Chapter 1: Introduction and literature review

1.1 The impact of inborn errors of metabolism

About two thirds of genetic disorders reduce life expectancy, of which about three quarters result in death before the age of 30 (Jimenez-Sanchez et al., 2001). About 5% of live-born individuals can be expected to experience disease with an important genetic component before the age of 25 years. This was based on a large study of over 1 000 000 consecutive live-births. Most of these diseases represent simple monogenic disorders (Baird et al., 1988). In a study on patients admitted to a paediatric hospital, it was found that about one third of admissions reflected an abnormal gene-environment interaction. This means that the disease is caused by either a simple Mendelian defect, or more complex congenital malformations (Scriver et al., 1973). These diseases result not only in suffering for the patient, but also places a tremendous burden on the family and society (Jimenez-Sanchez et al., 2001, Lee, 2001). This is because the management of genetic disease is difficult and expensive (Lee, 2001).

Major infectious diseases such as AIDS and tuberculosis kill millions of people every year. Consequently, considerable research effort is directed at treatment for these diseases, and this is justified. It is, however, becoming clear that heritable diseases also pose a considerable burden to society. It has been estimated that about two thirds of individuals in developed societies will in their life experience a complex disease with some genetic component. This is a major impact. Therefore, complex genetic diseases have to be taken seriously, even if the threat posed seems less urgent than that of infectious diseases. More importantly, heritable diseases will become ever more important in the future, due to increasingly improved health care and treatment of infectious disease. Because people live longer, complex diseases that appear with old age will also become a more prominent problem (Baird et al., 1988, Jimenez-Sanchez et al., 2001, Lee, 2001).

If disease is reconsidered from an evolutionary perspective, a general model of pathogenesis emerges. This model is important to modern understanding of medicine. The main idea is simple, and will be very briefly introduced. In The Metabolic and Molecular Basis of Inherited Disease, Barton Childs develops a concept of disease as an "incongruence between the genome and its environment." Albeit that this is a broad and vague definition, it is biologically relevant (Childs, 2001, Childs et al., 2001). The human organism is a complex homeostatic system built by evolutionary adaption to its environment. Disease arises when an environment is encountered for which the genome does not have appropriate instructions to maintain homeostasis. All disease must thus have some genetic/heritable component. Even diseases that are caused by external agents (pathogens, toxins and physical conditions) are genetic in the sense that most environmental elements interact with the organism ultimately at the molecular level, a level maintained by the genome (Hill, 2001). This view is supported by the inter-individual variability in susceptibility to most environmental threats, including toxic exposure and infectious diseases such as tuberculosis, malaria, and HIV (Hill, 2001, Roy et al., 2007).

Ultimately, a complex disorder (one involving several genetic and environmental factors) must be a product of its components. As the study of monogenic disorders and their treatment lead to better understanding of the genetic/molecular components of disease, it should also enhance understanding and treatment of the more numerous complex disorders (Lanpher et al., 2006, Scriver, 2004, Treacy, 2001).

1.2 The scope of this study

Detoxification is an important physiological process as it serves to decrease the toxicity of compounds that are not catabolised (Liska, 1998). These compounds include several endogenous metabolites, such as steroid hormones, and exogenous toxins such as compounds in food or industrial chemicals (Campbell et al., 1988). Detoxification is mainly divided into three phases. Phase I detoxification activates metabolites by adding functional groups to them. Metabolites may be hydroxylated by the cytochrome P450 system, for example. The activated compounds that result are often more toxic than the original

compounds, and are rapidly acted on by phase II detoxification systems (Kinzig-Schippers et al., 2005). These reactions include methylation and conjugation to sulfate, glucuronide, and glycine. All these reactions serve to make compounds less toxic and more soluble for excretion in the urine and bile. Phase III detoxification is involved in the export of toxins from cells. Conjugation of glycine to organic acids is a detoxification mechanism of central importance to the management of a range of metabolic disorders (Tanaka & Isselbacher, 1967). The enzyme involved, glycine N-acyltransferase, is the topic of investigation in this study. It is an enzyme involved in phase II detoxification.

One of the factors that influence the clinical outcome of organic acidemias is the efficiency with which the toxic metabolites that accumulate are detoxified (Sweetman & Williams, 2001, Tanaka et al., 1966, Tanaka & Isselbacher, 1967). Glycine N-acyltransferase (GLYAT, EC 2.1.3.13) is an enzyme responsible for the detoxification of several such toxic organic acids by means of conjugation to glycine (Nandi et al., 1979, Schachter & Taggart, 1954a). It has a significant influence, for example, on the clinical outcome of isovaleric acidemia (Tanaka et al., 1966). Because it efficiently detoxifies isovaleric acid, the principle toxic metabolite. Glycine supplementation is often a successful treatment for this disease, as it enhances this detoxification process. Propionic acid, however, is not a good substrate for GLYAT, and is thus not very efficiently detoxified by conjugation to glycine. Propionic acidemia, a disease in which propionic acid accumulates, is thus not efficiently treatable using glycine supplementation. A satisfactory treatment for propionic acidemia has not yet been developed (Fenton et al., 2001).

Recently, however, we became aware of an individual with propionic acidemia whom is relatively healthy. The levels of propionylglycine excreted in the urine of this individual is three orders of magnitude greater than the normal or reference values for this disease. This individual was identified by the Metabolic Laboratory of the North-West University (Prof LJ Mienie, personal communication). This individual's urinary metabolite profile resembles that of isovaleric acidemia, in the sense that the accumulating toxic metabolite is excreted in high quantities as the glycine conjugate. Consequently, glycine supplementation seems to be beneficial for this individual.

Our hypothesis is that this individual may have a novel variant of the GLYAT enzyme, which is capable of efficiently detoxifying propionic acid, as the extremely elevated levels of propionylglycine excreted is otherwise difficult to explain. One of our long term goals is to develop a recombinant therapeutic GLYAT enzyme with these properties, with the aim of using it to treat propionic acidemia. If the molecular mechanisms of substrate binding and specificity are understood, it may even be possible to manipulate the enzyme to use other substrates efficiently and, consequently, to treat a variety of organic acidemias. The propionic acidemia patient mentioned may very well provide *in vivo* support for our hypothesis. Unfortunately, it has not been possible to study the GLYAT gene of this individual, as a DNA sample has not been obtained, but this will be done in a future study.

In the remainder of this chapter the literature relevant to glycine N-acyltransferase, the GNAT superfamily of acetyltransferases, and recombinant protein expression and purification is very briefly reviewed. Instead of an encyclopaedic coverage of these topics, only the information relevant to discussions in later chapters is presented here.

1.3 Inborn errors of metabolism

This field is very large and complicated. For simplicity, the focus here will be limited to inborn errors of organic acid metabolism (organic acidemias) and the available treatment strategies.

Inborn errors of metabolism (IEMs) are genetic diseases caused by a mutation or some other heritable structural anomaly in the genome (Childs et al., 2001, Garrod, 1902, Hurles et al., 2008). These genetic aberrations lead to biochemical abnormalities and disease (Childs et al., 2001). Metabolism can be seen as a very complex network of chemical reactions needed to make life possible (Childs, 2001, Lanpher et al., 2006). These reactions have to be catalysed in order to proceed at sufficient rates at physiological temperatures. In living organisms, the catalysts for these various chemical reactions are complex protein molecules called enzymes. Enzymes catalyse reactions by lowering the activation energies of rate limiting steps and by coupling energetically unfavourable reactions to ones that are energetically favourable (Palmer, 2001). The nature of these

catalysts also makes complex regulation of metabolic flux possible. The levels of specific enzyme molecules in the cell can be regulated, and several enzymes can be allosterically regulated by metabolites from either the same or different metabolic pathways (Lanpher et al., 2006, Palmer, 2001).

Since enzymes are coded for by the genes of an organism, it follows that a genetic aberration can lead to the malfunctioning of an enzyme. This then results in dysfunction of the respective metabolic pathway, disrupting homeostasis and causing disease (Lanpher et al., 2006, Palmer, 2001). This concept is demonstrated in Figure 1.1. In the past, focus has been mostly on simple one-gene disorders, but it is becoming clear that multiple partially defective genes may conspire to cause a complex disease phenotype. For example, it has recently been demonstrated that compound heterozygosity (heterozygosity for a defective allele at more than one point in a metabolic pathway, such as the beta-oxidation pathway) can result in complex disease phenotypes (Vockley, 2008).

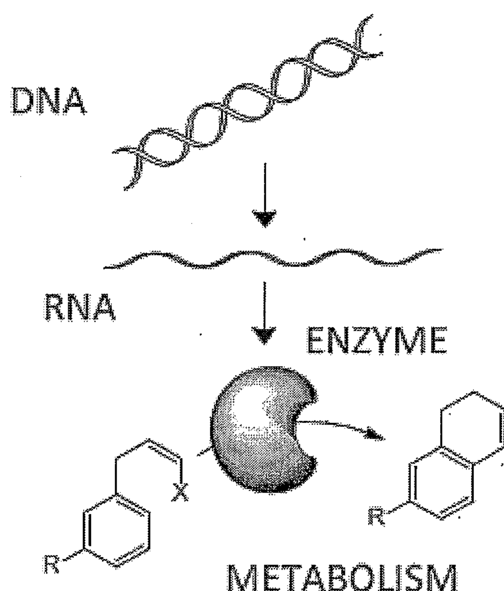

Figure 1.1 The relationship between DNA and metabolism. The figure demonstrates the flow of information from the DNA sequence, through RNA, to the structure of an enzyme that makes a biochemical reaction possible. A defective DNA sequence leads to a defective enzyme and this leads to deregulation of metabolic homeostasis.

Pathogenesis of an inborn error of metabolism is influenced by four main factors, as demonstrated in Figure 1.2. These are the inherent toxicity of metabolites that accumulate

upstream of the deficient enzyme, deficiency of essential metabolites formed downstream of the defective enzyme, the influence of accumulating metabolites on other metabolic pathways and the extent to which these metabolites are diverted to secondary pathways (which can form more or less toxic compounds). Usually some combination of these effects contributes to the disease pathology (Childs, 2001, Lanpher et al., 2006).

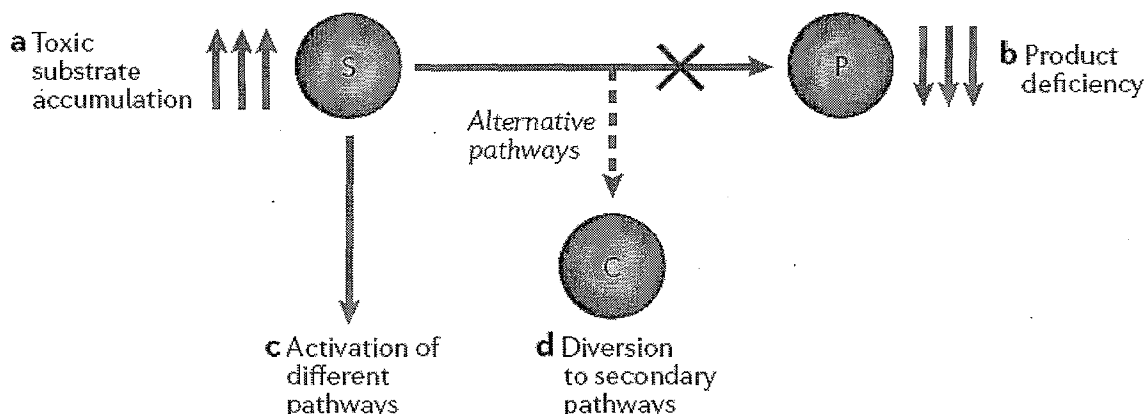

Figure 1.2 The four main contributing factors to pathogenesis in metabolic disorders. A genetic defect, leading to an enzyme deficiency, symbolised by the cross, has four main pathogenic effects: a) the metabolite upstream of the defective enzyme may accumulate to toxic levels; b) pathogenesis can also result from the deficiency of a metabolite formed downstream of the defective enzyme; c) the accumulating substrate of the defective enzyme may activate or inhibit other metabolic pathways, by allosteric regulation, for example; d) the accumulating metabolites may be diverted to secondary metabolic pathways, which may result in the formation of either more or less toxic metabolites (Lanpher et al., 2006).

1.3.1 Factors that influence the severity of inborn errors of metabolism

Understanding the mechanism of pathogenesis is of central importance in the treatment of disease. There are three principle factors determining the severity and clinical outcome of an IEM. The first is the severity of the primary enzyme deficiency; homozygosity or heterozygosity, total or partial loss of activity, or inefficient expression, all of which result in different levels of residual enzyme activity. The second is the extent to which accumulating metabolites are converted to toxic secondary metabolites. The third, and most important for this study, is the efficiency with which toxic metabolites are detoxified and excreted (Bartlett & Gompertz, 1974, Childs, 2001, Lanpher et al., 2006, Tanaka et al., 1966).

1.3.2 Treatment of inborn errors of metabolism

The implications of the definition of disease mentioned in the introduction are twofold: not only does every disease have a genetic component, but every genetic aberration will find itself modulated by the environment (Childs, 2001, Lanpher et al., 2006). Treatment of metabolic disorders is based on this philosophy. By modifying the environment, including diet, medication and physical activity, the disease phenotype can be manipulated (Treacy, 2001).

Accumulation of toxic metabolites can be prevented by several means. Dietary restriction can reduce the load on a defective pathway, limiting the accumulation of toxic metabolites. This strategy is employed for several disorders, including phenylketonuria, galactosemia and several organic acidemias (Sweetman & Williams, 2001, Treacy, 2001). A different approach is to divert accumulating metabolites to alternative pathways that form less toxic products which can then be excreted from the body. This is commonly the main therapeutic strategy for disorders of nitrogen, fatty acid, amino acid and organic acid metabolism. For example, patients with urea cycle defects or non-ketotic hyperglycinemia are treated with benzoic acid or 4-phenylbutyric acid, since conjugation of these compounds to glycine and consequent excretion removes excess nitrogen (or glycine) from the system. In defects of organic acid or fatty acid metabolism, supplementation with carnitine and glycine usually leads to increased synthesis and excretion of acyl-carnitines and acylglycines, effectively detoxifying substantial quantities of the toxic organic acids (Brusilow & Horwich, 2001, Dakin, 1910, Kanazu & Yamaguchi, 1997, Sweetman & Williams, 2001, Tanaka & Isselbacher, 1967, Treacy, 2001).

Other means of reducing toxic metabolite levels are more complicated and focus on the defective genes and proteins themselves. These include protein/enzyme replacement therapy, protein enhancement therapy, gene therapy and cell- or organ transplants (Treacy, 2001). Of particular importance for this study is enzyme replacement therapy, as one of the long term goals of our research is the development of a therapeutic enzyme for the treatment of organic acidemias.

Enzyme replacement therapy has been used to effectively treat several lysosomal storage diseases (for example Gaucher disease, Fabry disease and Pompe disease), adenosine deaminase deficiency, cystic fibrosis and other diseases (Brady et al., 1974, Cantz & Kresse, 1974, Chan et al., 2005, Munck et al., 2009). The treatment strategy usually involves intravenous infusion with an enzyme, targeted to the desired sub-cellular location where it performs the function of a deficient enzyme (Barton et al., 1990). The enzymes used for therapy were initially purified from human material, but several recombinant therapeutic proteins have now been developed (Desnick, 2004, Martiniuk et al., 2000).

Enzyme replacement therapy was first proposed by de Duve in 1964 for the treatment of lysosomal storage diseases (Desnick & Schuchman, 2002). The discovery that lysosomal enzymes are targeted to lysosomes by the mannose-6-phosphate receptor-mediated pathway led to cell culture studies. Cultured patient fibroblasts can take up exogenously supplied enzyme from the medium, resulting in the degradation of accumulated lysosomal substrates (Cantz & Kresse, 1974). Consequently, this approach was used with considerable success in the 1980s in clinical trials for the treatment of Gaucher disease. The effects were limited, however, presumably because of the low doses that were administered (Brady et al., 1980). When it was reported in the early 1990s that administration of large doses of 1-3 mg/kg of purified enzyme had clinical benefits, such as reversal of hepatomegaly, enthusiasm for this treatment strategy started to grow (Barton et al., 1990). Acid β-glucosidase, the enzyme used for treatment of Gaucher disease, was initially industrially purified from human placenta, but a recombinant enzyme expressed in Chinese hamster ovary cells, shown to be as effective, has been developed (Grabowski et al., 1995, Grabowski et al., 1998). This enzyme is now used in therapeutic regimes. In 2000, a recombinant acid alpha glucosidase (GAA) was produced and shown to be effectively taken up by cell cultures and the tissues of GAA deficient mice. This recombinant enzyme was approved for the treatment of human glycogen storage disease type II (Martiniuk et al., 2000).

Recently, enzyme replacement therapy was used to treat deficiency of a mitochondrial protein, lipoamide dehydrogenase. This was done by recombinant expression of the lipoamide dehydrogenase (LAD), fused to a signal peptide that directs the protein into hepatocytes and then targets the mitochondria for entry. This approach has been shown to

correct the enzyme deficiency in cultured patient hepatocytes. Preliminary trials with LAD-deficient mice show positive results (Rapoport et al., 2008). This is encouraging news, since it suggests that, at least in principle, it should be possible to direct a recombinant therapeutic glycine N-acyltransferase to liver mitochondria for the treatment of organic acidemias.

This brief glance shows that enzyme replacement therapy is an exciting and rapidly developing field of study. In the near future many more diseases may be treatable using this strategy. The next section discusses the organic acidemias which are defects of organic acid metabolism (Ogier & Saudubray, 2002, Sweetman & Williams, 2001).

1.3.3 Organic acidemias

A significant group of inborn errors of metabolism are the organic acidemias, disorders involving the catabolism of various organic acids. These diseases often present at birth or early infancy and are usually severely debilitating or fatal. Normal development is usually hampered and mental retardation frequently occurs (Ogier & Saudubray, 2002). Pathogenesis in these disorders usually results from accumulation of organic acids that cannot be further metabolised, to toxic levels. Apart from the metabolic acidosis caused, certain organic acids can have specific toxic effects (Bartlett & Gompertz, 1974, Fenton et al., 2001, Sweetman & Williams, 2001). Examples of organic acidemias are propionic acidemia, isovaleric acidemia, glutaric acidemia, and methylmalonic acidemia (Fenton et al., 2001, Sweetman & Williams, 2001). Because accumulating organic acids are often substrates for glycine N-acyltransferase, the enzyme investigated in this study, two of these disorders will be discussed in detail to demonstrate the principles of pathogenesis and treatment (Bartlett & Gompertz, 1974, Ogier & Saudubray, 2002).

1.3.3.1 Isovaleric acidemia

Isovaleric acidemia (MIM 243 500) is a congenital defect of leucine catabolism that results from any of several defects in the enzyme isovaleryl-coenzyme A dehydrogenase, as demonstrated in Figure 1.3 (Mohsen et al., 1998, Sweetman & Williams, 2001, Tanaka et

al., 1966, Tanaka & Isselbacher, 1967, Tanaka et al., 1972). The defect leads to accumulation of isovaleryl-coenzyme A in cells, which is subsequently deacylated to form the toxic isovaleric- and hydroxy-isovaleric acids, as shown in Figure 1.3. Accumulation of these toxic organic acids is largely responsible for pathogenesis in this disease (Sweetman & Williams, 2001, Tanaka et al., 1966). The accumulation of isovaleryl-coenzyme A also results in depletion of free coenzyme A, with negative consequences for cellular physiology (Mitchell et al., 2008, Sweetman & Williams, 2001).

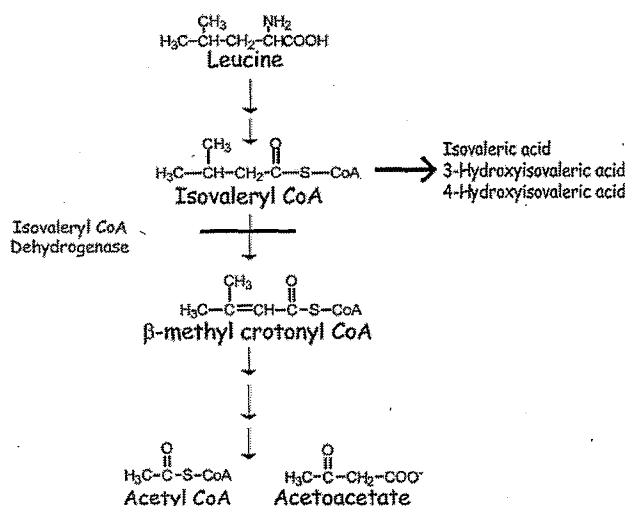

Figure 1.3 The leucine catabolic pathway, showing a defect of isovaleryl-coenzyme A dehydrogenase, the cause of isovaleric acidemia. The horizontal bar represents a defect of isovaleryl-coenzyme A dehydrogenase, which leads to accumulation of isovaleryl-coenzyme A. This metabolite is then converted to isovaleric acid and the hydroxy-isovaleric acids, the principal toxic metabolites of the disease (Sweetman & Williams, 2001).

The disease may present with either severe acute neonatal illness followed by coma and death, or a mild intermittent form characterised by recurrent episodes of metabolic acidosis throughout life. The clinical presentation does not seem to be correlated to the extent of the enzyme deficiency. Rather, the distinction between the two situations seems to depend, in part, on timing of the first episode of catabolic stress such as high protein intake, starvation, viral infection and other events leading to high rates of protein catabolism (Sweetman & Williams, 2001).

The disease is one of the few great success stories in the treatment of inborn errors of metabolism. A diet restricted in leucine decreases the load on the leucine catabolic

pathway, and decreases accumulation of isovaleryl-coenzyme A. Additionally, glycine and carnitine supplementation is prescribed. This results in increased flux of isovaleryl-coenzyme A to isovaleryl-carnitine and isovalerylglycine. These are less toxic and more soluble metabolites that are excreted in the urine. The improvement in clinical presentation is remarkable and results from the decreased levels of toxic metabolites and the increased release of free coenzyme A, which normalises cellular metabolism (Budd et al., 1967, Itoh et al., 1996, Mitchell et al., 2008, Sweetman & Williams, 2001, Tanaka & Isselbacher, 1967). The glycine supplementation strategy was implemented after the finding that large amounts of isovalerylglycine are excreted in the urine of patients with the disease (Tanaka et al., 1966, Tanaka & Isselbacher, 1967). It was argued that under stress conditions glycine may become limiting to the formation of isovalerylglycine and that supplementation could prevent this. This is supported by the observation that glycine limitation limits the synthesis of hippurate, the glycine conjugate of benzoic acid, in rats (Beliveau & Brusilow, 1987, Tanaka & Isselbacher, 1967). Situations that lead to catabolic stress, such as fasting and infection, may still cause an acute episode of metabolic decompensation and acidosis, but the severity and time of recovery are both decreased by rapid administration of glycine (Sweetman & Williams, 2001).

1.3.3.2 Propionic acidemia

Propionic acidemia (MIM 232 000 and MIM 232 050) is caused by a variety of defects in either of the two subunits (a and β) of propionyl-coenzyme A carboxylase (Campeau et al., 1999, Desviat et al., 2004, Perez-Cerda et al., 2003). Propionyl-coenzyme A is derived from many sources, including valine and isoleucine catabolism, odd-chain fatty acid degradation, nucleotide metabolism and the metabolism of intestinal microflora. A defect of propionyl-coenzyme A carboxylase results in accumulation of propionyl-coenzyme A (with concomitant depletion of free coenzyme A, as described above), which gives rise to propionic acid, 3-hydroxypropionic acid and several other toxic derivatives that accumulate in the tissues and blood. These compounds, demonstrated in Figure 1.4, and the associated acidosis, are responsible to a large extent for the disease pathology (Bonafe et al., 2000, Fenton et al., 2001, Mitchell et al., 2008, Ogier & Saudubray, 2002).

Most patients present in the early neonatal period with severe metabolic acidosis manifested by refusal to feed, vomiting, lethargy, hypotonia and developmental delay. The clinical presentation is very heterogeneous, ranging from virtually asymptomatic (relatively rare) to death in early infancy (more common). The basis for this phenotypic heterogeneity is not fully understood at present, but heterogeneity in the residual levels of carboxylase activity seems to explain only a small part of the variation between individuals (Fenton et al., 2001).

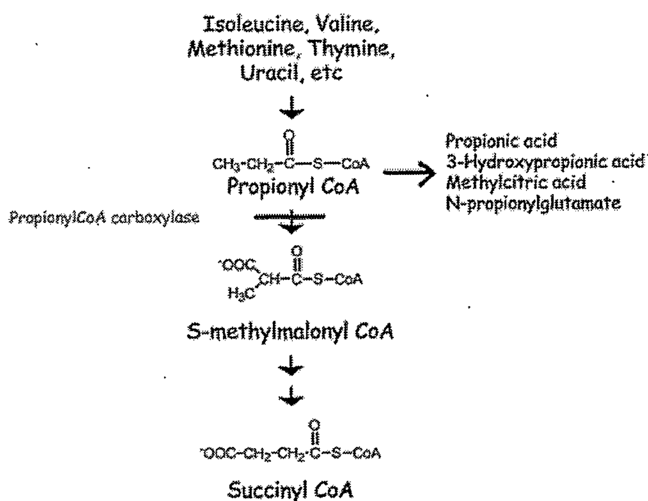

Figure 1.4 The major catabolic pathway for the metabolism of propionic acid. Propionate precursors lead to the formation of propionyl-coenzyme A, which is then metabolised to succinyl-coenzyme A. Propionic acidemia is caused by a defect of propionyl-coenzyme A carboxylase, represented by the horizontal bar. Consequent accumulation of propionyl-coenzyme A leads to formation of the toxic metabolites shown on the right (Fenton et al., 2001).

Propionate precursors are almost ubiquitous and treatment by dietary restriction is extremely difficult. Frequent feeding of a low protein diet (to avoid catabolic stress and unnecessary protein break down), peritoneal dialysis and supplementation with L-carnitine, all decrease levels of toxic metabolites, but the improvement in clinical presentation is usually not satisfactory. Treatment with metronidazole, an antibiotic that reduces the production of propionic acid by gut microflora, which accounts for up to 25% of the propionate load, is effective for only a short time (before the microbes adapt). In the severe cases, treatment does not seem to be very efficacious, and neurological abnormalities appear to be inevitable (Fenton et al., 2001).

The treatment strategy employed for isovaleric acidemia, supplementation with glycine to enhance detoxification, is not very successful for the treatment of propionic acidemia, since

propionyl-coenzyme A is not efficiently detoxified by conjugation to glycine. Glycine supplementation would also not help much because one of the natural consequences of propionic acidemia is hyperglycinemia. Supplementation with carnitine does have some beneficial effects, due to formation and excretion of C3-carnitine (Fenton et al., 2001, Ogier & Saudubray, 2002).

As mentioned in the introduction, this difference in the effect of glycine supplementation can be attributed to the efficiency of formation of the N-acylglycine corresponding to the accumulating acyl-coenzyme A, a process catalysed by the enzyme glycine N-acyltransferase. The formation of isovalerylglycine is better catalysed by this enzyme than the formation of propionylglycine, which explains the inefficiency of detoxification in the case of propionic acidemia (Bartlett & Gompertz, 1974). Glycine N-acyltransferase and its substrate specificity will be discussed in the following sections.

1.4 The properties of glycine N-acyltransferase

1.4.1 Enzymatic reaction and physiology

Glycine N-acyltransferase is an enzyme involved in phase II detoxification, a set of conjugation reactions that serve to make compounds less toxic and more soluble for excretion in the urine and bile. GLYAT detoxifies a wide range of xenobiotic and endogenous metabolites. These include benzoic acid (a compound found in fruits and vegetables and used in medicine and foodstuffs as a preservative), salicylic acid (a metabolite of aspirin), methyl-benzoic acid (a product of the metabolism of toluene, an industrial solvent) and several endogenous metabolites. The latter include isovaleric acid and propionic acid (Bartlett & Gompertz, 1974, Campbell et al., 1988, Fenton et al., 2001, Kolvraa & Gregersen, 1986, Nandi et al., 1979, Ogier & Saudubray, 2002, Schachter & Taggart, 1954a, Sweetman & Williams, 2001, Tanaka & Isselbacher, 1967, Temellini et al., 1993, van der Westhuizen et al., 2000). The diversity of endogenous compounds detoxified is demonstrated by the wide range of acylglycines excreted in the urine of patients with defects of organic acid metabolism (Table 1.1) (Bartlett & Gompertz, 1974, Blau et al.,

2003, Gregersen et al., 1976, Ogier & Saudubray, 2002, Sweetman & Williams, 2001, Tanaka & Isselbacher, 1967). No defect of GLYAT has yet been described, but it has been demonstrated that there is significant inter-individual variation in glycine conjugation capacity (Temellini et al., 1993). The molecular and biochemical basis for this variability is not well characterised, but factors that may influence the rate of conjugation include availability of glycine and transcriptional or translational regulation of the gene. Enzymes that influence acyl-coenzyme A metabolism, including acyl-coenzyme A synthetases, acyl-coenzyme A hydrolases, carnitine-acyltransferase and coenzyme A biosynthetic enzymes, may also contribute to the variability (Beliveau & Brusilow, 1987, Chantrenne, 1951, Loots et al., 2005, Loots et al., 2007, Mitchell et al., 2008, Ogier & Saudubray, 2002, Sweetman & Williams, 2001). This is demonstrated in Figure 1.5.

Table 1.1: Glycine conjugates detected in the urines of patients with various inborn errors of organic acid metabolism

Glycine conjugate	Disease or enzyme deficiency	Urine levels (mmol/mol creatinine)*
Isovalerylglycine	Isovaleric acidemia	2000-9000 0-1000
3-Methylcrotonylglycine	3-methylcrotonyl-coenzyme A carboxylase deficiency	400-1000 30-260
Tiglylglycine	Ketothiolase deficiency	0-1000
Hexanoylglycine	MCAD	2-730
Suberylglycine	MCAD	6-2200
Butyrylglycine	SCAD	Not reported
Isobutyrylglycine	3-Hydroxyisobutyryl-coenzyme A dehydrogenase	0-200
2-Methylbutyrylglycine	Multiple acyl-coenzyme A dehydrogenase deficiency	0-200
Propionylglycine	Propionic acidemia Methylmalonic acidemia	0-450
Methylmalonylglycine	Methylmalonic acidemia	Not reported

*Values were compiled from (Blau et al., 2003, Sweetman & Williams, 2001).

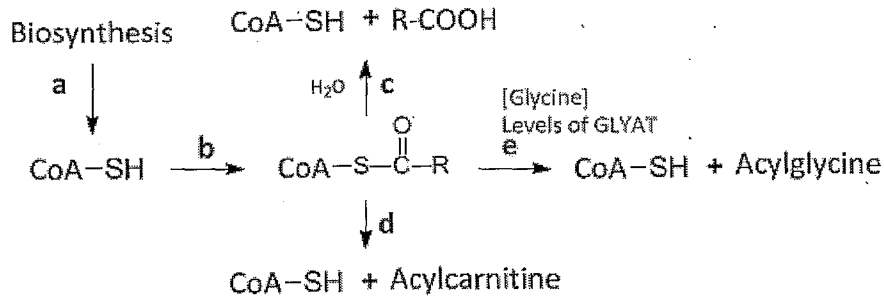

Figure 1.5 The major processes in acyl-coenzyme A metabolism. This diagram represents the elements of acyl-coenzyme A metabolism that are relevant to understanding variability in glycine conjugation capacity: a) there may be variation in the coenzyme A biosynthetic pathway itself, influencing the levels of free coenzyme A; b) acyl-coenzyme A species are formed by a variety of metabolic processes, and by acyl-coenzyme A synthetases such as benzoic acid-coenzyme A ligase; c) acyl-coenzyme A hydrolases release free coenzyme A; d) carnitine-acyltransferases transfer the acyl group of an acyl-coenzyme A to carnitine to form an acyl-carnitine; e) glycine N-acyltransferase forms acylglycines from acyl-coenzyme A esters, and this is may be influenced by the availability of glycine, the levels of the GLYAT enzyme, and polymorphisms of the GLYAT gene.

The exact physiological role of GLYAT is hard to define, since the origins of its various substrates are so diverse (Bartlett & Gompertz, 1974, Bonafe et al., 2000, Brandt et al., 1968, Gatley & Sherratt, 1977, Gregersen et al., 1986, Gregersen et al., 1976, Gron et al., 1978a, Gron et al., 1978b, Kelley & Vessey, 1986, Kolvraa & Gregersen, 1986, Kolvraa et al., 1980, Merkler et al., 1996, Merkler et al., 1999, Ogier & Saudubray, 2002, Schachter & Taggart, 1954a, Tanaka & Isselbacher, 1967, van der Westhuizen et al., 2000). It appears that the principal substrate for the enzyme is benzoyl-coenzyme A, as it has one of the lowest K_M values and has the highest reaction velocity listed in Table 1.2 (gives the largest V_{max} values). The common presence of hippuric acid in herbivore and omnivore urine can be explained by the abundance of benzoic acid and related phenolic compounds in plant material, a significant dietary source (Nandi et al., 1979, Schachter & Taggart, 1954a). Whether the ability to also utilise endogenous acyl-coenzyme A substrates is a chance consequence of active site architecture, or has been selected for because it confers some advantage to the organism, is not known. It has been suggested that GLYAT may be involved in the transfer of acyl groups (fatty acids) to amino groups of proteins, perhaps a mechanism of post-translational modification or a regulatory mechanism (similar perhaps to protein myristoylation). There may also be a role for GLYAT in hormone metabolism. By transferring glycine to medium- or long-chain fatty acids, fatty-acylglycines are formed. These compounds are metabolised by oxidase enzymes to form the corresponding fatty-

acid amides. These compounds act as hormones with diverse physiological effects, including the induction of sleep, analgesic effects, and the potentiation of the response of 5-HT2 receptors to serotonin. One such hormone, anandamide, has been shown to arrest the development of pre-implantation mouse embryos (Merkler et al., 1996, Merkler et al., 1999).

Despite the current lack of understanding of the evolutionary physiology of GLYAT, as discussed above, it is clear that the wide substrate range of the enzyme has significant implications for the clinical outcome and treatment of organic acidemias (Bartlett & Gompertz, 1974, Itoh et al., 1996, Ogier & Saudubray, 2002, Sweetman & Williams, 2001).

1.4.2 Enzyme localisation, kinetics, reaction mechanism and pH dependence

Using photoaffinity labelling experiments, it was determined that GLYAT has only one active site per enzyme molecule and functions as a monomer in the mitochondrial matrix (Gatley & Sherratt, 1977, Lau et al., 1977). The catalytic mechanism has not been studied at the molecular level. This may be due to lack of a recombinant expression system for GLYAT. The kinetic parameters have been studied in some detail on preparations of the enzyme purified from human, bovine, murine, and ovine liver or kidney. The reaction has a sequential mechanism, with the acyl-coenzyme A binding before glycine to form a ternary complex. This is followed by nucleophilic attack and direct acyl transfer, as demonstrated in Figure 3.21 (the reaction mechanism will be discussed at length in Chapter 3). The coenzyme A product then leaves first, followed by the peptide conjugate. Product release is the rate limiting step of the reaction (Brandt et al., 1968, Kelley & Vessey, 1993, Kelley & Vessey, 1994, Mawal & Qureshi, 1994, Nandi et al., 1979, Schachter & Taggart, 1954b, van der Westhuizen et al., 2000). Solution pH has a significant influence on the rate of reaction (determined for the human and bovine enzymes), with the rate increasing from near inactive at pH 6.0 to a maximum from pH 7.5 to pH 9.2, as demonstrated in Figure 1.6 (Mawal & Qureshi, 1994, Schachter & Taggart, 1954a).

Table 1.2: K_M values for acyl-coenzyme A substrates and glycine for the human and bovine GLYAT enzymes

	K_M for acyl-coenzyme A (µM)		K_M for glycine (mM)	
	bovine	human	bovine	human
Acetyl-coenzyme A	210			
	850			
Propionyl-coenzyme A	180			
Butyryl-coenzyme A	130	2400 ± 880	50	970 ± 210
Isobutyryl-coenzyme A		5580 ± 1440		2880 ± 1190
Valeryl-coenzyme A	490			
Isovaleryl-coenzyme A	180	672 ± 164		523 ± 206
		124		
2-Methylbutyryl-coenzyme A	110	490 ± 150		1170 ± 290
Tiglyl-coenzyme A	110			
3-methylcrotonyl-coenzyme A	14			
Hexanoyl-coenzyme A		2680 ± 770		1150 ± 210
Octanoyl-coenzyme A		322 ± 87		770 ± 110
		198		
Decanoyl-coenzyme A		2408 ± 887		690 ± 210
Benzoyl-coenzyme A	9	67 ± 5	15	6.5 ± 1
	16	57.9	2 ± 0.02	6.4
	27	13	3	
	15 ± 5		6.2	
	310			
	37 ± 5			
Salicylyl-coenzyme A	31 ± 7	83.7	2 ± 0.4	
			8	
2,4-D-benzoyl-coenzyme A	95		118	
2,4,5-T-benzoyl-coenzyme A	60		100	
Malonyl-coenzyme A	360			

*Values were compiled from (Bartlett & Gompertz, 1974, Gregersen et al., 1986, Gron et al., 1978a, Kelley & Vessey, 1986, Kelley & Vessey, 1993, Kelley & Vessey, 1994, Kolvraa & Gregersen, 1986, Mawal & Qureshi, 1994, Nandi et al., 1979, Schachter & Taggart, 1954a, van der Westhuizen et al., 2000).

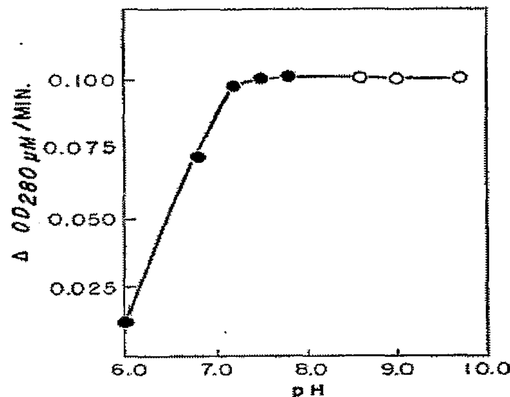

Figure 1.6 The pH dependence of the bovine GLYAT catalysed reaction. The plot shows the change in initial velocity of the bovine GLYAT reaction with change in pH of the reaction mixture. Closed circles represent the use of a potassium phosphate buffer, and open circles the use of a Tris-Cl buffer. (Schachter & Taggart, 1954a).

1.4.3 Substrate specificity

1.4.3.1 The acyl donor substrate

The substrate specificity of GLYAT has been studied by a number of investigators. The results are summarised in Table 1.2 and Figure 1.7. Most investigators determined the K_M and V_{max} values for various substrates, but some only report relative reaction velocities using different substrates (at maximal substrate concentrations). The kinetic constants are not very comparable between studies, as demonstrated by the widely differing values listed in Table 1.2. This can probably be explained by variations in the quality of enzyme preparations used, that the same reaction temperature was not always used, that the same buffer or buffer concentration was not always used, that some researchers added potassium ions to the reaction mixtures and differences in the quality of substrate used (Bartlett & Gompertz, 1974, Gregersen et al., 1986, Gron et al., 1978a, Kelley & Vessey, 1986, Kelley & Vessey, 1993, Kelley & Vessey, 1994, Kolvraa & Gregersen, 1986, Mawal & Qureshi, 1994, Nandi et al., 1979, Schachter & Taggart, 1954a, van der Westhuizen et al., 2000).

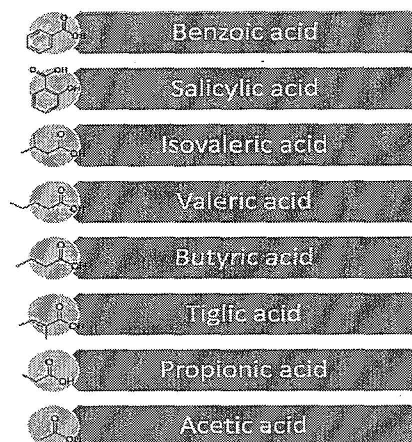

Figure 1.7 Substrate specificity of the glycine N-acyltransferase reaction. The figure approximately represents the substrate specificity of the GLYAT reaction. The acyl groups are arranged in descending order of preference by GLYAT. Thus benzoyl-coenzyme A is one of the best substrates and acetyl-coenzyme A is one of the poorest substrates (Nandi et al., 1979).

A variety of short and medium chain aliphatic acyl groups, as well as aromatic or substituted aromatic groups can be transferred by GLYAT. The preference is for four carbon, five carbon and benzoyl groups, with affinity for the shorter chain substrates acetyl-coenzyme A (C2) and propionyl-coenzyme A (C3) being much lower. Most studies agree to some extent on the order of substrate preference put forward in Figure 1.7, although the lack of consistency in the reaction conditions employed and the values reported limit the resolution with which this order can be defined. For example, if maximal substrate concentrations, above which the reaction rate no longer increases, are used, benzoyl-coenzyme A does not give the highest rate. Butyryl-coenzyme A, isovaleryl-coenzyme A and tiglyl-coenzyme A give rates of 223%, 103% and 115%, relative to benzoyl-coenzyme A at 100%. The concentrations used for these substrates is, however, much higher than for benzoyl-coenzyme A and this limits the interpretability of these results (Nandi et al., 1979).

1.4.3.2 The amino acid substrate

The preferred acyl acceptor substrate for both human and bovine GLYAT is glycine (Brandt et al., 1968, Gregersen et al., 1986, Nandi et al., 1979, Schachter & Taggart, 1954a, van der Westhuizen et al., 2000). The bovine enzyme can also utilise asparagine, glutamine, glutamate, alanine and serine, but at rates orders of magnitude lower than for glycine. The human enzyme can use alanine and glutamate in addition to glycine, again at significantly

lower rates. This phenomenon has been studied in detail. The kinetic parameters for the various amino acids (with benzoyl-coenzyme A as the acyl-donor) for both the human and bovine enzymes, are reported in Table 1.3 (van der Westhuizen et al., 2000).

Table 1.3: K_M values for benzoyl-coenzyme A and various amino acids for the human and bovine GLYAT enzymes

	K_M for acyl-coenzyme A (µM)		**K_M for glycine (mM)**	
Peptide conjugate	**Bovine**	**Human**	**Bovine**	**Human**
Benzoylglycine	160	13	6.2	6.4
Benzoylglutamine	105		353	
Benzoylasparagine	157		129	
Benzoylalanine	41	15.2	1573	997
Benzoylglutamic acid	998		1148	

Values taken from (van der Westhuizen et al., 2000).

Use of these alternative amino acids is likely not to be physiologically significant under normal conditions, judging by their slow formation. Under the conditions of metabolic stress experienced by sufferers of organic acidemias, where there is an overload of acyl-coenzyme A and limiting amounts of glycine, these alternative amino acid conjugations may become more significant (Beliveau & Brusilow, 1987). This is demonstrated by the observation that patients with isovaleric acidemia excrete isovaleryl conjugates of almost all amino acids under acute conditions (excreting for example isovaleryl-alanine, isovaleryl-glutamate, and other such conjugates) (Lehnert, 1981, Lehnert, 1983, Loots et al., 2005, Loots et al., 2007). Patients with urea cycle defects who are treated with benzoic acid excrete the alanine conjugate of benzoic acid in addition to the normal glycine conjugate (Lehnert, 1983, van der Westhuizen et al., 2000). In propionic acidemia, propionyl-serine and propionyl-glutamate are excreted in addition to propionyl-glycine (Fenton et al., 2001).

1.4.4 Inhibition of GLYAT by metal ions, sulfhydryl reagents and reaction products

GLYAT is resistant to treatment with a range of sulfhydryl reagents, including 5,5'-dithiobis(2-nitrobenzoic acid), iodoacetamide, p-chloromercuribenzoate and N-ethylmaleimide. These are chemicals that react readily and irreversibly with sulfhydryl

groups, including the side chains of cysteine residues. These observations suggest that a cysteine residue is neither involved in the catalytic mechanism, nor present in the active site of the enzyme (Nandi et al., 1979, Schachter & Taggart, 1954a).

GLYAT activity is inhibited by high, non-physiological concentrations of divalent metals, including zinc, nickel and magnesium ions. The inhibitory effect is concentration dependent and is more pronounced for nickel and zinc than for magnesium (Nandi et al., 1979). The inhibitory effect may be due to a change in enzyme structure induced by the high divalent cation concentration. For example, a number of enzymes are known to be allosterically regulated by magnesium ions. Another possibility is that the metal ions form coordinate complexes with groups that are critical to enzymatic catalysis, including the side chains of glutamate, aspartate, cysteine, and histidine residues.

Both products of the GLYAT reaction, free coenzyme A and the peptide product, act as inhibitors to the reaction (Schachter & Taggart, 1954b, van der Westhuizen et al., 2000). Since the free energy change for the GLYAT reaction is large, approximately 8 000 calories per mole, competitive inhibition by low concentrations of products is not likely to be the result of a significant change in equilibrium position of the reaction. It clearly results from the active site of the enzyme being occupied by the inhibitors, competitive inhibition at the molecular level. Inhibition by coenzyme A is independent of the nature of the sulfhydryl group. Treatment of coenzyme A with iodoacetate prior to analysis results in no significant difference in inhibition (Schachter & Taggart, 1954b). This implies that it is binding of the coenzyme A to the active site and not a chemical mechanism that is responsible for inhibition.

1.4.5 The GLYAT gene and its splice variants

The bovine GLYAT gene is located on chromosome 15 and has six exons. Post-transcriptional splicing results in a mature mRNA of 1095 nt. Translation of this mRNA results in a protein of 295 amino acids. The gene for human GLYAT is located on chromosome 11 and is approximately 23 kb in size. GLYATL1 and GLYATL2, two genes with appreciable sequence similarity to human GLYAT, but still ill defined function, are

found proximal to the GLYAT gene on chromosome 11 (Zhang et al., 2007). The human GLYAT gene has six exons, which are transcribed to form a mature mRNA of 2052 nt. In humans and chimpanzees, there are two mRNA species produced by alternative splicing of the primary transcript. The significance of the splice variants is not understood. The longer mRNA (splice variant 1), similar to that observed in the other species, includes exons one to six and encodes a protein 296 amino acids in length, called isoform a. The other mRNA (splice variant 2) is shorter because exon 6 is excluded, and a truncated protein of 163 amino acids, called isoform b, is produced (ENSEMBL, January 2009). This is demonstrated in Figure 1.8.

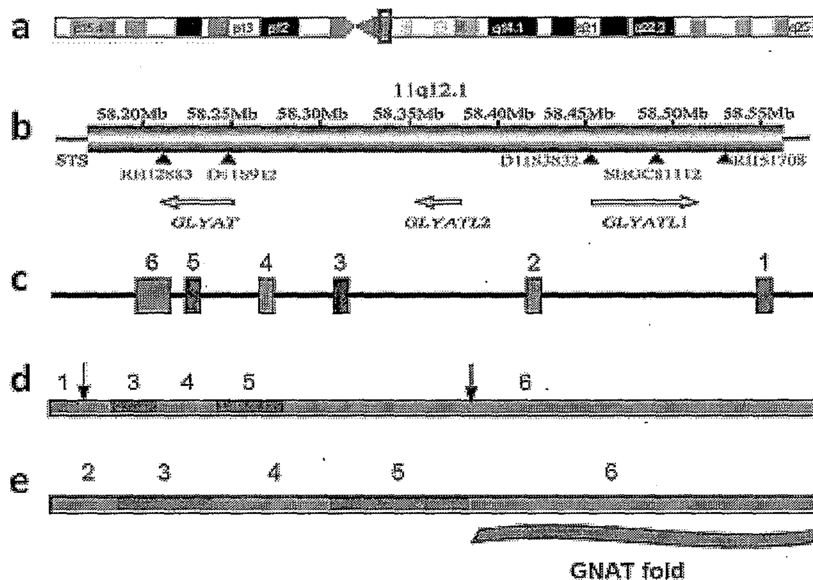

Figure 1.8 The human GLYAT gene, splice variants and protein isoforms. The diagram demonstrates the relationship between the human GLYAT gene, its splice variants and the resulting isoforms of the enzyme: a) the GLYAT gene is found on chromosome 11, on the q12 band; b) an enlarged view of the 11q12.1 band of chromosome, showing the GLYAT gene and the related genes GLYATL1 and GLYATL2; c) the primary transcript of the GLYAT gene, showing the six exons; d) mRNA splice variant 1, including all six exons. Splice variant 2 differs in that exon 6 is excluded. The arrows indicate approximately where the translation for isoform a starts and ends; e) translation of splice variant 1 to form GLYAT isoform a. Translation of splice variant 2 results in GLYAT isoform b, which is half the size of isoform a, excluding the GNAT domain encoded by exon 6. ENSEMBL, January 2009, and (Zhang et al., 2007).

1.4.6 Molecular weight and post-translational modification of the GLYAT enzymes

The literature reports experimentally determined sizes for bovine GLYAT, including approximately 33 kDa, 35.5 kDa and 36 kDa (Kelley & Vessey, 1993, Nandi et al., 1979,

van der Westhuizen et.al., 2000). The size calculated from the amino acid sequence (gi 29135315), obtained by translation of the nucleotide sequence (nm 177513.2), is 33.864 kDa, which compares well to the experimentally estimated sizes. The values reported for the size of human GLYAT are approximately 27 kDa and 30 kDa (Nandi et al., 1979, van der Westhuizen et al., 2000). The predicted size of this protein, based on its sequence (gi 11038137), is 33.885 kDa. The discrepancy between the predicted and experimentally estimated sizes is not understood, but may be attributed to inaccuracy of the size estimations (use of different size markers, for example), post-translational modifications (such as post-translational cleavage), or partial protein degradation. This matter has, however, not received much attention in the literature.

1.4.6.1 Cleavage of the mitochondrial signal peptide

Nuclear encoded proteins that function in mitochondria are targeted to these organelles by means of a signal peptide at the amino-terminus. Because it is in most cases cleaved upon entry into the mitochondria, it is often called the mitochondrial leader-peptide. It is important to consider the nature of the signal peptide of GLYAT, as its cleavage would have implications for the size of the mature protein and perhaps for enzyme activity.

No literature has discussed this topic directly, but enough information is available to make some deductions. Since GLYAT functions in the mitochondrial matrix, it is expected to have a mitochondrion-targeted signal peptide (Brandt et al., 1968, Gatley & Sherratt, 1977, Kelley & Vessey, 1994, Kolvraa & Gregersen, 1986, Nandi et al., 1979, Schachter & Taggart, 1954a).

It appears, however, that the signal peptide of GLYAT (bovine and human) is not cleaved upon entry into the mitochondrion. The N-terminal amino acid sequence obtained on translation of the bovine GLYAT nucleotide sequence (nm 177513.2) is shown in Figure 1.9. The protein isolated from bovine liver mitochondria has been subjected to Edman sequencing (Kelley & Vessey, 1992, van der Westhuizen et al., 2000), and is also shown in Figure 1.9. Since the two sequences are the same and the protein investigated was isolated from bovine liver mitochondria, it can be concluded that the leader peptide is probably not cleaved upon entry into the mitochondria.

The human GLYAT N-terminal sequence, expected from translation of the transcript sequence (np 964011.2) is shown in Figure 1.10. Human GLYAT has been shown to be acetylated on the lysine residue in this sequence, as demonstrated in Figure 1.10. This modification was discovered in a large study wherein GLYAT was only one of several lysine-acetylated proteins identified. The proteins were isolated by immunoprecipitation of N-acetylated lysine residues and then identified by mass spectrometry (Kim et al., 2006). If the assumption is made that the GLYAT identified was the mature form of the protein, then it follows that the N-terminal residues are all intact in the mature protein and that the mitochondrial signal peptide is not cleaved. It may of course be that the protein is cleaved, but that only the immature, un-cleaved, lysine-acetylated form was detected in this study.

M F L L Q G A Q M L Q M L E K

M F L L Q G A Q M L Q M L E K S L R K S L P M S L . . .

Figure 1.9 Edman sequencing of the N-terminus of the bovine GLYAT enzyme. The upper sequence represents that obtained upon Edman sequencing of the bovine GLYAT enzyme, purified from bovine liver mitochondria. The lower sequence represents that obtained on translation of the bovine GLYAT mRNA sequence (Kelley & Vessey, 1994).

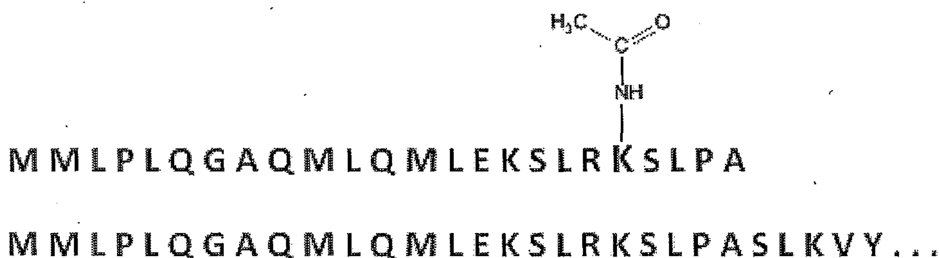

Figure 1.10 Acetylation of a lysine residue near the N-terminus of human GLYAT. The upper sequence represents the sequence obtained on mass spectrometric sequencing of a lysine-acetylated peptide fragment of human GLYAT. The lower sequence represents the translation of the GLYAT mRNA (Kim et al., 2006).

1.4.6.2 Other post-translational modifications

A near homogenous preparation of GLYAT from bovine liver mitochondria displayed anomalous electrophoretic behaviour in polyacrylamide-disc gel electrophoresis. The

protein separated into three discrete bands, with enzymatic activity in each band. The highest protein content and enzymatic activity were both found in the middle band. This behaviour was not investigated further (Nandi et al., 1979). Whether these different bands represent post-translational modifications of the enzyme is not understood.

As mentioned above, the human enzyme is acetylated at lysine 20 in the sequence SLRKSLPASL. Lysine acetylation plays various roles in the biology of proteins. It influences, amongst other things, the sub-cellular trafficking and mitochondrial entry of proteins and can influence enzyme activity or protein function (Kim et al., 2006). The lysine residue in question is not present in the bovine enzyme and the significance of this modification specifically for GLYAT is not understood.

1.5 The GNAT superfamily of N-acyltransferases

GLYAT is a member of the GNAT superfamily, an annotation made on the basis of sequence homology (ENSEMBL January 2009). This is supported by the similarity of the reactions catalysed by GLYAT and the GNAT enzymes and the similar reaction kinetics. The superfamily will be briefly reviewed, focussing on aspects that are important for understanding the biology of GLYAT.

1.5.1 Structural and functional conservation in the superfamily

The GNAT superfamily is one of the largest superfamilies of proteins identified to date. It has more than 10 000 members. These proteins are found everywhere in nature and all share a universally conserved acyltransferase activity, with an acyl-coenzyme A being used to acylate an amine substrate. The vast majority of GNAT enzymes are acetyltransferases, enzymes which use acetyl-coenzyme A as the acyl donor. However, there are members that use other acyl donors, such as protein N-myristoyltransferase, which uses myristoyl-coenzyme A, and GLYAT, which uses several acyl-coenzyme A substrates (Dyda et al., 2000, Vetting et al., 2005). A large variety of acyl acceptor substrates are found, but every enzyme is remarkably selective for its specific substrate.

Examples of GNAT enzymes are aminoglycoside N-acetyltransferases, which are medically important as they impart antibiotic resistance to microorganisms, and histone N-acetyltransferases, which are important in eukaryotic gene regulation (Berndsen et al., 2007, Berndsen & Denu, 2005, Berndsen & Denu, 2008, Bewley et al., 2006, Dyda et al., 2000, He et al., 2003, Marmorstein & Roth, 2001, Scheibner et al., 2002, Tanner et al., 2000, Vetting et al., 2005, Vetting et al., 2002, Watson et al., 2002).

There is as little as 3% sequence conservation between different members of the superfamily. Despite the lack of significant sequence conservation, though, these proteins display remarkable structural conservation, as demonstrated in Figure 1.11 (Dyda et al., 2000, Vetting et al., 2005). This is an excellent demonstration of the principle that in evolution, structure and function are better conserved than sequence (Rost, 1999). The structures solved to date, when superimposed, reveal a conserved GNAT fold of approximately 120 amino acids, as demonstrated in Figures 1.11 and 1.12. Binding of the acyl-coenzyme A substrate induces dramatic changes in conformation of parts of the fold, especially the α1-α2 motif. This alters the binding site geometry in a way that is essential for binding of the amine substrate. It is interesting to note that the conformation of the bound coenzyme A substrate is as conserved as the GNAT structures themselves, as demonstrated in Figure 1.13. The relative orientation of the amine nitrogen and the coenzyme A thioester also appears to be well conserved. Binding of the acyl-coenzyme A substrate is mediated mostly by contacts with the protein backbone, not amino acid side chains. This may account to some extent for the lack of sequence conservation between these enzymes (Dyda et al., 2000, Vetting et al., 2005).

1.5.2 Reaction kinetics and catalytic mechanisms in the GNAT superfamily

The GNAT enzymes studied to date all use a compulsory order ternary complex reaction mechanism. The acyl-coenzyme A substrate binds first, inducing structural changes that enable binding of the amine substrate, to form a ternary complex. Direct acyl transfer then occurs, followed by release of the products, the rate limiting step. The catalytic mechanisms of these enzymes are very similar, and involve simple acid-base catalysis

(Dyda et al., 2000, Vetting et al., 2005). Because this is important for discussion of the catalytic mechanism of GLYAT, this topic will now be presented in more detail.

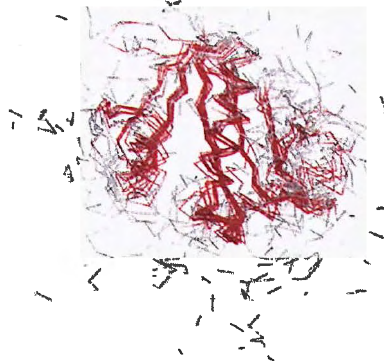

Figure 1.11 Superposition of the crystal structures of 15 GNAT enzymes. The core GNAT fold, approximately 120 amino acids in length, is demonstrated by the superposition of 15 GNAT structures. The fold is represented by the red core in the figure, which represents the positions where divergence between the 15 structures is minimal. (Vetting et al., 2005).

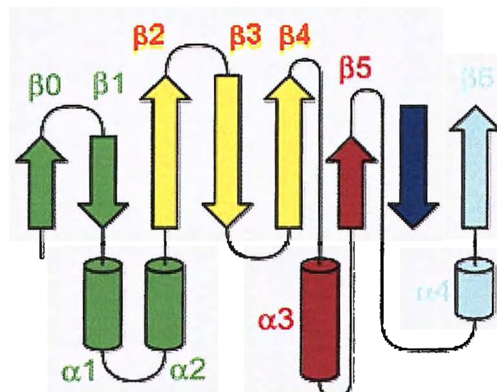

Figure 1.12 Topology of the core GNAT fold. This diagram demonstrates the topology, or order of secondary structure elements, in the GNAT fold. The β0 strand is not as well conserved as the rest of the fold. Most variation in the GNAT structures is found in this N-terminal region. The unlabelled strand between β5 and β6 is also not conserved. (Vetting et al., 2005).

Figure 1.13 Conservation of the conformation of bound acyl-coenzyme A between different members of the GNAT superfamily. The conformation of binding is most conserved for the pantetheine arm of coenzyme A. The circle indicates the orientation of the thioester bond, which is very well conserved. The conformation of the adenosine moiety, which is involved in binding of the coenzyme A, but not in catalysis, is most variable, as seen on the right. (Dyda et al., 2000).

1.5.2.1 The general reaction mechanism of GNAT enzymes

There are two reaction mechanisms commonly employed by acyltransferase enzymes. One is a ping-pong mechanism where the acyl donor binds first and the acyl group is transiently transferred to a cysteine residue to form a covalent acyl-enzyme intermediate. The first product (coenzyme A) is then released before the second substrate (the amine) is bound. The acyl group is then transferred to the amino group of the acceptor substrate to form the second product, as demonstrated in Figure 1.14. As demonstrated in Figure 1.14, the catalytic mechanism usually involves a conserved catalytic triad composed of a cysteine, a histidine, and an aspartic acid residue. The histidine is polarised by the aspartate residue and interacts with the cysteine to form a thiolate-imidazolium couple. The thiolate is thus very reactive and can attack the thioester of the acyl-coenzyme A to form the acyl-enzyme intermediate. The amine substrate then binds, followed by nucleophilic attack on the new thioester, resulting in acyl transfer to the amino group (Berndsen & Denu, 2005, Dyda et al., 2000, Vetting et al., 2005, Wang et al., 2005). To date, not one of the acyltransferases belonging to the GNAT superfamily which have been studied uses such a mechanism (Berndsen et al., 2007).

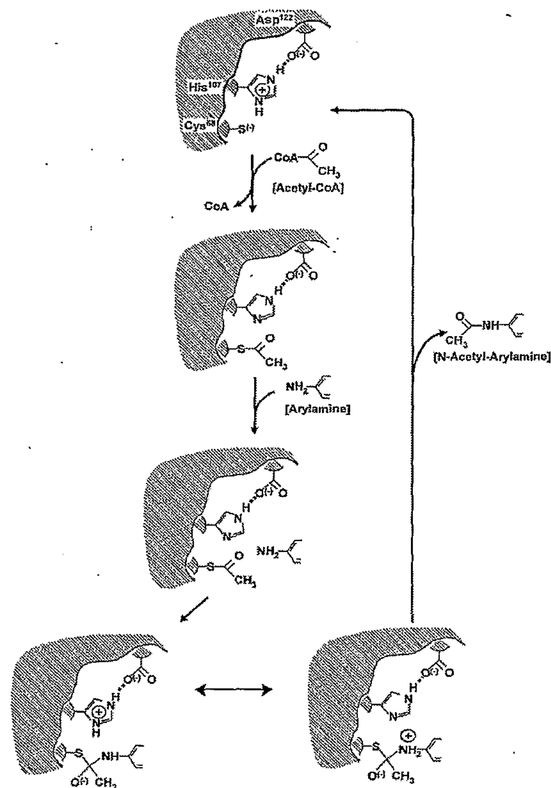

Figure 1.14 The ping-pong mechanism of acyl transfer by hamster NAT2. The figure demonstrates the involvement of a catalytic triad, composed of an aspartate residue, a cysteine residue and a histidine residue. The aspartate polarises the histidine imidazolium ring, which in turn polarises the cysteine sulfhydryl, to form a thiolate-imidazolium couple. The thiolate, a strong nucleophile, attacks the acyl-coenzyme A and the acyl group is transferred to the cysteine residue. The amine substrate binds next, upon which the acyl group is transferred from the enzyme cysteine to the substrate amino group, forming the acetylated product. (Wang et al., 2005).

All GNAT enzymes studied to date employ the alternative strategy, the compulsory order ternary complex mechanism, depicted in Figure 1.15. The acyl-coenzyme A or acyl donor substrate binds first, followed by the amine, to form a ternary complex. Direct acyl transfer then takes place as the nucleophilic amino group of the amine attacks the electrophilic carbonyl group of the acyl-coenzyme A thioester. Kinetic, mutagenic and structural evidence support this mechanism. The insensitivity of GNAT enzymes to cysteine-modifying reagents and the inability to observe covalent substrate-enzyme intermediates argue against a ping-pong mechanism (Berndsen et al., 2007, Berndsen & Denu, 2005, Berndsen & Denu, 2008, Dyda et al., 2000, Tanner et al., 1999, Vetting et al., 2005).

One MYST family histone acetyltransferase, Esa1 from Saccharomyces cerevisiae, was reported to have an enzyme acyl-cysteine intermediate, casting doubt on the universality of

GNAT mechanisms (the MYST family belongs to the GNAT superfamily). However, it was later demonstrated that the direct transfer mechanism is employed, and that the covalent intermediate was an artefact (mutagenesis studies showed that the cysteine involved was not catalytically important and identified a catalytic glutamate residue) (Berndsen et al., 2007). It would then appear that the mechanism employed by the GNAT superfamily of acyltransferases is indeed universally conserved.

1.5.2.2 The general base catalyst

In the direct nucleophilic attack mechanism, the amino group of the amine substrate needs to be in the unprotonated form. The protonated and positively charged form is not a nucleophile, as demonstrated in Figure 1.15. Having pKa values near 10, most primary amines are protonated at physiological pH, and can thus not participate in an acyl-transfer mechanism. This suggests that a GNAT enzyme should provide a general base residue, capable of deprotonating the amino group, as part of its catalytic mechanism (Dyda et al., 2000, Vetting et al., 2005).

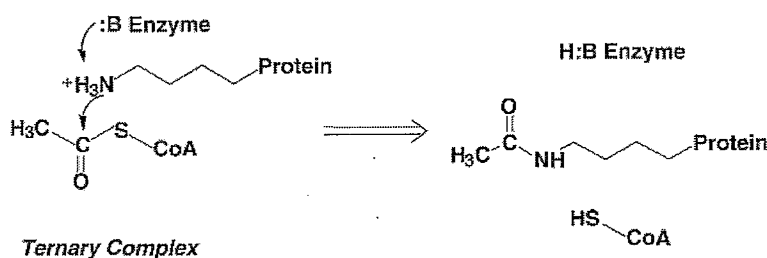

Figure 1.15 The ternary complex, direct acyl transfer mechanism. In this mechanism, both substrates are bound to the enzyme simultaneously, forming a ternary complex. The nucleophilic amine substrate then directly attacks the thioester of the acyl-coenzyme A, upon which the acyl group is transferred to the amino group. The amino group must be deprotonated (NH_2) in order to behave as a nucleophile. (Berndsen & Denu, 2005).

Such general base catalysts have been investigated, supporting the proposed catalytic mechanism. In yeast GCN5 (a histone acetyltransferase from *S. cerevisiae*) a conserved glutamate residue, E173, has been shown to be crucial for enzyme function. Mutation of this residue to a glutamine residue results in a 320 fold decrease in activity (Tanner et al., 2000, Tanner et al., 1999). Similarly, in Esa1, a MYST family histone acetyltransferase

from *S. cerevisiae*, the glutamate residue E338 is responsible for catalysis. When this residue is mutated to a glutamine residue, a 200 fold decrease in activity is observed (Berndsen et al., 2007). These two residues are in the same position in a structural superposition. In several GNAT structures there is an equivalent residue. It is, however, not absolutely necessary for the residue to occupy exactly the same position, as it appears that it only needs to be in the vicinity of the thioester and amine (Vetting et al., 2005). Examples of structurally identified catalytic residues include E92 of human spermine/spermidine N-acetyltransferase (Hegde et al., 2007), E92 of the *P. syringae* tabtoxin resistance protein (He et al., 2003) and several others (Bewley et al., 2006, Marmorstein & Roth, 2001, Scheibner et al., 2002, Watson et al., 2002).

In serotonin N-acetyltransferase (SNAT) a conserved histidine residue, H120, is situated in a similar position. Mutation of this residue of the ovine enzyme to a glutamine residue does not significantly influence the rate of the reaction (although it does increase the Km for the amine substrate). There is also another conserved histidine residue, H122, spatially close to H120 in the crystal structure. Mutation of both histidine residues to glutamine residues results in a 60 fold decrease in reaction rate. This suggests that H120 may be the catalytic base, but that its mutation can be rescued by H122 (Hickman et al., 1999, Scheibner et al., 2002).

It seems then that no structurally conserved residue will serve as the hallmark of catalysis, but that some general base catalyst is used in most GNAT enzymes (Vetting et al., 2005). In several enzymes the catalytic residue does, however, align with the E173 residue of yGCN5. This is supported by the observation that this glutamate residue is conserved in 36% of GNAT protein sequences. An aspartate residue that aligns to Y212 of yGCN5 (see above) is conserved in 24% of GNAT sequences, and this represents another major position for potential catalytic residues (He et al., 2003).

Because of the nature of catalysis provided by the general base residue, the reaction should be predicted to be pH dependent. This is in fact what is observed for the GNAT enzymes. The mutants that have been studied are inactive at low or physiological pH, as would be expected. At these pH levels the amine substrates are protonated and cannot participate in nucleophilic attack without a general base catalyst to remove the protons.

When pH is increased to high levels, more and more amine becomes deprotonated and capable of nucleophilic attack. It is then less crucial to have a catalytic residue to remove protons and the activity of the mutant increases, approaching wild type activity at very high pH (Dyda et al., 2000, Hegde et al., 2007, Scheibner et al., 2002, Tanner et al., 1999, Vetting et al., 2002). This principle is demonstrated for yGCN5 in Figure 1.16.

1.5.2.3 The general acid catalyst

In the direct transfer mechanism, nucleophilic attack is followed by the release of a coenzyme A thiolate ion (coenzyme A with an unprotonated sulfhydryl group). A thiolate ion is not a very good leaving group and has to be protonated as soon as possible. It may be expected that a GNAT enzyme would provide a general acid catalyst to protonate the leaving thiolate ion as part of the catalytic mechanism. A protonated sulfhydryl is a better leaving group and the reaction will thus be more favourable and more rapid. In several of the GNAT structures determined to date there is a tyrosine residue approximately 3 Å from the sulfur atom, which could act as the proton donor. In fact, this tyrosine is conserved in 62% of GNAT sequences (He et al., 2003). Mutation of this residue to a phenylalanine in SNAT results in almost complete loss of catalytic activity, as shown in Figure 1.17 (Scheibner et al., 2002). It has been proposed that the tyrosine may also be important for correct substrate positioning, by forming hydrogen bonds with the thioester (Vetting et al., 2005).

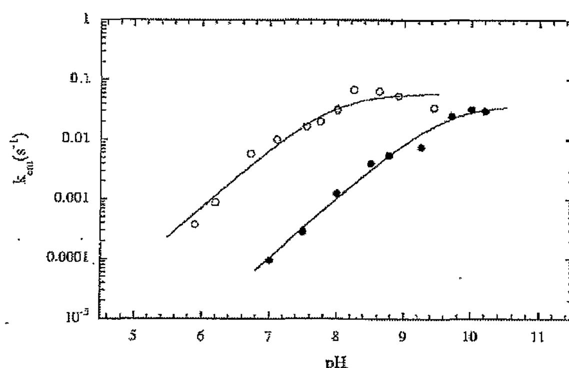

Figure 1.16 The dependence of the reaction rate of wild type and mutant yGCN5 acetyltransferases. The plot with the empty circles represents the wild type yGCN5 enzyme. The full circles represent the E173Q mutant, where the catalytic residue is changed. This mutant has lower activity than the wild type at low pH, but as pH increases, the rate of the mutant approaches that of the wild type enzyme. (Tanner et al., 1999).

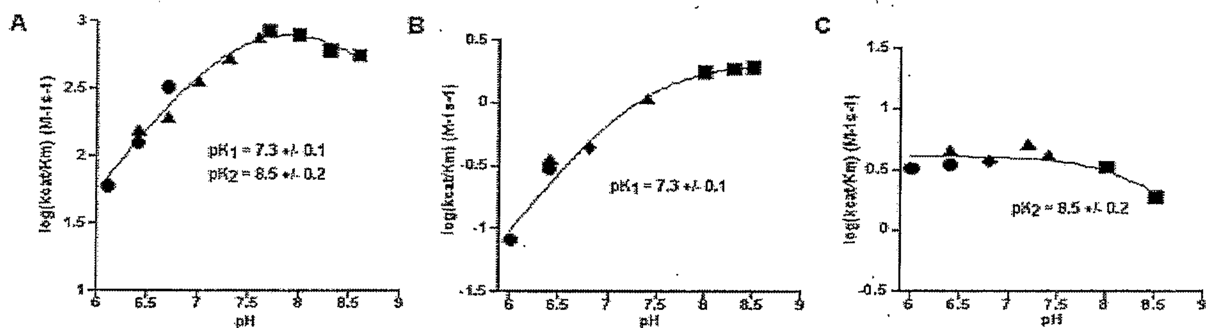

Figure 1.17 The pH dependence of the reaction rate of wild type and mutants of serotonin N-acetyltransferase. The plots show the dependence of reaction rate on the pH of the reaction mixture: a) the wild type enzyme, with its general acid- and base catalysts, has an optimal pH, from which the rate decreases as the pH is either increased or decreased; b) the general acid catalyst, a tyrosine residue, has been mutated to a phenylalanine. The rate now no longer responds to an increase in pH in a negative manner; c) the general base catalysts, two histidine residues, have been mutated to glutamine residues. The rate now no longer responds negatively to a decrease in pH. (Scheibner et al., 2002).

As for the base catalysts, the tyrosine residue is not universally conserved, suggesting that a general acid catalyst is not always needed. Some enzymes either have another residue for protonating the thiolate or do not need one. This is supported by pH dependence studies that argue against the presence of a general acid catalyst for some members of the family, such as yGCN5 (as demonstrated in Figure 1.16) (Tanner et al., 1999).

Figure 1.18 (next page) shows the mechanism of spermine/spermidine N-acetyltransferase (SSAT), which utilises both a catalytic acid (tyrosine) and base (glutamate). This mechanism is representative of that employed by most members of the GNAT superfamily, although some members will not use both the acid and the base catalysts (Dyda et al., 2000, Hegde et al., 2007, Vetting et al., 2005).

1.6 Recombinant protein expression in *Escherichia coli*

1.6.1 General principles of recombinant protein expression

The gram negative bacterium *Escherichia coli* is the most popular bacterial system for heterologous protein expression, because of its rapid growth to high culture densities in inexpensive media (Baneyx, 1999). Several proteins have been successfully expressed in

soluble and biologically active form using this host and the multitude of expression vectors available (Andersen & Krummen, 2002, Baneyx, 1999). Many complex proteins, often of mammalian origin, can not at present be successfully expressed in this host. Insolubility and biological inactivity of the recombinant proteins is a frequent problem (Sorensen & Mortensen, 2005). In these cases mammalian expression systems, especially the Chinese hamster ovary (CHO) expression system, have been employed with considerable success. Compared to bacterial systems, however, these systems are not only inconvenient (slow growth, low densities and contamination), but also expensive in terms of the growth medium and culture conditions used. For this reason, efforts to enhance the efficiency of bacterial systems, both with regards to the expression vectors used and mutant bacterial strains with altered properties, are constantly being undertaken (Baneyx, 1999).

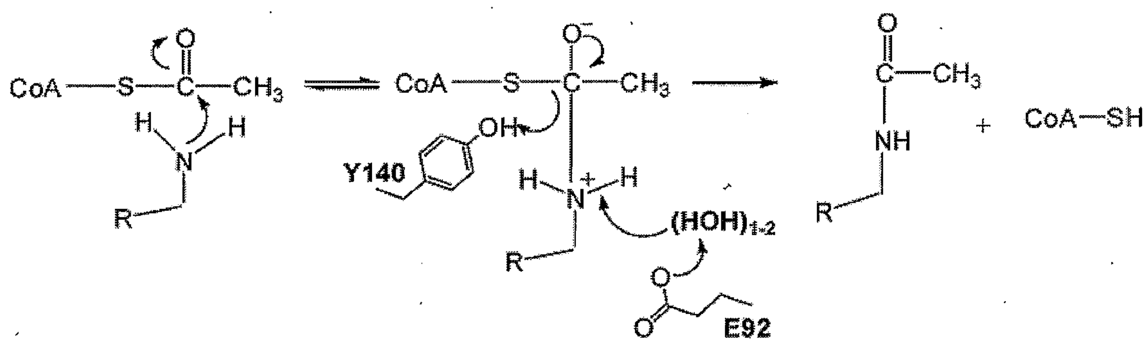

Figure 1.18 The catalytic mechanism of spermine/spermidine N-acetyltransferase. The catalytic mechanism of SSAT, a member of the GNAT superfamily, is demonstrated. Two catalytic residues are involved. E92 is responsible for proton removal from the amino group during the catalysis. Y140 is responsible for proton donation to the thiolate intermediate. (Hegde et al., 2007).

Often there are no problems with levels of transcription and translation, but the expressed protein is insoluble or incorrectly folded and thus biologically inactive (Baneyx, 1999, Sorensen & Mortensen, 2005). There are several strategies to moderate this problem (Hannig & Makrides, 1998). The more complicated strategies involve complex engineering of culture conditions and cellular physiology to optimise recombinant protein expression. Proteins may also be targeted to the periplasm or extra-cellular space, where advantages include the oxidising environment and the decreased toxic load on the cytoplasm. By expressing the protein as a fusion to a "solubilising partner" such as maltose binding protein, thioredoxin, or glutathione S-transferase, the yield of soluble and biologically active protein can be increased. This is presumably because the fusion tag rapidly folds into the

correct conformation, and this aids the correct folding of the rest of the protein. The problem with this approach, however, is the low yields obtained upon cleavage of the fusion tag and the cost of the proteases needed to do so (Baneyx, 1999). Two simpler strategies seem to be more generally successful. One is co-expression of the target protein with molecular chaperones (Nishihara et al., 1998, Nishihara et al., 2000, Walter & Buchner, 2002), and the other is expression at low temperature (Vasina & Baneyx, 1996).

1.6.2 Expression at low temperature and co-expression of chaperone proteins

When culture temperature is decreased to between 15 °C and 30 °C, the rate of protein expression decreases, accompanied by decreased synthesis of DNA and RNA, and other physiological changes. Part of this response is mediated by loss of normal promoter (e.g. tac) activity at these temperatures and an increase in transcription from a set of cold-shock induced promoters (such as cspA) (Vasina & Baneyx, 1996). Cold shock expression has advantages for the expression of aggregation prone proteins. Firstly, the rate of host protein synthesis decreases, making the transcriptional and translational machinery available for synthesis of recombinant proteins. Recombinant proteins tend to be less prone to aggregation and more soluble at these low temperatures. Secondly, heat shock proteases that are normally induced by over expression of foreign proteins are not expressed at high levels, resulting in decreased proteolytic degradation of the recombinant proteins (Vasina & Baneyx, 1996). This principle has been commercialised in the form of the pCold expression vectors from the company Takara. The vectors contain the cspA promoter region for transcriptional initiation at low temperatures (usually 15 °C). Downstream of this promoter region a lac operator is inserted to ensure tight regulation and inducible expression of the target protein.

An alternative strategy is the use of chaperone co-expression. The information a protein needs in order to fold into its correct three dimensional structure is contained in its primary amino acid sequence, provided it folds in the right environment. It turns out that the environment is not a simple solvent inside the cell, but involves several interactions with other proteins and small molecules. Chaperones are proteins which are specifically involved in protein folding and refolding. The chemistry of chaperone action is complex and

will not be reviewed here, but the important principle is that they either prevent premature folding, or unfold and refold poorly folded proteins (Baneyx, 1999, Nishihara et al., 1998, Nishihara et al., 2000, Walter & Buchner, 2002).

A set of plasmids containing different combinations of various chaperone proteins has been commercialised by Takara. The Takara pGTf2 plasmid used in this study contains the GroEL, GroES, and trigger factor (TF) chaperones. Trigger factor associates with ribosomes and prevents premature folding of newly translated polypeptides, by associating with stretches of hydrophobic amino acids. This prevents protein aggregation and provides time for proper folding. GroEL (as a homo-tetradecamer) has the effect of binding to and unfolding misfolded proteins, such as proteins stuck in local folding energy minima. The GroEL-peptide complex then associates with a heptameric GroES complex, upon which the protein is released into the lumen of the barrel formed by GroEL. Here it is capable of folding without risk of aggregation, as it is at "infinite dilution." Finally, ATP hydrolysis powers the release of the hopefully correctly folded protein from the chaperone complex (Nishihara et al., 2000, Walter & Buchner, 2002). These chaperones are all expressed from a single plasmid, pGTf2. A plasmid containing the GrpE, DnaK, and DnaJ chaperones is also available, but the literature suggests that over expression of either set of chaperones should be sufficient. In addition, the DnaK and DnaJ chaperones seem to have less capacity to refold improperly folded proteins (Nishihara et al., 1998). A chloramphenicol resistance gene on the chaperone expression plasmid enables simultaneous selection for this plasmid and an expression plasmid conferring resistance to another antibiotic such as ampicillin or kanamycin. The coexpression of these chaperone proteins has been demonstrated to be effective in increasing the yield of soluble and biologically active protein in several cases. However, there is no way of predicting whether the chaperones will contribute to the correct folding of a specific protein (Walter & Buchner, 2002).

1.6.3 Histidine tag affinity purification of proteins

To facilitate purification of a recombinant protein, a hexa-histidine tag may be fused to it, by engineering the coding sequence or cloning into a vector already encoding such a tag. Upon passage through a resin with nickel ions immobilised onto it, the histidine tags bind

tightly by forming coordinate bonds with the nickel ions, while most other proteins can be washed off. The tagged protein is then eluted with a buffer containing a high concentration of imidazole, which displaces the coordinate bonds between the histidine residues and nickel ions (Porath et al., 1975). In this study, the Protino Ni-TED columns from Machery Nagel will be used. The immobilised nickel in this resin has only one unoccupied coordination site, which gives lower yields but better purification than the Ni-IDA resin, which has four, and is thus prone to non-specific binding. This is demonstrated in Figure 1.19.

Occasionally the structure of a protein is such that the histidine tag is not accessible to the immobilised nickel on the column, being hidden in a cleft on the protein surface for example. To solve this problem the tag can be designed with a flexible serine-glycine linker spacing the protein from the histidine residues. This ensures that the tag is free to bind to the column, enhancing purification (Loughran et al., 2006). The protein may also benefit in terms of stability, as it has more freedom to move and is not forced into contact with the resin.

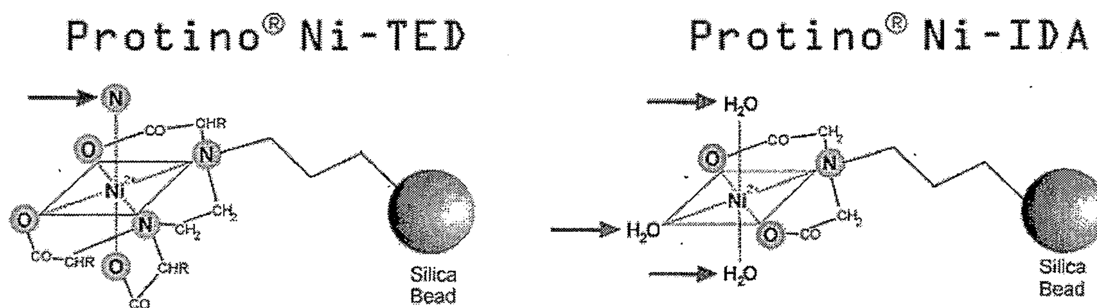

Figure 1.19 The difference between Protino Ni-TED and Protino Ni-IDA resins. Whereas Protino Ni-IDA has three unoccupied coordination sites for the binding of histidine residues, Protino Ni-TED has only one unoccupied coordination site. The effect is that the resin has a lower binding capacity, but much less non-specific binding to the resin occurs. Figure taken from the Protino Ni-TED product manual.

1.7 Site-directed mutagenesis

Site-directed mutagenesis is a technique used to directly study the importance of one or a group of amino acids in the function of an enzyme or protein (Carter, 1986). Using recombinant DNA technology a nucleotide change can be introduced into the coding

sequence of the gene of interest, causing an amino acid change in the translated protein. If a certain residue is thought to be functionally important, a reasonable hypothesis would be that changing the residue for another will alter or abolish enzyme function. This hypothesis can be experimentally tested by generating a mutant recombinant enzyme and investigating its biochemical properties.

An amino acid residue in a protein can have either or both structural and functional significance. Thus, when a mutation renders a protein dysfunctional, care should be taken not to overestimate the catalytic importance of that residue. The loss of function may very well be a consequence of altered protein folding, stability, or solubility, and not of altered catalysis as such. For this reason a few guidelines on designing appropriate mutations and for analysing their effect on protein function should be followed. Firstly, amino acids should, if possible, be replaced by residues that are chemically distinct but structurally similar. In following this rule, it can with greater certainty be said that a change in activity is due to loss of the specific chemistry of the residue, rather than loss of protein structure or stability. Secondly, the hypothesis can be tested by changing the relevant residue to another which is structurally distinct but chemically similar. This type of mutant would further demonstrate that it is the chemical nature of the amino acid, and not its structure, that is important for catalysis. The last guideline is to alter the conditions under which the assay is carried out in ways that will augment the effect of the mutation. For example, the effect of mutations that influence acid-base catalysis will depend on the pH of the reaction mixture; loss of a base catalyst can to some extent be compensated for by an increase in pH, and loss of an acid catalyst by a decrease in pH (Scheibner et al., 2002, Tanner et al., 1999). If these general guidelines are followed, site directed mutagenesis can illuminate a reaction mechanism to an extent that is hard to achieve using any other approach.

In this study the megaprimer method of site directed mutagenesis was used, and the details of this method are described in Chapter 3 (Aiyar & Leis, 1993, Carter, 1986). Site directed mutagenesis has been used to study the catalytic mechanisms of several enzymes, including the GNAT enzymes discussed above. Site directed mutagenesis can also be used in a process of rational design of novel enzyme functions, with the design of Subtilisin being a classic example (Saeki et al., 2007).

1.8 Problem formulation and aims of this study

As mentioned in the introduction, it is our long term goal to generate a recombinant therapeutic GLYAT enzyme, for the treatment of propionic acidemia and other organic acidemias. In order to achieve this goal, a recombinant enzyme expression system is needed to facilitate investigation of the mechanisms of substrate binding and catalysis. A system for the expression of a soluble and enzymatically active human GLYAT (EC 2.3.1.13) has not been reported in the literature. Several previous projects in our laboratory have failed to express enzymatically active human GLYAT in bacterial expression systems (DA Grundling and R van der Sluis, unpublished results).

Sometimes the use of the same enzyme, from another species, can result in expression of enzymatically active enzyme. It was recently shown in our laboratory that bovine GLYAT, expressed at 15 °C from pColdIII with chaperone co-expression, is partially soluble and enzymatically active (M Snyders, unpublished results). Although there are differences between the human and bovine GLYAT enzymes, the general pattern of substrate specificity is similar (Bartlett & Gompertz, 1974, Kelley & Vessey, 1994, Mawal & Qureshi, 1994, Nandi et al., 1979, van der Westhuizen et al., 2000). For this reason, investigation of the molecular mechanisms of substrate binding and catalysis can be initiated using a recombinant bovine GLYAT enzyme. When a system for the expression of enzymatically active human GLYAT is developed, the knowledge gained by studying the bovine enzyme could aid design of novel variants of human GLYAT.

The aims for this project were the following:

1) The first objective of the present study was to express and purify a recombinant bovine GLYAT with a C-terminal histidine tag. This recombinant GLYAT enzyme was then to be subjected to kinetic characterisation, and compared to the GLYAT enzyme extracted from bovine liver.

2) The second objective was to use the recombinant bovine GLYAT expression system to investigate the catalytic mechanism of the bovine GLYAT enzyme by means of site-directed mutagenesis.

3) The third objective of this study was to implement a system for the cost effective in-house synthesis of benzoyl-coenzyme A, which is used as a reagent in GLYAT enzyme assays.