and guanine depletion in mouse lymphoma cells, Cancer Res. 44, 2272-2277.

Nygaard, P. (1976) Functioning of purine salvage pathways in *Escherichia coli*, Adv. Exp. Med. Biol. 76A, 186-195.

Ohta, T. (1992) The nearly neutral theory of molecular evolution, Ann. Rev. Ecol. Syst. 23, 263-286.

Okada, G., I. Kaneko and H. Koyama (1986) Purification and characterization of adenine phosphoribosyl transferase from mouse mammary carcinoma FM3A cells in culture, Biochim. Biophys. Acta, 884, 304-310.

Olsen, A. and G. Milman (1974) Chinese hamster hypoxanthine-guanine phosphoribosyl transferase, J. Biol. Chem. 249, 4030-4037.

Pai, E.F., W. Sachsenheimer, R.H. Schirmer and G.E. Schulz (1977) Substrate positions and induced-fit in crystalline adenylate kinase, J. Mol. Biol. 114, 37-45.

Parissenti A,, L. Su, and H. Riedel (1993) Reconstitution of protein kinase C alpha function by the protein kinase C beta-I carboxy terminus, Mol. Cell. Endocrinol., 98, 9-16.

Parks, R.E., P.R. Brown and C.M. Kong (1973) Incorporation of purine analogs into the nucleotide pools of human erythrocytes, Adv. Exp. Med. Biol. 41A, 117-127.

Parry, D.A. (1982) Coiled-coils in alpha-helix-containing proteins: analysis of the residue types within the heptad repeat and the use of these data in the prediction of coiled-coils in other proteins, Biosci. Rep. 2, 1017-1024.

Parsell, D., Y. Sanchez, J. Stitzel and S. Lindquist (1991) Hsp104 is a highly conserved protein with two essential nucleotide-binding sites, Nature, 353, 270-273.

Parsell, D. and R. Sauer (1989) The structural stability of a protein is an important determinant of its proteolytic susceptibility in *Escherichia coli*, J. Biol. Chem. 264, 7590-7595.

Pascarella, S. and P. Argos (1995) Conservation of amphipathic conformations in multiple protein structural alignments, Protein Engineering, 7, 185-193.

Pauling, L. and R. Corey (1951) Configurations of polypeptide chains with favoured orientations around single bonds: two new pleated sheets, Proc. Nat. Acad. Sci. U.S.A. 37, 729-740.

Pauling, L., R. Corey and H. Branson (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain, Proc. Nat. Acad. Sci. U.S.A. 37, 205-211.

Peltomaki, P. (1991) DNA methylation changes in human testicular cancer, Biochimica Et Biophysica Acta, 1096, 187-196.

Perry, K., J. Onuffer, M. Gittelman, L. Barmat and C. Matthews (1989) Long-range electrostatic interactions can influence the folding, stability, and cooperativity of dihydrofolate reductase, Biochem. 28, 7961-7968.

Persson, B. and P. Argos (1997) Prediction of membrane protein topology utilizing multiple sequence alignments, J. Protein Chem. 1997. Jul. 16, 453-457.

Phear, G., W. Armstrong and M. Meuth (1989) Molecular basis of spontaneous mutation at the *aprt* locus of hamster cells, J. Mol. Biol. 209, 577-582.

Phear, G., J. Nalbantoglu and M. Meuth (1987) Next-nucleotide effects in mutations driven by DNA precursor pool imbalances at the aprt locus of Chinese hamster ovary cells, Proc. Natl. Acad. Sci. U.S.A. 84, 4450-4454.

Phear, G., N.P. Bhattacharyya and M. Meuth (1996) Loss of heterozygosity and base substitution at the *APRT* locus in mismatch-repair-proficient and -deficient colorectal carcinoma cell lines, Mol. Cell Biol. 16, 6516-6523.

Phillips, E.N., F. Xia, K.T. Kelsey and H.L. Liber (1995) Spectra of Spontaneous and X-Ray-Induced Mutations at the *hprt* Locus in, Radiat. Res., 143, 255-262.

Phillips, C.L., B. Ullman and R.G. Brennan (1996) Crystallization of the purine salvage enzyme adenine phosphoribosyl transferase, Proteins, 25, 510-513.

Pieters, R., D.R. Huismans, A.H. Loonen, G.J. Peters, K. Hahlen, A. van der Does-van den Berg, E.R. van Wering and A.J. Veerman (1992) Hypoxanthine-guanine phosphoribosyl transferase in childhood leukemia: relation with immunophenotype, in vitro drug resistance and clinical prognosis, Int. J. Cancer, 51, 213-217.

Pillwein, K., P. Chiba, A. Knoflach, B. Czermack, K. Schucter, E. Gersdorf, B. Ausserer, C. Murr, R. Goebl, G. Stockhammer, H. Maier and H. Kostron (1990) Purine metabolism of human glioblastoma *in vivo*, Cancer Res. 50, 1576-1579.

Popjak, G. (1970) Stereospecificity of enzyme reactions, in: P. Boyer (Ed.), The Enzymes, 3rd Edn., Academic Press, New York, pp. 115-215.

Prochownik, E., K. O'Rourke and V. Dixit (1989) Expression and analysis of COOH-terminal deletions of the human thrombospondin molecule, J. Cell Biol. 109, 843-852.

Queen, S.A., D.L. Vander Jagt and P. Reyes (1989) Characterization of adenine phosphoribosyl transferase from the human malaria parasite, *Plasmodium falciparum*, Biochim. Biophys. Acta, 996, 160-165.

Rao, S.T. and M.G. Rossman (1973) Comparison of super-secondary structures in proteins, J. Mol. Biol. 76, 241-256.

Reinstein, M. Brune and A. Wittinghofer (1988) Mutations in the nucleotide binding loop of adenylate kinase of *Escherichia coli*, Biochem. 27, 4712-4720.

Reinstein, I. Vetter, I. Schlichting, P. Rosch, A. Wittinghofer and R. Goody (1990)

Fluorescence and NMR investigations on the ligand binding properties of adenylate kinases, Biochem. 29, 7440-7450.

Roblin, R., J. Lampen, J. English, Q. Cole and J. Vaughan (1945) Studies in chemotherpay. VIII. Methionine and purine antagonists and their relation to the sulfonamides, J. Am. Chem. Soc. 67, 290-294.

Rossmann, M., D. Moras and K. Olsen (1974) Chemical and biological evolution of a nucleotide-binding protein, Nature, 250, 194-197.

Rost, B. and S. O'Donoghue (1997) Sisyphus and prediction of protein structure, Comput. Appl. Biosci. 13, 345-356.

Rost, B. , R. Schneider and C. Sander (1993) Progress in protein structure prediction, TIBS, 18, 120-123.

Russell, R.B. and G.J. Barton (1994) Structural features can be unconserved in proteins with similar folds. An analysis of side-chain to side-chain contacts secondary structure and accessibility, J. Mol. Biol. 244, 332-350.

Sage, E., B. Lamolet, E. Brulay, E. Moustacchi, A. Chteauneuf and E.A. Drobetsky (1996) Mutagenic specificity of solar UV light in nucleotide excision repair- deficient rodent cells, Proc. Natl. Acad. Sci. U.S.A. 93, 176-180.

Sahota, A., J. Chen, M. Behzadian, H. Takeuchi, P. Stambrook and J. Tischfield (1991) 2,8-dihydroxyadenine lithiasis in a Japanese patient heterozygous at the adenine phosphoribosyltransferase locus, Am. J. Hum. Genet. 48, 983-989.

Sanfilippo, O., M. Camici, M.G. Tozzi, M. Turriani, A. Faranda, P.L. Ipata and R. Silvestrini (1994) Relationship between the levels of purine salvage pathway enzymes and clinical/biological aggressiveness of human colon carcinoma, Cancer Biochem. Biophys. 14, 57-66.

Saraste, M., P.R. Sibbald and A. Wittinghofer (1990) The P-loop - a common motif iin ATP- and GTP-binding proteins, TIBS. 15, 430-434.

Scapin, G., D.H. Ozturk, C. Grubmeyer and J.C. Sacchettini (1995) The crystal structure of the orotate phosphoribosyl transferase complexed with orotate and alpha-D-5-phosphoribosyl-1-pyrophosphate, Biochem., 34, 10744-10754.

Scapin, G., J.C. Sacchettini, A. Dessen, M. Bhatia and C. Grubmeyer (1993) Primary structure and crystallization of orotate phosphoribosyl transferase, J. Mol. Biol. 230, 1304-1308.

Schopf, G., H. Rumpold and M.M. Muller (1986) Alterations of purine salvage pathways during differentiation of rat heart myoblasts towards myocytes, Biochim. Biophys. Acta, 884, 319-325.

Schulz, G., E. Schlitz, A. Tomaselli, R. Frank, M. Brune, A. Wittinghofer and R. Schirmer (1986) Structural relationships in the adenylate kinase family, Eur. J. Biochem. 161, 127-132.

Shao, C., P.K. Gupta, Y. Sun, A. Sahota and J.A. Tischfield (1996) Complex chromosomal mechanisms lead to *APRT* loss of heterozygosity in heteroploid cells, Cytogenet. Cell Genet. 75, 216-221.

Shin, D.H., S.J. Yoo, Y.K. Shim, J.H. Seol, M.S. Kang and C.H. Chung (1996) Mutational analysis of the ATP-binding site in HslU, the ATPase component of HslVU protease in *Escherichia coli*, FEBS Lett. 398, 151-154.

Shindalyov, J., N. Kolchanov and C. Sander (1994) Can three-dimensional contacts in protein structures be predicted by analysis of correlated mutations? Prot. Eng. 7, 349-358.

Simmonds, H., K. van Acker, and A. Sahota (1992) 2,8-Dihydroxyadenine urolithiasis, Lancet 339, 1295-1296

Smith, J. and E. Matthews (1957) The metabolism of 8-azapurines, Biochem. J. 66, 323-333.

Smith, L.E. and A.J. Grosovsky (1993) Genetic instability on chromosome 16 in a human B lymphoblastoid cell line, Somat. Cell Molec. Genet. 19, 515-527.

Smith, R.F., B.A. Wiese, M.K. Wojzynski, D.B. Davison and K.C. Worley (1997) BCM Search Launcher--an integrated interface to molecular biology data base search and analysis services available on the World Wide Web, Genome Res. 6, 454-462.

Smithers, G. and W. O'Sullivan (1982) ^{31}P nuclear magnetic resonance study of phosphoribosyldiphosphate and its interaction with magenesium ions, J. Biol. Chem. 257, 6164-6170.

Smolenski, R.T., C. Montero, J.A. Duley and H.A. Simmonds (1991) Effects of adenosine analogues on ATP concentrations in human erythrocytes: Biochemical Pharmacology, 42, 1767-1774.

Snyder, F.F., M.S. Hershfield and J.E. Seegmiller (1976) Purine toxicity in human lymphoblasts, Adv. Exp. Med. Biol. 76A, 30-39.

Soga, N., R. Suzuki, Y. Komeda and H. Kinbara (1995) A case of 2, 8-dihydroxyadenine ureteral stone, Japanese Journal of Clinical Urology, 49, 411-413.
Spector, L.B. (1982) Covalent Catalysis by Enzymes, Springer-Verlag, New York.

Srivastava, S. and E. Beutler (1971) Purification and kinetic studies of adenine phosphoribosyltransferase from human erythrocytes, Arch. Biochem. Biophys, 142, 426-434.

Stambrook, P.J., C. Shao, M. Stockelman, G. Boivin, S.J. Engle and J.A. Tischfield (1996) APRT: a versatile in vivo resident reporter of local mutation and loss of heterozygosity, Environ. Mol. Mutagen. 28, 471-482.

Stout, J.T. (1985) HPRT: Gene structure, expression, and mutation, Ann. Rev. Genet. 19, 127-148.

Sung, P., D. Higgins, L. Prakash and S. Prakash (1988) Mutation of lysine-48 to arginine in the yeast RAD3 protein abolishes its ATPase and DNA helicase activities but not the ability to bind ATP, EMBO, 7, 3263-3269.

Syed, D.B., R.S. Strauss and D.L. Sloan (1987) Orotate phosphoribosyltransferase and hypoxanthine/guanine phosphoribosyltransferase from yeast: Nuclear magnetic relaxation studies of the structures of enzyme-bound phosphoribosyl 1-pyrophosphate, Biochem. 26, 1051-1058.

Tao, W., C. Grubmeyer and J.S. Blanchard (1996) Transition state structure of *Salmonella typhimurium* orotate, Biochemistry, 35, 14-21.

Taylor, M., A. Simon and Kothari (1985) The APRT system, in: M. Gottesmann (Ed.), Molecular Cell Genetics, John Wiley & Sons Inc. pp. 311-332.

Taylor, W. (1992) Patterns, Predictions, and Problems, in: W. Taylor (Ed.), Patterns in Protein Sequence and Structure, Springer-Verlag, Heidelberg, pp. 1-9.

Thomas, C., W. Arnold and W. Kelley (1973) Human adenine phosphoribosyl transferase: Purification, subunit structure, and substrate specificity, J. Biol. Chem. 248, 2529-2535.
Thompson, J.D., D.G. Higgins and T.J. Gibson (1997) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nuc. Acids Res. 22, 4673-4680.

Thompson, L., S. Fong and K. Brookman (1980) Validation of conditions for efficient detection of *hprt* and *aprt* mutations in suspension-cultured Chinese hamster ovary cells, Mutat. Res. 74, 21-36.

Tian, G., H. Yan, R. Jiang, F. Kishi, A. Nakazawa and M. Tsai (1990) Mechanism of adenylate kinase. Are the essential lysine residues essential? Biochem. 29, 4296-4304.
Tindall, K.R. and T.A. Kunkel (1988) Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase, Biochem., 27, 6008-6013.

Tsai H., and J. Wilson (1995) Functional organization of mammalian hexokinases: characterization of chimeric hexokinases constructed from the N- and C-terminal domains of the rat type I and type II isozymes, Arch. Biochem. Biophys. 316, 206-214

Tso, J.Y., H. Zalkin, M. Cleemput, C. Yanofsky and J.M. Smith (1982) Nucleotide sequence of *Escherichia coli purF* and deduced amino acid sequence of glutamine phosphoribosylpyrophosphate amidotransferase, J. Biol. Chem. 257, 3525-3531.

Tullson, P. and R. Terjung (1991) Adenine nucleotide synthesis in exercising and endurance-trained skeletal muscle, Am. J. Physiol. 261, c342-c347.

Turker, M.S. and G.M. Martin (1985) Induction of adenine salvage in mouse cell lines deficient in adenine phosphoribosyl transferase, Mol. Cell. Biol. 5, 2662-2668.

Uhlen, M., T. Moks and L. Abrahmsen (1988) Protein engineering to optimize recombinant protein purification, Biochem. Soc. Trans. 16, 111-112.

Valencia, A., M. Kjelgaard, E. Pai and C. Sander (1991) GTPase domains of ras p21 oncogene protein and elongation factor Tu: analysis of three dimensional structures, sequence families, and functional sites, Proc. Nat. Acad. Sci. U.S.A. 88, 5443-5447.

Van Diggelen, O.P., T.F. Donahue and S.I. Shin (1979) Basis for differential cellular sensitivity to 8-azaguanine and 6-thioguanine, J. Cell. Physiol. 98, 59-72.

Vetter, I.R., Reinstein and P. Rosch (1990) Complexes of *Escherichia coli* adenylate kinase and nucleotides: ^{1}H NMR studies of the nucleotide sites in solution, Biochem. 29, 7459-7467.

Victor, J., L.B. Greenberg and D.L. Sloan (1979) Studies of the kinetic mechanism of orotate phosphoribosyl transferase from yeast, J. Biol. Chem. 254, 2647-2655.

Viswanadhan, V.N., B. Denckla and J.N. Weinstein (1991) New Joint Prediction Algorithm (Q7-JASEP) Improves the Prediction of Protein Secondary Structure, Biochemistry. 30, 11164-11172.

von Heijne, G. (1987) Sequence Analysis in Molecular Biology: Treasure Trove or Trivial Pursuit, Academic Press Inc. San Diego.

Vos, S., J. De Jersey and J.L. Martin (1996) Crystallization and preliminary X-ray crystallographic studies of *Escherichia coli* xanthine phosphoribosyl transferase, J. Struct. Biol. 116, 330-334.

Walker, B., M. Krishnasastry, L. Zorn and H. Bayley (1992) Assembly of the oligomeric membrane pore formed by Staphylococcal alpha-hemolysin examined by truncation mutagenesis, J. Biol. Chem. 267, 21782-21786.

Walker, J.E., M. Saraste, M.J. Runswick and N.J. Gay (1982) Distantly related sequences in the a- and b- subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold, EMBO J. 1, 945-951.

Wierenga, R.K., P. Terpstra and W.G.J. Hol (1985) Prediction of the occurence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint, J. Mol. Biol. 187, 101-107.

Wilson, J.M., R. Kobayashi, I.H. Fox and W.N. Kelley (1983) Human hypoxanthine-guanine phosphoribosyltransferase. Molecular abnormality in a mutant form of the enzyme (HPRT/Toronto), J. Biol. Chem. 258, 6458-6460.

Wilson, J.M. and W.N. Kelley (1984) Human hypoxanthine-guanine phosphoribosyltransferase. Structural alteration in a dysfunctional enzyme variant ($HPRT_{Munich}$) isolated from a patient with gout, J. Biol. Chem. 259, 27-30.

Wu, B., A. Wawrzynow, M. Zylicz and C. Georgopoulos (1996) Structure-function analysis of the *Escherichia coli* GrpE heat shock protein, EMBO J. 15, 4806-4816.

Yadollahi-Farsani, M., N.J. Gooderham, D.S. Davies and A.R. Boobis (1996) Mutational spectra of the dietary carcinogen 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine(PhIP) at the Chinese hamsters *hprt* locus, Carcinogenesis, 17, 617-624.

Yokota, N., H. Yamanaka, Y. Yamamoto, S. Fujimoto, T. Eto and K. Tanaka (1991) Autosomal dominant transmission of gouty arthritis with renal disease in a, Annals of the Rheumatic Diseases, 50, 108-111.

Yokuma, Y., S. Hashimoto, R. Roeder and M. Horikoshi (1992) Structural conservation of putative functional motifs between *Xenopus* and human TFIIE-b, Nuc. Acids Res. 20, 4363

Yuan, L., S.P. Craig,Iii, J.H. Mckerrow and C.C. Wang (1992) Steady-state kinetics of the schistosomal hypoxanthine-guanine, Biochem., 31, 806-810.

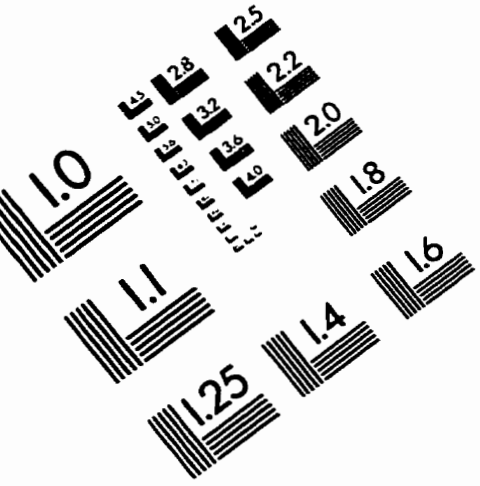
Zhang H., S. Malpure, Z. Li, H. Hiasa, and R. DiGate (1996) The role of the carboxyl-

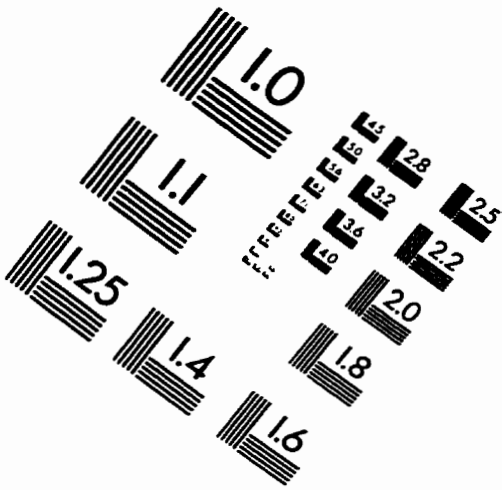
terminal amino acid residues in *Escherichia coli* DNA topoisomerase III-mediated catalysis, J. Biol. Chem. 271, 9039-9045
Zoellner, N. and U. Gresser (1990) Nephrolithiasis in twins with APRT-deficiency: Stones as a marker of an, Bildgebung, 57, 64-66.

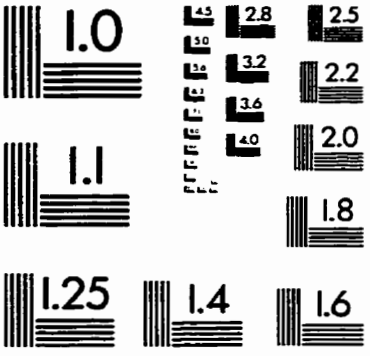
Zoller, M. and M. Smith (1982) Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any fragment of DNA, Nuc. Acids Res. 10, 6487-6500.

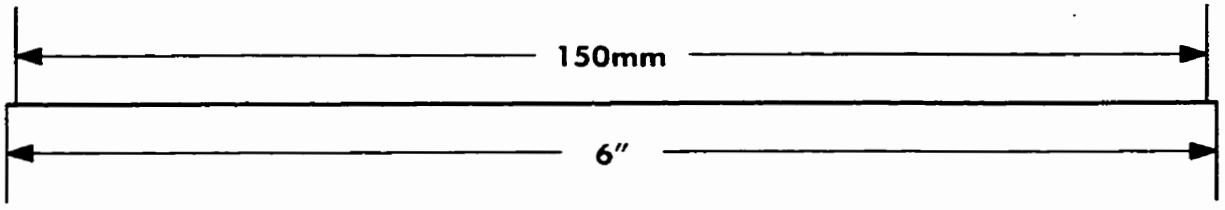
Zhou, N.E., B.Y. Zhu, C.M. Kay and R.S. Hodges (1992) The two-stranded alpha-helical coiled-coil is an ideal model for studying protein stability and subunit interactions, Biopolymers, 32, 419-426.

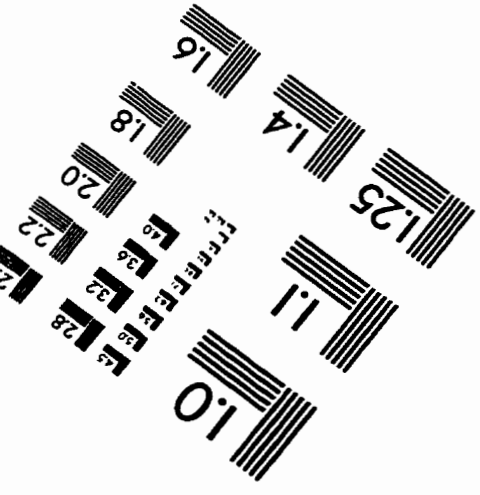
IMAGE EVALUATION
TEST TARGET (QA-3)

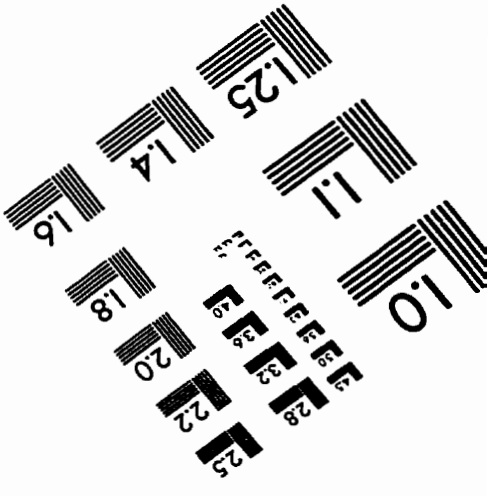












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989