NORTH-WEST UNIVERSITY
YUNIBESITI YA BOKONE-BOPHIRIMA
NOORDWES-UNIVERSITEIT
POTCHEFSTROOMKAMPUS

Molecular characterisation of glycine-N-acyltransferase from two primates: the vervet monkey and the chacma baboon

by

Cornelius Mthiuzimele Mahlanza Hons. B.Sc (Biochemistry)

Dissertation submitted in partial fulfilment of the requirements for an

Masters degree in Biochemistry

Division for Biochemistry, School of Physical and Chemical Sciences, North-West University, Potchefstroom Campus, Potchefstroom, 2520, South Africa

Supervisor: Prof. A.A. van Dijk

March 2011

Acknowledgements

I would like to thank God for the strength he gave me throughout this project.

I would like to thank my study leader (Prof Albie van Dijk) for her efforts in supervising this project and Mr Jaco Wentzel for helping with the translation of the abstract from English to Afrikaans.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

I thank NRF (FA grant: FA200503170001), DST-NRF BioPAD project (BPP007) and the North-West University Centre for Human Metabonomics (NWU-CHM) for their financial assistance.

I hereby appreciate all the help that I received from my fellow postgraduate students in the Biochemistry Department (North-West University).

This project started as a dream coming true. Then, it shifted from being my MSc project to being my introduction to strangers, my personality and ambition then finaly it became my life. Now, it is over. I have to dream again so that I can start a new life for myself.

Table of contents

Abstract

Glycine-N-acyltransferase (GLYAT, EC 2.3.1.13) has been characterised in a number of species including: humans, chimpanzees, rhesus monkeys and bovines. The characterisation of GLYAT from various species contributes to a better understanding of the diversity of the enzyme which in turn might help improve the current understanding of detoxification in mammals. The GLYAT enzyme of both the chacma baboon and vervet monkey has not been characterised. In this project, tissue samples were obtained from a chacma baboon (*Papio ursinus*) and a vervet monkey (*Chlorocebus pygerythrus*) to determine the nucleic acid sequence that encodes GLYAT in these two species to broaden our current understanding on the diversity of GLYAT in primates.

A liver of a chacma baboon was used to extract total RNA. Complementary DNA (cDNA) was synthesised using an oligo (dT) primer. An open reading frame (ORF) encoding GLYAT of the chacma baboon was amplified with a PCR (polymerase chain reaction) using primers designed from a human GLYAT transcript. The PCR product containing an ORF encoding GLYAT of the chacma baboon was cloned, sequenced and expressed. The recombinant GLYAT of the chacma baboon expressed well in bacteria, but was insoluble and did not have enzyme activity. A crude cytoplasmic extract was prepared from the liver of a chacma baboon. The objective was to compare enzyme activity between the native and recombinant GLYAT. The prepared liver extract from the chacma baboon was assayed for enzyme activity and compared to the activity

in a liver extract from bovine, previously prepared by Ms M Snyders. Both the chacma baboon and bovine liver extracts had GLYAT enzyme activity.

To obtain sequence information on vervet monkey GLYAT, leukocytes were isolated from blood obtained from a living vervet monkey. A human GLYAT gene sequence was used as a reference DNA sequence in the design of PCR primers that were used to amplify the exons of GLYAT of the vervet monkey. All six GLYAT exons were individually amplified and PCR products were sequenced. The sequences were combined to reconstruct an ORF encoding GLYAT of the vervet monkey.

The ORFs coding the GLYAT of both chacma baboon and vervet monkey were found to be 888 bp long (excluding stop codon) and encoded a protein of 296 amino acids. A fragment of 1256 bp of the chacma baboon GLYAT transcript was sequenced. The two GLYAT ORF sequences were translated to amino acid sequences and aligned to that of GLYAT of primates obtained from the Ensembl sequence database. The GLYAT amino acid sequences of the chacma baboon, vervet monkey and rhesus monkey formed a related group, distinct from other primates. The chacma baboon and vervet monkey sequences were 99 % identical to the rhesus monkey sequence and 92.6 % identical to the human sequence. There were 4 new variations introduced by GLYAT amino acid sequences from the chacma baboon and the vervet monkey. The vervet monkey introduced an isoleucine in place of a valine at position 32 and an arginine in place of a histidine or glutamine at position 224. The chacma baboon introduced a tyrosine in

place of isoleucine at position 201 and an arginine in place of histidine or glutamine at position 240.

The knowledge generated in this project will broaden the understanding of GLYAT diversity relating to GLYAT in primates.

Keywords: Glycine-N-acyltransferase, recombinant GLYAT, Chacma baboon GLYAT, Vervet monkey GLYAT and primate GLYAT nucleic acid sequencing.

Opsomming

Glisien-N-asieltransferase (GLYAT, EC 2.3.1.13) is al gekarakteriseer in verskeie spesies, soos mense, sjimpansees, rhesus ape en beeste. Karakterisering van hierdie spesies se GLYAT het bygedra tot ‘n beter begrip van die diversiteit van hierdie ensiem en kan ons begrip rakende soogdiere se detoksifisering verbeter. Alhoewel ‘n verskeidenheid spesies se GLYAT al gekarakteriseer is, is dié van die chacma bobbejaan en blou aap nie bekend nie. Die doel van hierdie projek was om die GLYAT nukleïensuurvolgorde van die chacma bobbejaan (*Papio ursinus*) en blou aap (*Chlorocebus pygerythrus*) vanaf weefsel monsters te bepaal en sodoende ons begrip van GLYAT diversiteit in primate te verbreed.

Die lewer van ‘n chacma bobbejaan is gebruik om RNA te isoleer. Komplementêre DNA (cDNA) is gesintetiseer deur gebruik te maak van oligo (dT) voorvoerders. ‘n Oop leesraam (OLR) wat kodeer vir die chacma bobbejaan GLYAT is vermeerder deur ‘n polimerase ketting reaksie (PKR) te gebruik met voorvoerders wat ontwerp is vanaf ‘n menslike afskrif. Die PKR produk, wat die OLR insluit wat kodeer vir die chacma bobbejaan, GLYAT is gekloneer. Die nukleïensuurvolgorde is bepaal en die proteïen is uitgedruk. Die rekombinante chacma bobbejaan GLYAT het goeie uitdrukking getoon in bakteriële selle, maar was onoplosbaar en het geen ensiem aktiwiteit gehad nie. Om die ensiem aktiwiteit van die natuurlike chacma bobbejaan GLYAT met die van ‘n rekombinante ensiem te vergelyk, is ‘n kru sitoplasmiese ekstrak voorberei van ‘n chacma bobbejaan lewer. Die natuurlike chacma bobbejaan GLYAT ensiemaktiwiteit is

vergelyk met dié van ‘n bees lewer-ekstrak, wat voorberei is deur Me M Snyders. Beide die chacma bobbejaan en die bees lewer-ekstrak het GLYAT aktiwiteit getoon.

Om die nukleïensuurvolgorde van die blou aap se GLYAT te bepaal, is leukosiete van die bloed verkry vanaf ‘n blou aap, geïsoleer. ‘n Menslike GLYAT nukleïensuurvolgorde is gebruik as ‘n raamwerk om PKR voorvoerders te ontwerp vir die vermeerdering van die blou aap eksons. Al ses die GLYAT eksons is individueel vermeerder en die PKR produkte se volgorde bepaal. Hierdie nukleïensuurvolgordes is gekombineer om ‘n OLR wat vir blou aap GLYAT kodeer saam te stel.

Beide die blou aap en die chacma bobbejaan se OLR wat kodeer vir GLYAT was 888 bp lank en kodeer vir 296 aminosure. ‘n Fragment van 1256 bp van die chacma bobbejaan GLYAT afskrif se nukleïensuurvolgorde is bepaal. Die twee OLR nukleïensuurvolgordes is omgesit na aminosuurvolgordes en het ooreengekom met die GLYAT volgorders van primate in die Ensembl volgordedatabasis. Die GLYAT aminosuurvolgordes van die chacma bobbejaan, blou aap en rhesus aap groepeer filogeneties saam en is onderskeibaar van ander primate. Die nukleïensuurvolgordes van die chacma bobbejaan en die blou aap is 99% identies aan die rhesus aap volgorde en 92.6% identies aan die mens volgorde. Daar is vier nuwe variasies gevind in die GLYAT aminosuur volgordes van die chacma bobbejaan en blou aap. Die blou aap het isoleusien in plaas van valien in posisie 32 en arginien in plaas van histidien of glutamien in posisie 224. Die chacma bobbejaan het 'n tirosien in posisie 201 in plaas van 'n isoleusien en in posisie 240 'n arginien in plaas van 'n histidien of glutamien.

Die kennis wat in hierdie projek gegenereer is, verbreed ons begrip van die diversiteit van GLYAT in primate.

Chapter 1
Introduction and literature review

1.1 Introduction

Nowadays, the entire genome of an organism can be sequenced relatively readily and compared with genomes from other organisms. The genome of humans has been sequenced (International human genome sequencing consortium, 2004). The genome of a person of African origin has also been sequenced recently to serve as a reference for African genome alignments (Schuster *et al*, 2010). Genomes of non-human primates such as the rhesus monkey and the chimpanzee have been sequenced as well (Rhesus Rhesus monkey Genome Sequencing and Analysis Consortium, 2007).

Glycine-N-acyltransferase (GLYAT, EC 2.3.1.13) is one of the enzymes responsible for the biotransformation of potentially harmful endogenous and xenobiotic acyl-Coenzyme As (acyl-CoAs) in mammals. There are no reports on the characterisation of genes encoding GLYAT of either the chacma baboon or the vervet monkey. For this project, a chacma baboon liver and a blood sample from a vervet monkey were donated to us to investigate the molecular characteristics of their GLYAT.

The Biochemistry Division of the North-West University (Potchefstroom campus) has an initiative to study:

1) Inborn errors of metabolism

2) Supplements that can serve as therapy for inborn errors of metabolism and

3) Compare enzyme characteristics of glycine-N-acyltransferase (GLYAT) of humans with that from other species.

The study reported here addresses one of these objectives by sequencing the GLYAT open reading frame of a chacma baboon and a vervet monkey. The two ORF sequences were compared to GLYAT ORFs from humans, chimpanzees and rhesus monkeys.

Monkeys are often used in laboratories as test animals because they are a very close substitute for humans. The chimpanzee has been the primate of choice when studying human disease. The chacma baboon also provides an excellent model for genetic studies because they exhibit the same physiological characteristics that are critical to common diseases in humans. Studies with primate tissue are rare because it is very difficult to find primate tissue and get ethical approval to conduct studies using samples obtained from primates. Therefore, this study was one of those rare opportunities.

1.2 Metabolism and inborn errors of metabolism

Metabolism is responsible for the maintenance of energy requirements of the body. It is composed of mainly the building of new molecules that store energy (anabolism) and breaking down molecules in order to release energy (catabolism). Defects in genes encoding metabolic enzymes may result in failure of the body to maintain its normal

function. Heritable defects in metabolic enzymes are collectively termed inborn errors of metabolism (IEMs). IEMs in the metabolism of branched chain amino acids (leucine, isoleucine and valine) may lead to a disease state called organic aciduria which is characterised by an accumulation of organic acids. Isovaleric acidemia is an example of a dicarboxylic aciduria which is caused by the defect in leucine catabolism on the enzyme isovaleryl-CoA dehydrogenase (Tanaka, 1966).

The biochemical profile of dicarboxylic acids presented by patients suffering from isovaleric acidemia can generally be eliminated by supplementing glycine to increase glycine conjugation (Gron *et al*., 1978). The initial treatment for the disease is the restriction of leucine from diet (Levy *et al*., 1973).

Patients diagnosed with isovaleric acidemia usually respond well to glycine and carnitine supplementation which results in the excretion of acylglycines and acylcarnitine respectively (Krieger and Tanaka, 1976; Roe *et al*., 1984). GLYAT joins the acyl group with the amino acid to form a peptide bond in the biochemical process called amino acid conjugation. Glycine conjugation is the major pathway for the clearance of (C6 to C8) acyl-CoAs in patients with medium-chain acyl-CoA dehydrogenase deficiency (Rinaldo *et al*., 1993). GLYAT from bovine was shown to use the same substrates as the metabolites excreted by patients suffering from isovaleric acidemia (Bartlett and Gompertz, 1974).

1.3 Detoxification enzyme systems

1.3.1 Description of detoxification

Detoxification is a series of biochemical reactions concerned with removing unwanted and potentially harmful compounds from mammalian circulation. Detoxification occurs mostly in the hepatocytes but does also occur in other tissues such as the kidneys, lungs, intestines, and brain (Zhang *et al*, 2007; Waluk *et al*, 2010). Endogenous toxins (endotoxins) are metabolic pathway intermediates that accumulate to reach toxic levels at a particular site, metabolic end products or bacterial endotoxins (Liska, 1998). In contrast, xenobiotics (exotoxins) are any unwanted compounds introduced from outside the body by means of consumption, injection or inhalation for example pharmaceutical drugs, agricultural products, environmental polutants and the vast amount of chemicals humans come into contact with on a daily basis (Caldwell, 1986).

1.3.2 Detoxification enzymes

Toxins found in mammalian bodies are divided into polar and non-polar compounds. The very polar xenobiotics are easily excreted into urine because they are not reabsorbed by the kidneys tubules. The lipid soluble (non-polar) compounds are reabsorbed and will undergo detoxification which is a biological reaction of converting them into polar compounds which can be excreted into urine (Stachulski and Lennard,

2000). Detoxification has four phases (phase 0, I, II and III). Phase 0 has been suggested to exist with a function of importing xenobiotic compounds into the detoxifying cell (Liska, 1998; Kohle and Bock 2007). Most xenobiotics do not have an active site that would make them vulnerable for attack in a chemical reaction which prepares them for excretion. Thus, detoxification has two main steps for activation of xenobiotics (Figure 1.1): a fuctionalization step (phase I) which uses oxygen to form a reactive site on an unwanted compound and a conjugation step (phase II) which results in the unwanted compound coupled to one of the detoxifying compounds (e.g. glycine, taurine, carnitine etc).

Phase I of detoxification is mainly composed of the cytochrome P450 and flavin-containing mono-oxygenases (FMO) enzyme families. The activated xenobiotic released can behave as extremely reactive electrophilic metabolites that can covalently react with proteins, RNA and DNA resulting in cell toxicity (Gonzalez, 2005). Some authors consider the cyt P450 enzyme family as the most important elimination pathway for lipophilic drugs (van der Waide and Steijns, 1999). Phase I reactions produce reactive groups on the xenobiotics.

The function of Phase II of detoxification is to neutralise reactive groups on the xenobiotics using transferase enzymes. The transferases conjugate the reactive xenobiotics to polar molecules (Jacoby and Ziegler, 1990). Phase II is composed of

multiple families of enzymes that catalyse reactions such as sulfation, glucuronidation, glutathione conjugation, acetylation and amino acid conjugation.

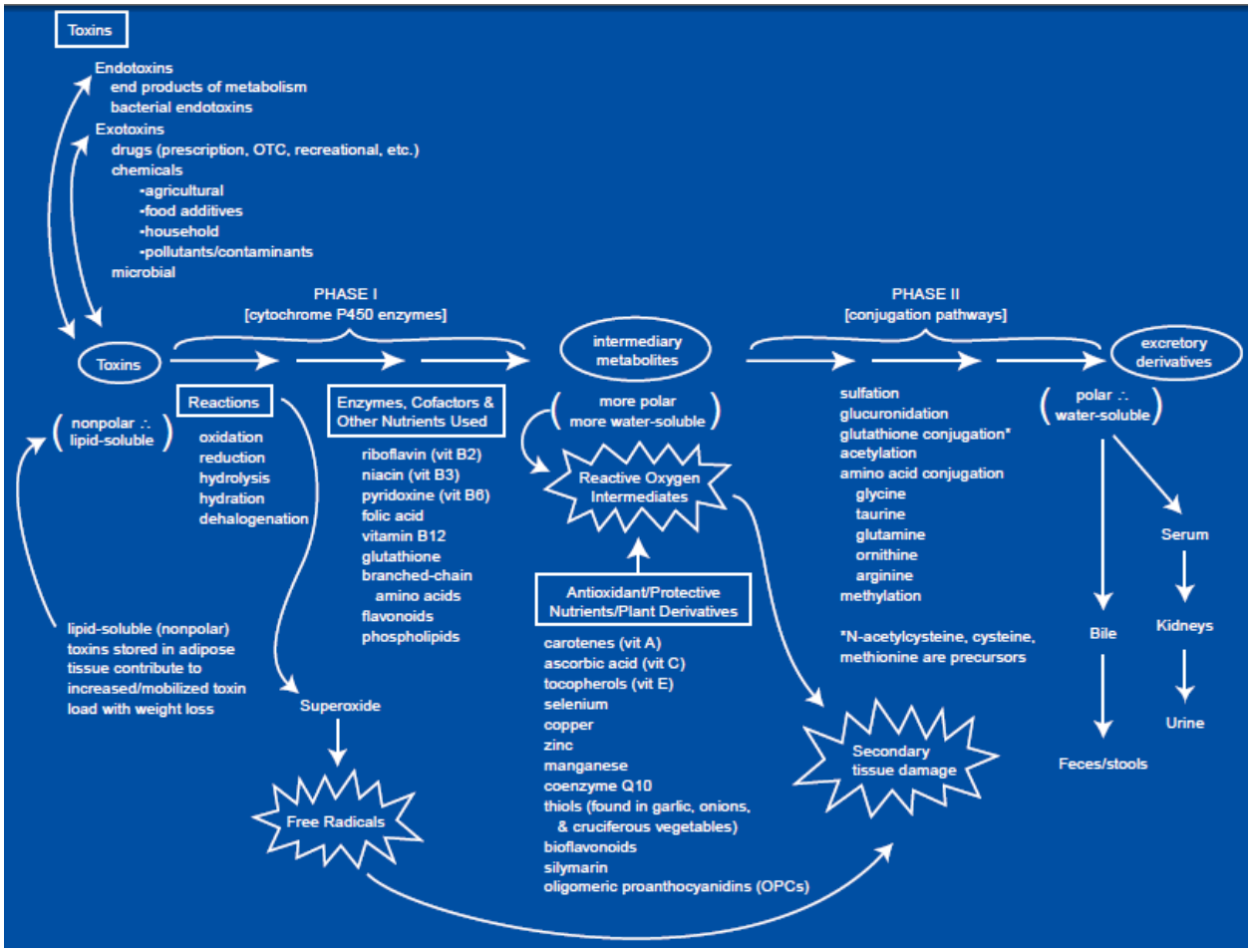


Figure 1.1: A schematic diagram illustrating phase I and II of detoxification pathway in mammals (Liska, 1998).

1.4 Single nucleotide polymorphism

Single nucleotide polymorphism (SNP) refers to a situation where there is a variation in a specific nucleotide position in a genome and the variation is present in at least 1 % of the entire population being studied. Less than 5 % of the mammalian genome codes for proteins. This explains the fact that most SNPs are found in the non-coding regions of

the mammalian genome. It is estimated that there are 60 000 SNPs that fall within the coding and untranslated regions in the human genome (Group, 2001). Evidence suggest that genetic polymorphism in detoxification enzymes does have a role in human susceptibility to life threatening diseases such as cancer, lupus erythematosus and Parkinson's disease (Coles and Kadlubar, 2003; Kang *et al*., 2005; Bandmann *et al*., 1997). The situation where an organism is likely to develop a disease due to the presence of SNPs is called genetic predisposition. It is thought that many complex diseases may be due to quantitative differences in gene products (Chakravarti, 2001). The individual's SNP map may affect drug metabolism. Knowledge about an individual's SNP map will enable physicians to prescribe medicine in more effective doses to patients and this creates a field known as personalized medicine.

There are 276 SNPs along the human GLYAT gene of which only 10 are in the coding region (GenBank: accessed 13 January 2011). It has been proposed that two of the 10 SNPs which are found in the coding region (Ser17Thr and Arg199Cys) could have an effect on GLYAT activity (Cardenas *et al*., 2010). Both the serine at position 17 and arginine at position 199 are found on the proteins' α-helix which is important for GLYAT structure and appeared to be conserved in GLYAT orthologous proteins (Cardenas *et al*., 2010). Thus, it was suggested that the serine at position 17 on the surface of the protein may interact with the surrounding water molecules while substitution of arginine with cysteine may result in addition of a positive charge in the overall protein charge disrupting the folding of the protein (Cardenas *et al*., 2010).

1.5 Protein classification and the GNAT superfamily

A protein family is defined as a group of proteins with both similar structure and functions or has amino acid sequence identity of 30 % or above (Murzin *et al*., 1995; Hubbard *et al*., 1997). A protein superfamily is defined as a group of protein families whose amino acid sequence may have low sequence identity but their structure, and sometimes function, suggest a common evolutionary history (Hubbard *et al*., 1997). Enzyme superfamilies that have the same major secondary structure in the same arrangement and the same topological connections are described as proteins with a common fold and these protein common folds are grouped into protein classes (Murzin *et al*., 1995 and Hubbard *et al*., 1997).

The GCN5-related N-acetyltransferase (GNAT) superfamily has over 10 000 families across all kingdoms of life (Vetting *et al*., 2005). The GLYAT-like family is grouped under the GNAT superfamily because of possessing a characteristic conserved GNAT fold (Figure 1.2). This fold is composed of a N-terminal strand followed by two helices, three antiparallel β strands, followed by a ‘‘signature’’ central helix, a fifth β strand, a fourth α helix and a final β strand. The function of the GNAT fold is to bind to the pantatheine arm of the acyl-CoA and to polarize the carbonyl of the thioester (Vetting *et al*., 2005).

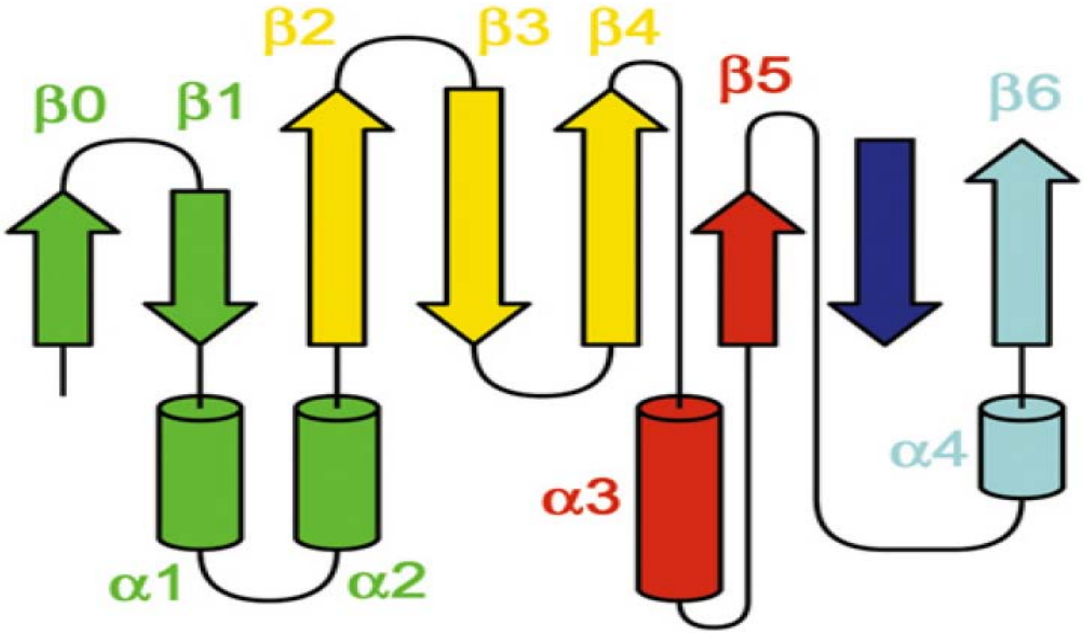


Figure 1.2: Topology of the conserved core GNAT fold. The GNAT fold starts with the β0 representing the N-terminal. The three antiparallel beta strands are followed by a signature central helix (α3). Then follows a fifth β strand, a fourth α helix and a final β strand (Vetting *et al*., 2005).

1.6 Glycine-N-acyltransferase

1.6.1 Reaction localisation and description

GLYAT (EC 2.3.1.13) is a mammalian detoxification enzyme active in phase II detoxification where it drives glycine conjugation of acyl-CoAs. In the literature, GLYAT is known with many names such as glycine acylase, ACGNAT and glycine benzoyltransferase (EXPASY enzyme database). The enzyme commission has allocated a number (EC 2.3.1.13) that groups GLYAT distinct from other enzymes. The GNAT superfamily is represented by the first digit (2) which is for all transferase enzymes. The acyltransferase family is represented by the second digit (3) which means that an acyl group is transferred. The third digit refers to the type of hydrogen or electron acceptor and in this case it is the cofactor Coenzyme A which is represented by the digit one (1). The fourth digit signifies that the glycine is transferred to the acyl and it is represented by the number 13. Hence, the complete GLYAT enzyme commission number is 2.3.1.13.

Initially, it was thought that there was only one enzyme in the liver producing both glutamate and glycine conjugates (Maldave and Meister, 1957). However, two separate acyltransferases were found in the liver of mammals: glutamine acyltransferases (GAT) and glycine benzoyltransferases (GBT) (Webster *et al*., 1976). The reason for the confusion was illustrated by Nandi and colleagues when they showed that glycine was the preferred acceptor for the enzymes benzoyl-CoA: glycine-N-acyltransferase and

phenylacetyl-CoA: glycine-N-acyltransferase (Nandi *et al*., 1979). Glycine conjugation only occurs in mammals (Vessey, 1978). Glycine conjugation was shown to occur exclusively in mitochondria of rat liver (Kolvraa and Gregersen, 1986).

1.6.2 GLYAT substrates

GLYAT was shown to be the enzyme that catalyses the transfer reactions of various aliphatic acyl groups containing 2 carbons up to 10 carbons and aromatic acyl groups from the acyl thio esters of Coenzyme A (Schachter and Taggart, 1954a and b). GLYAT binds to two substrates at a time before releasing a product (Nandi *et al*., 1979). A good example that can be used to illustrate glycine conjugation is that of the detoxification of benzoate by conjugation to glycine releasing hippurate (Figure 1.3). It has been shown that Coenzyme A is required to form a benzoyl-Coenzyme A intermediate before glycine is joint to benzoyl (Chantrenne, 1951). The benzoyl-CoA ligase catalyses the substitution of the carboxyl group with CoA at the expense of ATP. Glycine transfer is the rate limiting step in the formation of the product hippurate (Beliveau and Brusilow, 1986).

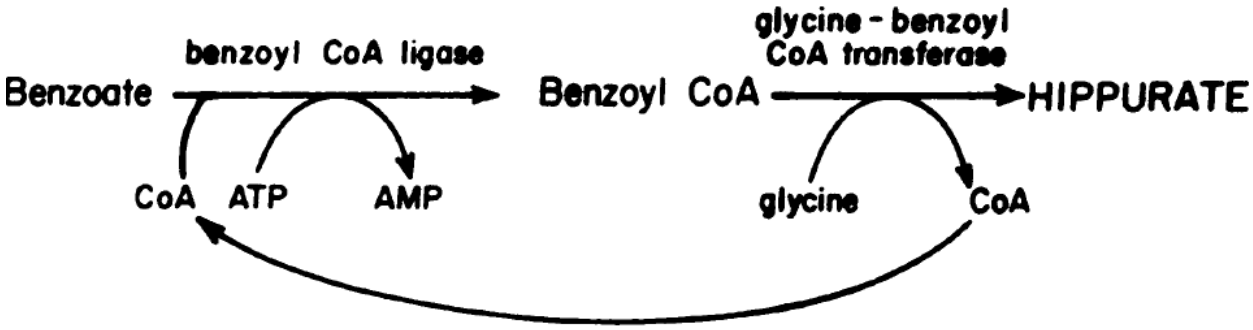


Figure 1.3: The detoxification of benzoate to release hippurate in hepatocytes. Benzoyl-CoA ligase must first substitute the carboxyl group of benzoate with Coenzyme A (CoA) while consuming ATP to release AMP. GLYAT transfers glycine to substitute CoA to produce hippurate. The glycine transfer reaction is the rate limiting step of the pathway while the CoA is recycled (Beliveau and Brusilow, 1986).

The GLYAT enzyme has a preference for substrates in the following descending order: benzoyl-CoA, salicyl-CoA, isovaleryl-CoA and octanoyl-CoA (Mawal and Qureshi, 1994). GLYAT was shown to utilise amino acids such as alanine, serine and glutamic acid in addition to glycine (van der Westhuizen *et al*., 2000). Glycine is the best amino acid substrate for the GLYAT reaction. The kinetic parameters for GLYAT when using different substrates (Table 1.1) reveal benzoyl-CoA to be the best substrate for glycine conjugation.

Table 1.1: Kinetic parameters for GLYAT (Mawal and Qureshi, 1994)

Substrate	**V_{max} (µmol/min/mg protein)**	**K_m (mM)**
Benzoyl-CoA	17.1	57.9
Salicylyl-CoA	10.1	83.7
Isovaleryl-CoA	7.64	124
Octanoyl-CoA	3.3	198

1.6.3 The role of GLYAT and glycine in non-detoxification pathways

Some GLYAT conjugates (acylglycines) have been identified to be precursors of fatty acid amide biosynthesis whose products serve as mammalian hormones (Merkler *et al*., 1996; Farrel and Merkler, 2008). Glycine has been used as pharmacotherapy for schizophrenia (Heresco-levy *et al*., 1999). Glycine acts as an anti-inflammantory immunonutrient (Wheeler *et al*., 1999) and is known to play a role as a neurotransmitter (Bowery and Smart, 2006). GLYAT has been shown to convert the benzoate and p-aminobenzoate to hippurate and by so doing removing the porphyric state (Piper *et al*., 1973).

1.6.4 GLYAT family members

There are four genes encoding GLYAT-like family members found in humans (Table 1.2). The GLYAT family members were named by adding the suffix “like plus a number” for example GLYAT like 1 is abbreviated as GLYATL1. Three of the four genes (GLYAT, GLYATL1 and GLYATL2) have been located on chromosome 11 while GLYATL3 has been located on chromosome 6. The human GLYATL1 and GLYATL2 genes have been cloned and sequenced (Zhang *et al*., 2007; Waluk *et al*., 2010). The human GLYATL1 was shown to activate the heat shock element (HSE) pathway (Zhang *et al*., 2007). The GLYATL1 was detected in liver, kidneys, pancreas, testis, ovary and stomach (Zhang *et al*., 2007). The human recombinant GLYATL2 has been expressed in the endoplasmic reticulum, salivary glands and trachea where it did show enzymatic activity (Waluk *et al*., 2010). The GLYATL2 was also detected in spinal cord, lung tissue and skin fibroblasts (Waluk *et al*., 2010).

Table 1.2: GLYAT family members

Member	Isoforms	Ensembl protein accession ID	GenBank protein accession ID
GLYAT Variant 1 Variant 2	 a = 296 aa b = 163 aa	 ENSP00000340200 ENSP00000278400	 NP_964011 NP_005829
GLYATL1 Variant 1 Variant 2 Variant 3	 a = 333 aa b = 302 aa c = 279 aa	 ENSP00000300079 ENSP00000322223 ENSP00000401353	 NP_542392 Q969I3 BAG62195
GLYATL2	294 aa	ENSP00000287275	NP_659453
GLYATL3	288 aa	ENSP00000360240	NP_001010904

1.7 Recombinant protein expression

1.7.1 The expression system

There is a variety of protein expression systems available commercially, which include fungi, plants, cultured insects, yeast, mammalian and bacterial cells. Each of these expression hosts has their own advantages. Recombinant proteins need to be expressed in an appropriate host to fold into an active protein. Fungi's biggest advantage is the ability to secrete large amounts of product into the surrounding media, but this is overshadowed by the fact that high level expression of product has not been achieved. Plants have a disadvantage of growing very slowly and have low transformation efficiency. Cultured insect cells have many processing mechanisms similar to that of eukaryotes and give high levels of expression of the product. The greatest drawback of the cultured insect cells is their lack of adequate glycosylation and therefore expressed protein is not always functional. Yeast cells do have glycosylation and formation of disulfide bond formation but the glycosylation is different from that of mammals. The mammalian cells can produce a product with the same biological activity as the native protein but with the disadvantage of slow generation time. In addition, mammalian cells can be difficult to grow and very expensive. Bacterial cells have an advantage of fast generation time, large choice of cloning vectors plus the gene expression is easily controlled. The bacterial species (*Esherichia coli, E. coli)* was selected as both the cloning and expression host because it has an added advantage of being capable of expressing recombinant protein to a total of 50 % total cellular protein (Baneyx *et al*., 1999; Sorensen and Mortensen, 2005). The biggest disadvantage of the

E. coli host is the inability to conduct post-translational modifications which may lead to a recombinant protein with different or no biological activity compared to the natural protein. However, there are reports of human transferase enzymes that have been expressed successfully in *E. coli* (Honchel *et al*., 1993; Grant *et al*., 1992).

1.7.2 The pCold protein expression system and molecular chaperones

Growing cultures at lower temperature produces a special response called the “cold shock” response from the cspA gene (Jones *et al*., 1987) which is a condition where cells produce “cold shock” proteins. The “cold shock” proteins help the cell survive the cold (Goldberg *et al*., 1997). The commercial pCold system from TAKARA™ makes use of a vector that expresses proteins using the cspA promoter. Molecular chaperones (or simply chaperones) are proteins which are co-expressed with target proteins to help with the proper folding and avoid protein aggregation to finally deliver an active enzyme.

1.7.3 Protein purification

Often, it may be required to isolate a single type of protein in order to study its properties. Protein purification is a step-by-step process of separating a single type of protein from biological samples. The initial matter resulting from the disruption of a cell or tissue is commonly called a homogenate. Total protein extracted from the homogenate is commonly called a crude cytoplasmic extract. The crude cytoplasmic extract contains between 10 000 and 20 000 different proteins. In order to separate a single type of protein from biological samples, it may be required to exploid the

characteristics of proteins such as protein size, charge, binding affinity and biological activity. Proteins which do not contain a purification tag are generally separated by ammonium sulphate first, then followed by either one of the chromatographic techniques i.e. size exclusion, ion exchange and affinity chromatography. Recombinant proteins can be marked by way of flagging (Flag™) or tagging (Histidine tag™) which means that a series of amino acids is attached to one end of the expressed protein of interest. The protein of interest can then, for example, be selected by use of a monoclonal antibody against the tag attached to an immobilized support for use in affinity chromatography.

1.7.3.1 Ammonium sulphate precipitation

The ammonium sulphate precipitation method has been used for protein precipitation for a long time (Watson and Langstaff, 1927). It is generally used to prepare native proteins from biological samples. The ammonium sulphate precipitation procedure is a more specific method of the broader concept of salting out. The concept of salting out is based on the addition of salt to absorb water in solution. This leaves the hydrophilic amino acid residues of proteins exposed to interact with each other. Proteins then start to precipitate at different rates.

1.7.3.2 Nickel affinity histidine tag-based purification

The histidine tag (His-tag) purification system allows recombinant proteins to be selectively purified from biological samples. I planned to use it to purify the chacma baboon recombinant GLYAT after being expressed in Origami® cells. In this approach, recombinant proteins are expressed containing an amino terminal tag of six histidines (Janknecht *et al*., 1991). Crude protein extracts are loaded on a column made up of nickel nitrilotriacetic acid (Ni^{2+}-NTA) and His-tagged recombinant proteins are selectively eluted with a buffer containing amidazole.

1.8 Problem formulation and aims

GLYAT is a mammalian enzyme responsible for the phase II detoxification of toxic acyl-CoAs from a variety of origins i.e. endogenous toxic metabolites and xenobiotic compounds in preparation for their excretion in urine. GLYAT is most active in mammalian liver and kidneys.

Some GLYAT conjugates (acylglycines) have been identified to be precursors of fatty acid amide biosynthesis whose products serve as mammalian hormones (Merkler *et al*., 1996; Farrel and Merkler, 2008). However, a lot is still unknown about GLYAT. Currently, there are no inborn errors of metabolism associated with human GLYAT that have been reported. The human GLYAT gene has been sequenced (Vessey and Lau, 1998; Kelley and Vessey, 1992). There are 276 SNPs along the human GLYAT gene and only 10 SNPs are in the coding region (GenBank: accessed on 13 January 2011). It has been suggested that only 2 of the 10 SNPs (Ser17Thr and Arg199Cys) may have an effect on GLYAT enzyme activity (Cardenas *et al*., 2010).

Neither the GLYAT nucleic acid nor amino acid sequences of the chacma baboon nor that of vervet monkey has been reported in the literature. Characterisation of GLYAT from additional non-human primates is necessary to help with the understanding of GLYAT diversity and gain more insight into mammalian detoxification systems. This will pave the way for future manipulation of the enzyme's primary structure for the rational

design of a therapeutic recombinant GLYAT with broader substrate specificity for acyl-CoAs. Currently, an ideal GLYAT is required that could conjugate a much larger spectrum of toxic metabolites resulting from inborn errors of metabolism.

Aim:

The main aim of this project was to determine the nucleic acid sequence encoding GLYAT open reading frame from the vervet monkey (*Chlorocebus pygerythrus*) and chacma baboon (*Papio ursinus*).

The specific objectives of the project were to:

(a) Clone and sequence the cDNA derived from the mRNA of GLYAT from the chacma baboon

(b) Express the cloned amplicon encoding GLYAT from chacma baboon and test if the recombinant enzyme has activity

(c) Sequence the vervet monkey GLYAT open reading frame from genomic DNA

(d) Compare the nucleic acid and amino acid sequence of GLYAT originating from humans, chimpanzee, rhesus monkey, chacma baboon and vervet monkey

(e) Extract chacma baboon GLYAT from liver tissue and test it for enzyme activity

Chapter 2
DNA sequencing and analysis of the open reading frames that encode GLYAT of the chacma baboon and the vervet monkey

2.1 Introduction

Glycine-N-acyltransferase (EC 2.3.1.13, GLYAT) is one of the enzymes from the GCN5-related N-acetyltransferase (GNAT) superfamily responsible for the biotransformation of endogenous and xenobiotic toxic compounds. GLYAT conjugates glycine to acyl-CoA in mammalian mitochondria, mainly in the liver and kidneys (Schachter and Taggart, 1953).

Patients with organic acidemias such as isovaleric acidemia, 3-methylcrotonylglycinuria and propionic acidemia usually respond well to glycine therapy and excrete acylglycine (Krieger and Tanaka, 1976). The Biochemistry Department at the North-West University (Potchefstroom campus) has an initiative to study detoxification profiling using substrate loading tests. Analytical methods are used to detect metabolites associated with detoxification. The aim is to eventually be able to correlate patients' genotype to their biochemical detoxification profile. Hence, the genes encoding GLYAT enzymes need to be characterised.

The entire genes encoding GLYAT have been sequenced for humans (ENSG00000149124), the chimpanzee (ENSPTRG00000003683), rhesus monkey (ENSMMUG00000011718) and bovine (ENSBTAG00000038323). The aim of this project was to characterize the GLYAT enzyme in two additional non-human primates so that we can start to understand its diversity and gain more inside into the mammalian detoxification system. The core GNAT fold encoded by exon 6 of the GNAT superfamily has been shown to be conserved in all kingdoms of life (Vetting *et al*., 2005, Dyda *et al*., 2000). Thus, sequencing the ORF encoding GLYAT from the chacma baboon and vervet monkey will enable comparison of this exon 6 in two more primates. The alignment of GLYAT sequences from human, chimpanzee, rhesus monkey, chacma baboon and vervet monkey is expected to show similarities and differences on nucleic acid and amino acid level. These similarities and differences can eventually be used to rationally design a recombinant GLYAT enzyme, with altered substrate specificity. A GLYAT enzyme with altered substrate specificity would be able to biotransform a much larger substrate spectrum. As a result such an enzyme can be used for the treatment of patients with a wide range of organic acidemias. Neither the genomes nor the GLYAT genes of both the chacma baboon and vervet monkey have been sequenced. Tissue samples from chacma baboon and vervet monkey were available to start this project. This chapter reports, for the first time, nucleic acid sequences of the open reading frame encoding GLYAT for both the chacma baboon and vervet monkey.

The objectives of the work described in this chapter were to:

1) Sequence the open reading frame encoding GLYAT from both the chacma baboon and vervet monkey

2) Compare and analyse the GLYAT deduced amino acid sequences from humans, chimpanzee, rhesus monkey, chacma baboon and vervet monkey

2.2 Materials and methods

2.2.1 Cloning of cDNA encoding chacma baboon GLYAT

2.2.1.1 Ethical approval

The chacma baboon liver tissue was obtained from an project that was approved by the Ethics Committee of the North-West University (ethical approval number: NWU-00005-09-A1). A blood sample of a vervet monkey was obtained from a living vervet monkey from another project approved by the Ethics Committee of the North-West University (with an ethical approval number of NWU-0022-09-S5).

2.2.1.2 Extraction of total RNA from chacma baboon liver

An extract of total RNA from mammalian cells contains many RNAs including: ribosomal RNA (rRNA), transfer RNA (tRNA) and messenger RNA (mRNA). The mRNA, which is destined to be translated into polypeptides, is characterised by a stretch of adenine repeats known as the poly (A) tail which is between 150-200 nucleotides at the 3'-ends (Edmonds and Caramela, 1969). The main reason for using total RNA from the liver is to obtain an mRNA that will eventually provide a sequence of the open reading frame encoding for the GLYAT that was expressed in the liver of chacma baboon.

The principle of the kit used to extract total RNA (QIAGEN, cat # 74104) is based on a chromatographic purification of RNA where the selective binding property of nucleic acids to silica based membrane, in the presence of high salt buffer and ethanol, is

exploited (Boom *et al*., 1990). The principle for the binding of nucleic acid acids to silica is thought to be due to the dehydration of the phosphodiester backbone by the chaotropic salt, which allows the phosphodiester to adsorb to silica (Russell and Sambrook, 2001). RNAs less than 200 nucleotides and DNA contaminants are excluded. RNases are non-specific RNA digesting enzymes that will degrade total RNA as soon as it is released from the cells. Thus, a buffer containing RNases deactivators such as guanidine thiocyanate and 2-mercaptoethanol is used to deactivate the RNases (Chirgwin *et al*., 1979). The chromatographic column is washed with ethanol and total RNAs with a length greater than 200 nucleotides are eluted with autoclaved milliQ (18 Ω) RNase free water.

A 30 mg piece of chacma baboon liver was homogenised in RNAlater® on a Heidolph silent crusher S (01-005-002-74-1). Total RNA was extracted using the QIAGEN kit (cat #74104). The quantity and quality of the extracted total RNA was determined first on a NanoDrop® ND-1000 Spectrophotometer (as described in section 2.2.1.3) and then loaded on a 1.2 % agarose formaldehyde denaturing gel (as described in section 2.2.2.4).

2.2.1.3 Determination of the concentration and purity of nucleic acid samples

The concentration and purity of nucleic acid extracts was determined on a NanoDrop® ND-1000 Spectrophotometer. The pinciple of the spectrophotometer takes advantage of the fact that nucleic acids absorb most ultra-violet (UV) light at optical density (OD) of 260 nm (OD_{260}) and aromatic proteins absorb most ultra-violet light at OD_{280}. One OD_{260}

of single stranded RNA corresponds to a concentration of 40 ng/µl and a pure sample has an OD_{260}/OD_{280} value of 2.0. One OD_{260} of double stranded DNA corresponds to a concentration 50 ng/µl and a pure sample has an OD_{260}/OD_{280} ratio of 1.8. Sample purity for nucleic acids is determined by dividing the absorbance values obtained at OD_{260} by those at OD_{280} to give what is known as an OD_{260}/OD_{280} ratio (Russell and Sambrook, 2001).

2.2.1.4 Characterisation of nucleic acids on agarose gel electrophoresis

An agarose gel is composed of a linear polysaccharide (M_r = 12000) made up of repeats of agarobiose. The agarobiose is actually a unit composed of galactose and 3, 6-anhydrogalactose. The gelling properties of agarose are due to intermolecular and intramolecular hydrogen bonding that occurs within and between agarose chains. When the gel cools down (settles) pores are formed within the gel. The sizes of these pores are determined by the concentration of the agarose i.e. the higher the concentration the smaller the pores will be and vice versa. Nucleic acid molecules are separated based on the pore size i.e. smaller fragments of nucleic acids will travel faster while larger fragments will follow behind at a slower rate when an appropriate voltage is applied. Both the 1 % agarose gel electrophoresis and 1.2 % formaldehyde agarose denaturing gel electrophoresis were performed as described by Russell and Sambrook (2001).

2.2.1.4.1 Agarose gel electrophoresis for DNA separation

A 1 % agarose gel: TAE buffer was prepared by dissolving agarose in a 1 X TAE buffer (40 mM Tris acetate and 1 mM EDTA, pH 8) to a final concentration of 10 mg/ml. The

mixture was boiled in a microwave oven until agarose was completely dissolved. The solution was cooled to about 60 °C before ethidium bromide was added to a final concentration of 1 µg/ml. The solution was mixed by swirling. The 50 ml solution was poured on a gel casting stand (length = 14 cm, width = 10 cm and height = 0.7 cm) and allowed to solidify with combs (depth = 6 mm, base width = 5 mm and base thickness = 1 mm) inserted for an hour. After the gel had cooled, combs were removed and the gel was immersed in a reservoir containing a 1 X TAE buffer. To prepare samples for loading on the gel, DNA samples were mixed with a 6 X DNA loading dye (10 mM Tris-HCl pH 7.6, 0.15 Orange G, 0.03% xylene cyanol FF, 60 % glycerol and 60 mM EDTA) from Fermentas™. A constant 70 volts and an electric current of 25 mA were applied for 90 min (powered by BIO-RAD PowerPac™ HC power supply cat #164-5052). The resulting gel was visualised after ethidium bromide staining using a Dark reader (DR88M) transilluminator purchased from Clare Chemical Research. All agarose gels were documented under ultra violet light using a SYNGENE™ Chemni Genius Bio Imaging System.

2.2.1.4.2 Agarose formaldehyde denaturing gel electrophoresis for RNA separation

Formaldehyde denatures RNA secondary structures. This will allow RNA to separate according to size on the gel. A 1.2 % agarose formaldehyde denaturing (FA) gel was prepared by dissolving agarose in a 10 x FA gel buffer (200 mM 3-N-morpholino propanesulfonic acid (MOPS), 50 mM sodium acetate and 10 mM EDTA at pH 7). The mixture was boiled in a microwave oven until agarose was completely dissolved. The

solution was cooled to 65 °C then formaldehyde and ethidium bromide were added to final concentrations of 220 mM and 1 mg/ml respectively then mixed by swirling. The solution was poured on a gel casting stand (length = 14 cm, width = 10 cm and height = 0.7 cm) and allowed to solidify with combs (depth = 6 mm, base width = 5 mm and base thickness = 1 mm) inserted for an hour. Then, combs were removed and the solid gel was submerged in a gel reservoir containing 1 x FA gel buffer (20 mM 3-N-morpholino propanesulfonic acid (MOPS), 5 mM sodium acetate and 1 mM EDTA at pH 7). To prepare samples for loading on the gel, one volume of 5 x RNA loading buffer (0.03 % bromophenol blue (w/v), 0.004 M EDTA, 0.8856 M formaldehyde, 20 % glycerol, 31 % formamide, 40 % 10 X FA gel buffer) was mixed with 4 volumes of RNA samples. The RNA/dye solution was incubated for 3 to 5 min at 65 °C, chilled on ice for a minute and then a total of 10 µl was loaded on the FA gel. The RNA was separated at a constant 70 volts and an electric current of 25 mA for 90 min (powered by BIO-RAD PowerPac™ HC power supply, cat #164-5052). The resulting gel was visualised after ethidium bromide staining using a Dark reader (DR88M) transilluminator purchased from Clare Chemical Research. All agarose gels were documented under ultra violet light using a SYNGENE™ Chemni Genius Bio Imaging System.

2.2.1.5 cDNA synthesis

Complementary DNA (cDNA) is usually synthesised in a single step where the reaction components of both cDNA synthesis and polymerase chain reaction (PCR) are incubated in the same tube at the same time. Although this method has an advantage of using the entire cDNA synthesised it has a major drawback which is the fact that

reaction conditions can not be optimised for either the cDNA synthesis or the PCR reaction. However, in this project a two step approach was used were cDNA was synthesised separately from the PCR.

The polyadenylated mRNA represents 1 to 5 % of total RNA from a eukaryotic cellular preparation. It needs to be converted to cDNA before it is amplified. The method of converting mRNA into cDNA employs a reverse transcriptase. A region of DNA composed of repeats of thymine (T) base (oligo [dT]) bind to the poly (A) tail of the mRNA in order to prime transcription. The reverse transcriptase is actually a RNA dependant-DNA polymerase. A cloned moloney murine leukemia virus reverse transciptase (MMLV-RT), which lacks activity for RNase H that could degrade RNA, is often used to synthesise cDNA (Kotewicz *et al*., 1988).

A cDNA synthesis kit, MMLV High Performance Reverse Transcriptase, (EPICENTRE® cat #RT80125K) was used to synthesise chacma baboon GLYAT cDNA. Since the reagents used for cDNA synthesis were from the same kit, they shared the same catalogue number (RT80125K) except if stated otherwise. A cDNA synthesis reaction was set up as follows: 100 pg of total RNA, 10 pmol oligo $(dT)_{18mer}$ and autoclaved milliQ (18 Ω) RNase free water to final volume of 10 µl. The reaction was incubated at 65 °C for 2 min (in a BIO-RAD MJ Mini™ Gradient Thermal cycler (PTC-1148)) then chilled on ice for a minute. The following components were added to the reaction: 1 X reaction buffer (5 % glycerol, 5 mM Tris-Cl, pH 7.5), 0.5 mM for each of the dNTPs (TAKARA™ lot number BF2601A), 100 mM DTT, 5 U RNase inhibitor, 100 U MMLV HP

reverse transcriptase and RNase free water to final volume of 20 µl. The reaction was incubated at 37 °C for 60 min and the reaction was terminated by incubation at 85 °C for 5 min then chilled on ice for a minute. The cDNA was stored at -20 °C.

2.2.1.6 Polymerase chain reaction

DNA can be amplified from a single copy, of a specific region on DNA template, to billions of copies by means of a polymerase chain reaction (PCR). A PCR is a biochemical reaction whereby DNA is synthesised by a thermostable DNA dependent DNA polymerase enzyme *in vitro*. A PCR has three main steps: DNA denaturing step, primer binding step and elongation step. The DNA is completely denatured by a single cycle by heating the reaction up to temparatures of between 90 °C and 100 °C. There is a cyclic step whose purpose is to allow: DNA denaturing, primers to bind on template DNA and the elongation of the newly synthesised DNA strand. During the cyclic step, the reaction temperature is decreased (immediately after final denaturation) and this allows primers to bind (temperature varies depending on primer set used) then increased to 72 °C where the Taq polymersase has optimum activity to elongate the new DNA strand. At the end of the cyclic step, the reaction is kept at 72 °C for a few minutes to extend the new DNA copies. The general PCR composition is made up of a buffer suitable for the polymerase enzyme used, DNA template, deoxynucleotide triphosphates (dATP, dCTP, dGTP and dTTP) in equimolar ratio, primer set (forward and reverse primers), DNA polymerase, magnesium (Mg^{2+}), and water.

All reagents used for PCR were from the TAKARA™ *Ex Taq* kit (cat #PR001A). The PCR was performed in a BIO-RAD MJ Mini™ Gradient Thermal cycler (PTC-1148). The PCR components were: a 1 X *Ex Taq* PCR buffer, 0.2 mM dNTPs, 0.2 µM each primer, 1.25 U TAKARA™ *Ex Taq* DNA polymerase, 2 mM Mg^{2+} and sterile water to final volume of 50 µl. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is known to be present in all living animals and for that reason primers known to amplify GAPDH were used as positive control for PCR. PCR was done under the following conditions: the initial denaturation was done at 94 °C for 30 sec. The cyclic step was repeated 30 times and involved: final denaturation at 94 °C for 30 sec, primer annealing at 60 °C (depending on the primer set used; see Table 2.1) for 30 sec and elongation at 72 °C for 30 sec. The extension step was at 72 °C for 10 min. The reaction was held at 4 °C.

Table 2.1: Primers used to amplify and sequence GAPDH, vervet monkey GLYAT exons and ORF encoding chacma baboon GLYAT

Region amplified	Primer direction	Nucleotide sequence	Annealing temperature
GAPDH	Forward Reverse	5'- gaaggtgaaggtcggagtc-3' 5'-gaagatggtgatgggatttc-3'	62 °C
Exon 1	Forward Reverse	5'-cagattcttttgccagcctagtac-3' 5'-cactcatgtagcatggatcccatataca-3'	56 °C
Exon 2	Forward Reverse	5'-cagctcgttctcagaggagtcag-3' 5'-gcagtgtttagactaagg-3'	60 °C
Exon 3	Forward Reverse	5'-agtggttgtctgcctctctgtg-3' 5'-gccctggctctaccatattgc-3'	65 °C
Exon 4	Forward Reverse	5'-caggatatgacagatgaccttgat-3' 5'-tctggagcttggaggaag-3'	60 °C
Exon 5	Forward Reverse	5'-ggaaagccagagtgaatgcag-3' 5'-tagcaccaagcccagaacc-3'	65 °C
Exon 6	Forward Reverse	5'-gattctcacagacaccaaatctgctg-3' 5'-cttcactctgttcctctttcatca-3'	56 °C
ORF	Forward	NdeI 5'-aattcatatgatgttaccattacaaggtgc-3'	62 °C

	Reverse	XhoI 5’-aattctcgagcagaggcacacagttccac-3’	
5’-Baboon GLYAT_UTR	Forward Reverse	5’-caaaatttactctgcgagactctgctg-3’ 5’-gtgcatgggtaacatccatg-3’	62 °C
3’-Baboon GLYAT_UTR	Forward Reverse	5’-ccccaaaactgtcaggaattc-3’ 5’-gagctcgagcaatatgtatctgatg-3’	62 °C
M13 (sequencing across the multiple cloning site of the TA vector)	Forward Reverse	5’-gttttcccagtcacgac-3’ 5’-caggaaacagctatgac-3’	

2.2.1.7 PCR clean up and gel extraction

A PCR reaction contains various components that may interfere with downstream application of the PCR product such as the DNA polymerase, primers, PCR buffer, $MgCl_2$ and especially incorrectly synthesised DNA fragments. A PCR amplicon needs to be purified from these contaminants before cloning. The Macherey-Nagel Nucleospin Extract II™ kit’s principle is based on the property of nucleic acids of binding to a silica based membrane in the presence of 6 M sodium perchlorate (Marko et al, 1982). The principle for the binding of nucleic acid acids to silica is thought to be due to the dehydration of the phosphodiester backbone by the chaotropic salt, which allows the phosphodiester to adsorb to silica (Russell and Sambrook, 2001).Contaminants such as rRNA, proteins, agarose and nucleic acid fragments below 165 bases are washed away during the washing step.

The amplicons resulting from vervet monkey GLYAT exons amplification were PCR cleaned directly after PCR. The PCR amplicon of the ORF encoding chacma baboon

GLYAT was first electrophoresed on a 1 % agarose gel (as in section 2.2.1.4.1) then gel extracted. Both PCR clean up and gel extractions were performed using the Macherey-Nagel Nucleospin Extract II™.

2.2.1.8 Cloning of a PCR product containing an ORF encoding for GLYAT of chacma baboon

2.2.1.8.1 TA cloning of the PCR product containing an ORF encoding for GLYAT of chacma baboon

The TA vector is designed to be a plasmid having a 3'-ddT overhang as cloning sites. The 3'-ddT overhang of the TA vector complements the 3'-dA overhang of the PCR product produced by Ex Taq™ DNA polymerase (Figure 2.1 illustrates how the PCR product is cloned into the TA vector).

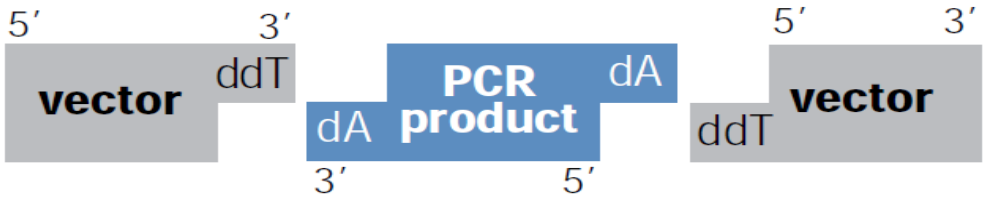


Figure 2.1: A schematic representation of the TA vector cloning site with a PCR product simulated

The principle of the Fermentas InsTaclone cloning kit (cat #K1214) involves the use of a specialized type of a TA vector called pTZ57R/T (see Figure 2.2). The pTZ57R/T is 2886 bp and has a multiple cloning site as shown in Figure 2.2, lac operator that regulates expression of the insert from the T7 promoter and the bla gene that expresses resistance to ampicillin.

The PCR product containing the chacma baboon GLYAT ORF was cloned into the pTZ57R vector (InsTAclone™ cloning vector cat #K1214). DNA ligation reaction specific to TA cloning was set up as follows: 5 U T4 DNA ligase, 0.15 µg pTZ57R DNA, 1 x ligation buffer, GLYAT DNA and water to final volume of 30 µl. The reaction was incubated at 22 ºC for one hour.

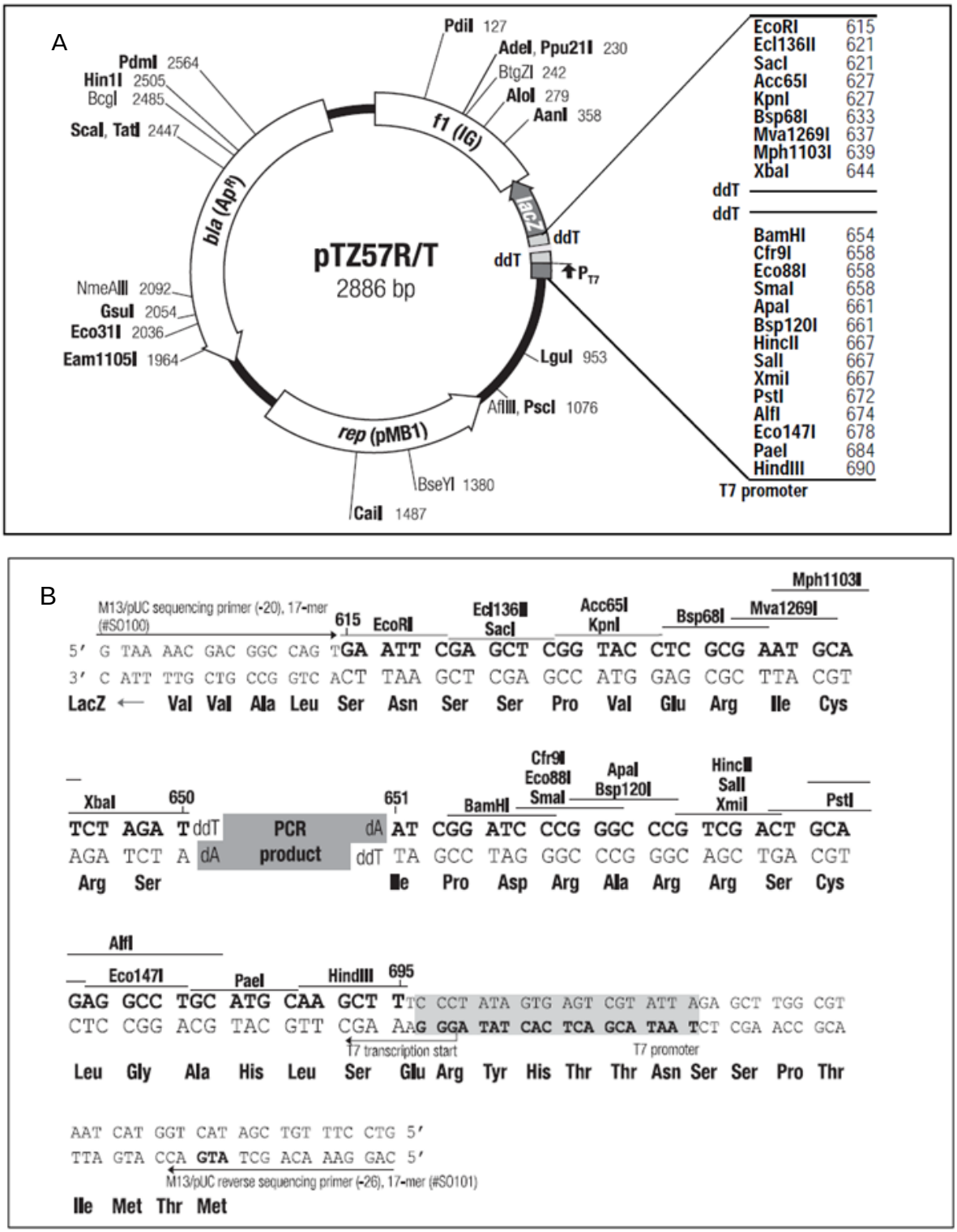


Figure 2.2: (A) A vector map of pTZ57R. (B) The base sequence of pTZ57R's multiple cloning site.

2.2.1.8.2 Directional cloning of PCR amplicons containing the ORF that encodes for GLYAT of chacma baboon into pColdIII vector

Directional cloning refers to the type of cloning that involves restriction enzyme digestion of both the vector and the insert with the same restriction enzymes to give compatible ends for their ligation. A PCR product containing an ORF encoding GLYAT of chacma baboon was cloned into pColdIII in preparation for expression in bacteria. The pColdIII is a commercial vector that is 4377 bp long and has the multiple cloning site shown in Figure 2.3. The pColdIII expresses the insert in the multiple cloning site from the cspA promoter that is induced by cold environments. The cspA promoter is regulated by the *lac* operator. The pColdIII expresses resistance to ampicillin.

Separate restriction enzyme digestion of pColdIII and PCR product containing ORF encoding for GLYAT of chacma baboon was done using the restriction enzymes NdeI and XhoI as stated in section 2.2.1.8.7. Reactions were incubated at 37° C for 16 hours. The appropriate fragments were gel extracted separately (as stated in section 2.2.1.7). The linear pColdIII was ligated to the digested PCR product containing the GLYAT ORF of chacma baboon using a T4 ligase enzyme as shown in section 2.2.1.8.8.

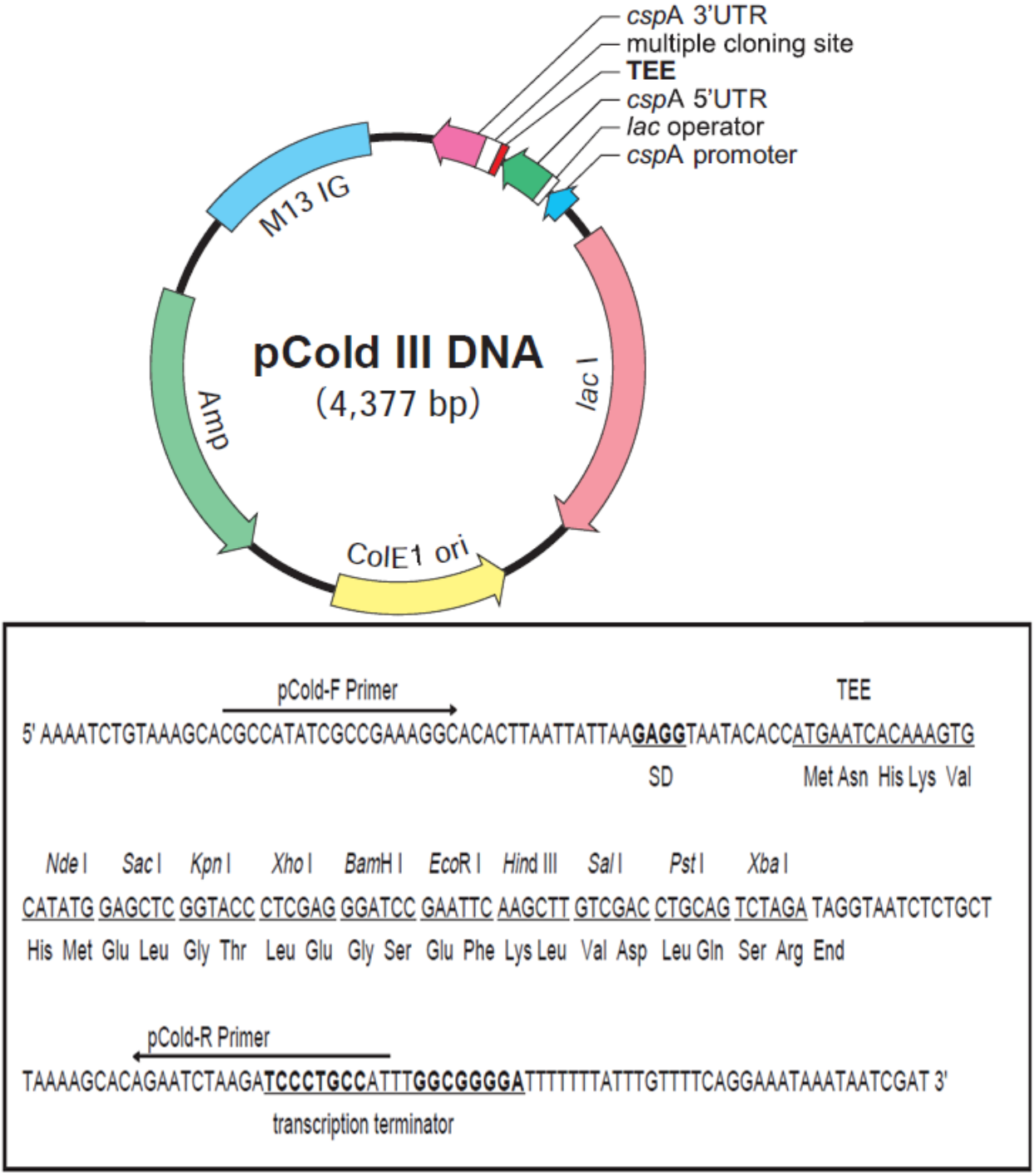


Figure 2.3: A diagrammatic illustration of pColdIII with the DNA sequence of its multiple cloning site below

2.2.1.8.3 Preparation of chemically competent *Esherichia coli* cells

The *E. coli* cells need to be prepared for the uptake of foreign DNA (transformation) in a procedure that is commonly termed “preparation of competent cells”. The principle of the preparation of *E. coli* chemically competent cells is based on the observation that *E. coli* cells tend to be susceptible to take up foreign DNA from their environment after been treated with a buffer containing calcium chloride ($CaCl_2$) (Hanahan, 1983; Mandel and Higa, 1970). Competent cells can be stored in 15 % glycerol at -80 °C for months without losing their competency (Morrison, 1977). The best chemical method for preparation and transformation of ultra-competent *E. coli* was used (Inoue *et al*., 1990). This method is similar to the chemical method first described by the heat shock method (Hanahan, 1983) but the main difference is that the culture is grown at temperatures between 18 ºC and 22 °C instead of 37 ºC. The decrease in temperature is suspected to favour the efficiency of the transformation. The method for preparation of ultra competent *E. coli* cells (Inoue *et al*., 1990; outlined by Russell and Sambrook, 2001) was used to prepare competent JM109™ and Origami™ cells.

A bacterium was inoculated in a 25 ml Luria-Bertani (LB) broth in a large flask (250 ml) for better aeration. The inoculated broth was incubated at 37 °C, shaking at 225 rpm, for 8 hours. The culture was divided into three volumes: 10 ml, 4 ml and 2 ml. These three cultures were each inoculated in 250 ml LB broth and incubated at 18 °C, shaking at 140 rpm, until density of one of the cultures reached an OD_{600} of 0.55 (measurements were done using a Biochrom™ Novaspec II® visible spectrophotometer). The culture

that reached an OD_{600} of 0.55 was selected and placed in an ice water bath for 10 min while the other two were discarded. The cells were harvested by centrifugation at 2500 x g for 10 min at 4 °C. The supernatant was discarded and the pellets were suspended in 80 ml ice-cold Inoue transformation buffer (55 mM $MnCl_2 \cdot 4H_2O$, 15 mM $CaCl_2 \cdot 2H_2O$, 250 mM KCl, 10 mM PIPES and all dissolved in autoclaved milliQ (18 Ω) water). The PIPES [piperazine-1,2-bis(2-ethanesulfonic acid)] stock solution (0.5 M pH 6.7) was prepared by dissolving 15.1 g of PIPES in 100 ml autoclaved milliQ (18 Ω) water then filtered through a disposable prerinsed Nalgene filter (0.45 µm pore size). The filtered PIPES solution was aliquoted and stored at -20 °C. The cell pellet was suspended by swirling then they were harvested by centrifugation at 2500 x g for 10 min at 4 °C. The supernatant was discarded and the cell pellet was suspended in 20 ml ice-cold Inoue transformation buffer. The suspension was mixed with 1.5 ml DMSO by swirling and incubated in ice-water bath for 10 min. The suspended cells were aliquoted in 2 ml volumes (polypropylene tubes) and snap frozen in liquid nitrogen then stored in -80 °C, for long term storage**.**

2.2.1.8.4 Transformation

The principle of transformation is based on the brief and sudden heat shock of competent cells at 42 °C in the presence of DNA and chilled to facilitate transformation (Hanahan, 1983). The transformed cells are grown on agar plates with selective antibiotics that are used to select the successfully transformed cells. Transformation efficiency of 2 x 10^9 cfu/ µg is considered ideal (Russell and Sambrook, 2001). The transformation was conducted as follows: all cultures were grown at 37 °C (in a

Multitron™, INFORAG™ 2000, incubator shaker) to reach OD_{600} reading of 0.5 (measurements were done using a Biochrom™ Novaspec II® visible spectrophotometer) before transformation. One OD_{600} of an actively growing *E. coli* culture contains ~10^9 bacteria/ ml. DNA mass of 1 ng was used to transform a total volume of 50 µl competent cells using the heat shock method (Inoue *et al*.,1990; as outlined in Russell and Sambrook, 2001).

$$\text{Transformation efficiency} = \frac{\text{number of colonies on plate (successful transformants)}}{\text{mass of DNA used to transform in ng/ volume of cells in µl}} \times 1000 \text{ ng/µg.}$$

Equation 1.1

Three 1.5 ml Eppendorf tubes each containing 50 µl competent cells was set up: one had 1 ng of transforming DNA (experiment), the other had 1 ng of a known plasmid (positive control) and the last had no plasmid (negative control). Tubes were stored on ice for 30 min. Cells in the three tubes were heat shocked by suspending the tubes in water bath at 42 °C for 90 sec. Then, the cells were cooled in an ice water bath for 2 min. A volume of 800 µl SOC medium (Russel and Sambrook, 2001) was added to each tube and incubated at 37 °C, shaking at 180 rpm, for 45 min. The three cultures were separately spread on SOB agar (Russel and Sambrook, 2001) with appropriate antibiotic and incubated overnight at 37 °C.

2.2.1.8.5 Selection of successfully transformed bacteria

Two approaches were used to screen for successfully transformed cells: the antibiotics and “blue/white” screening. The antibiotics approach involved the use of antibiotics in media for growing cells. A positive control was defined as the transformation of *E. coli* cells with a known amount of a known plasmid used to test if cells were competent and calculate their transformation efficiency. A positive control provides a reference that should give an indication whether transformation was successful. A negative control was defined to be growing *E. coli* cells without the plasmid that can give them selective advantage over the antibiotics used. This negative control should indicate whether the antibiotics used were effective. Luria-Bertani (LB) media and antibiotics concentrations were prepared as in Russell and Sambrook (2001).

The success of the transformation can be screened by including a substrate for β-galactosidase in the medium of the transformed culture (Lim and Chae, 1989). The β-galactosidase has a normal function of breaking down lactose to glucose and galactose. The β-galactosidase is a tetramer with one of the monomers (amino-terminal) linking the three other monomers. The two fragments need to associate to produce an active β-galactosidase (Ullmann *et al*., 1967). Most plasmid vectors are designed to have a regulatory region such as the lac operon. The *lac* operon has the *lacz* gene that regulates the expression of the β-galactosidase. Therefore, inactivation of the *lacz* repressor by the IPTG results in increased expression of the β-galactosidase. The multiple cloning sites of many plasmid vectors carry a DNA sequence encoding for the

first 146 amino acids of the β-galactosidase gene. Thus, cloning an insert into the multiple cloning site will disrupt the expression of the amino-terminal fragment that complements the other three monomers in the formation of an active α-galactosidase. The β-galactosidase can convert X-gal (5-bromo4-chloro-3-indolyl-β-D-alactopyranoside) to an insoluble dense blue compound (Horwitz *et al*., 1964). In α-complementation screening, the expression of an active β-galactosidase from a plasmid vector is used to indicate failure to clone into the multiple cloning sites. An active β-galactosidase is indicated by the conversion of the X-gal (present in medium) to an insoluble blue compound which will result in colonies turning blue. In contrast, white colonies would mean that the α-complementing fragment failed to express because an insert was successfully cloned into the multiple cloning site.

Transformed cells were spread on Luria-Bertani agar (LB-agar) plates containing: 100 ug/ml ampicillin, 0.027 mg/ml X-gal (Promega cat #3941), and 0.13 mM IPTG (Promega cat #V395A). Plates were incubated at 37 °C overnight. The next day, white colonies were selected (to make pure colonies, “master plate”) and streaked on LB agar containing 100 µg/ml ampicillin, 27 mg/ ml X-gal and 0.13 mM IPTG and incubated at 37 °C overnight.

2.2.1.8.6 Extraction of plasmids from transformed cells using the PureYield™ Plasmid Midi preparation system

The PureYield™ Plasmid Midiprep system (Promega™ cat #A2496) purifies plasmids by selectively binding them to a silica-membrane column. The system uses an endotoxin removal wash step to remove a substantial amount of protein, RNA and DNA contaminants. Purification does not use the isopropanol precipitation of purified DNA nor does it require extensive centrifugation which translates into a faster method with a higher concentration of purified plasmid. Hence, plasmids were midi extracted using this system.

Transformed cells were grown in 100 ml LB broth overnight at 37 °C. The next day, cells were centrifuged at 4 000 x g for 10 min (Heraeus Multifuge® 1L/1L-R centrifuge from Kendro). Cell pellets were resuspended in 3 ml cell resuspension buffer (50 mM Tris-HCl, pH 7.5; 10 mM EDTA; 100 µg/ml RNase A) then lysed with 3 ml cell lysis buffer (0.2 M NaOH; 1% SDS). The mixture was mixed by inversion 3 to 5 times by hand then incubated at room temperature for 3 min. The pH of the mixture was neutralized by addition of 5 ml neutralising buffer (4.09 M guanidine hydrochloride, pH 4.2; 759 mM potassium acetate; 2.12 M glacial acetic acid) and inverting 3 to 5 times then incubation at room temperature for 3 min so that a white flocculent precipitate can form. To remove cellular debris, the mixture was first centrifuged at 15 000 x g for 15 min then the supernatant was loaded on the PureYield™ clearing column which blocked the cellular debris but allowed DNA, RNA and small proteins to pass through. The filtered lysate

was loaded on the PureYield™ binding column to bind plasmid DNA and centrifuged at 1 500 x g for 3 min. The endotoxin removal wash solution was added to remove contaminating endotoxin, proteins, RNA and endonucleases. The solution was centrifuged for 1 500 x g for 3 min. The column was washed by adding column wash buffer (60 % ethanol; 60 mM potassium acetate; 8.3 mM Tris-HCl; 0.04 mM EDTA.) and centrifuging at 1 500 x g for 3 min. Then the column was centrifuged again at 1 500 x g for 10 min to dry the column. Plasmid DNA was eluted with autoclaved milliQ (18 Ω) water.

2.2.1.8.7 Restriction enzyme digestion reaction

Unless indicated otherwise a general restriction enzyme digestion reaction was set up as follows: 1 µg of DNA, 5 U restriction enzyme, 1 X reaction buffer and autoclaved milliQ (18 Ω) water to a final volume of 15 µl. The reaction was incubated overnight.

2.2.1.8.8 Ligation reaction

Ligation was done with the Promega™ T4 ligase kit (Promega cat #9PIM180). A general ligation reaction composed of 1 X ligation buffer (40 mM Tris-HCl, pH 7.8; 10 mM $MgCl_2$; 10 mM DTT; 1 mM ATP), 100 ng vector DNA, 10 ng insert DNA, 1 U T4 ligase enzyme and autoclaved milliQ (18 Ω) water to a final volume of 20 µl. The reaction was incubated at 15 °C for 4 hours.

2.2.2 Reconstruction of the vervet monkey GLYAT open reading frame

2.2.2.1 Preparation of leukocytes from whole blood of vervet monkey

Fresh blood was obtained from a vervet monkey and collected in an anticoagulant ethylene diamine tetra-acetate (EDTA) containing tube and transported to the laboratory. An equal volume of histopaque was poured on top of the blood (without mixing) and centrifuged at 3000 x g for 10 min. Leukocytes will aggregate and be cushioned in the histopaque. Three layers were observed: the top layer had plasma, the intermediate layer had buffy coat (leukocyte-enriched fraction) and the bottom layer had concentrated erythrocytes. The buffy coat layer was transferred into a new 2 ml Eppendorf tube. Both the top and bottom layers were discarded. The buffy coat was washed with isotonic phosphate buffered saline (PBS) and centrifuged at 12 000 rpm for 5 min on a bench centrifuge (Eppendorf bench centrifuge 5415D). The pellets were used to extract total genomic DNA following the FlexiGene™ kit protocol for extracting genomic DNA from buffy coat.

2.2.2.2 Genomic DNA extraction from leukocytes of a vervet monkey

The FlexiGene® mammalian genomic DNA extraction kit (QIAGEN cat# 51206) was used to extract genomic DNA from leucocytes of vervet monkey (prepared in section 2.2.2.1). The method starts by lysing the leukocyte cells and sedimenting the cell nuclei and mitochondria. The pellets are resuspended and incubated in a denaturing buffer containing chaotropic salts (Chirgwin *et al*., 1979) and QIAGEN® protease. DNA is then recovered by isopropanol precipitation and washed with 70% ethanol (Boom *et al*., 1990). The DNA is resuspended in hydration buffer (10 mM Tris.Cl, pH 8.5). The total genomic DNA was separated on a 1 % agarose gel and its concentration determined on the NanoDrop spectrophotometer.

2.2.3 DNA sequencing

All DNA sequencing was done at the Central Analytical Facility (CAF) at the University of Stellenbosch, South Africa. The recombinant plasmid (pTZ57R/T with chacma baboon GLYAT open reading frame) was sequenced using the M13 primers. The vervet monkey GLYAT PCR amplicons of the various exons were sequenced with the same primer pairs which they were amplified with.

The resulting sequence data was analysed using DNASTAR® Lasergene 8.0 software suite for sequence analysis. This software was used to translate nucleic acid sequences

to amino acid sequences, align nucleic sequences and amino acid sequences and to construct phylogenetic trees.

2.3 Results

2.3.1 Extraction of total cellular RNA from chacma baboon liver

The objective of the experiment was to extract total RNA from liver tissue of a chacma baboon. A piece of 30 mg chacma baboon liver tissue was homogenised. Total RNA was extracted using a kit from QIAGEN™ (described in section 2.2.1.2) then the RNA was separated on a 1.2 % formaldehyde agarose denaturing gel (as described in section 2.2.1.4.2). The results are shown in Figure 2.4.

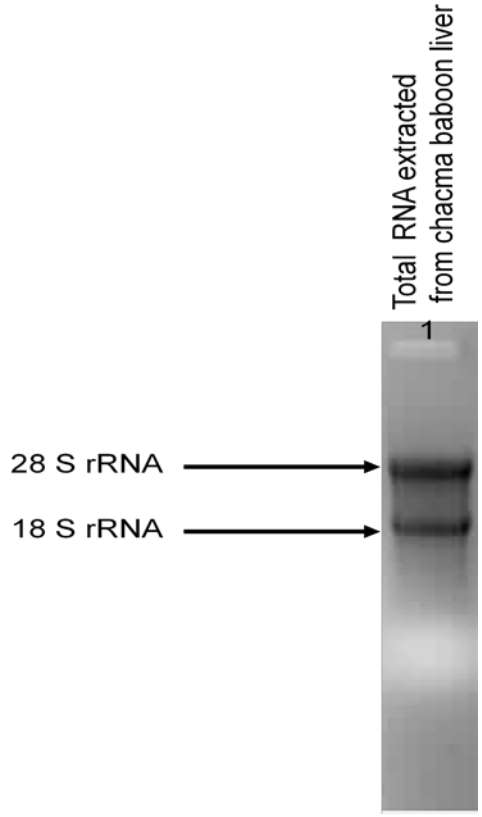


Figure 2.4: A 1.2 % formaldehyde agarose gel analysis of total RNA extracted from a liver of a chacma baboon

Humans and chacma baboons are classified as members of the primate family. Therefore, the total RNA of the chacma baboon was expected to show the same characteristics as that of human RNA. It is characteristic of mammalian total RNA to show two bands on the gel where the upper band appears to be denser than the lower band. The reason for this difference is that the 28 S rRNA binds to more ethidium bromide than the 18 S rRNA which results in the difference in fluorescence. The concentration of RNA in the total RNA preparation was found to be 108 ng/µl and since 80 µl was eluted it means that the yield was 8.64 µg and the A_{260}/A_{280} ratio was 1.9. The expected yield, as according to the manufacturer, was between 40 µg and 60 µg total RNA with an A_{260}/A_{280} ratio of 2. The $A_{260}/_{280}$ ratio of 1.9 is acceptable but the yield of 8.64 µg was poor relative to the expected amounts since this represents only 14 % of the expected total RNA amount. Lower yields do occur when the column is overloaded. The total RNA preparation was used to synthesise cDNA in the next step.

2.3.2 cDNA synthesis using oligo (dT) primers

A chacma baboon cDNA was synthesised using 0.9 µg of total RNA extracted from chacma baboon liver in a final reaction volume of 20 µl to synthesise cDNA (described in section 2.2.1.5). The cDNA was used in PCR to amplify an open reading frame encoding GLYAT of chacma baboon (section 2.2.1.6).

2.3.3 Amplification and gel extraction of an amplicon containing the open reading frame encoding GLYAT of the chacma baboon

The human GLYAT transcript (ENST00000344743) was used to design primers to amplify cDNA containing an ORF that encodes GLYAT of chacma baboon. Primers designed from human genes have been used successfully to amplify baboon DNA before (Hampf *et al*., 1996; Bowcock *et al*, 1989). Baboon antithrombin III cDNA has been used as a specific hybridisation probe to select for human antithrombin III cDNA (Chandra *et al*, 1983). The primer pair ORF FWD and ORF REV was used to amplify the ORF encoding GLYAT of chacma baboon (as illustrated in Figure 2.5 and primer sequence shown in Table 2.1). The start codon (ATG) was included in the forward primer sequence but the stop codon (TGA) was excluded in the reverse primer. The stop codon (TGA) was excluded because the PCR amplicon was going to be cloned in an expression vector that already had a stop codon (see chapter 3).

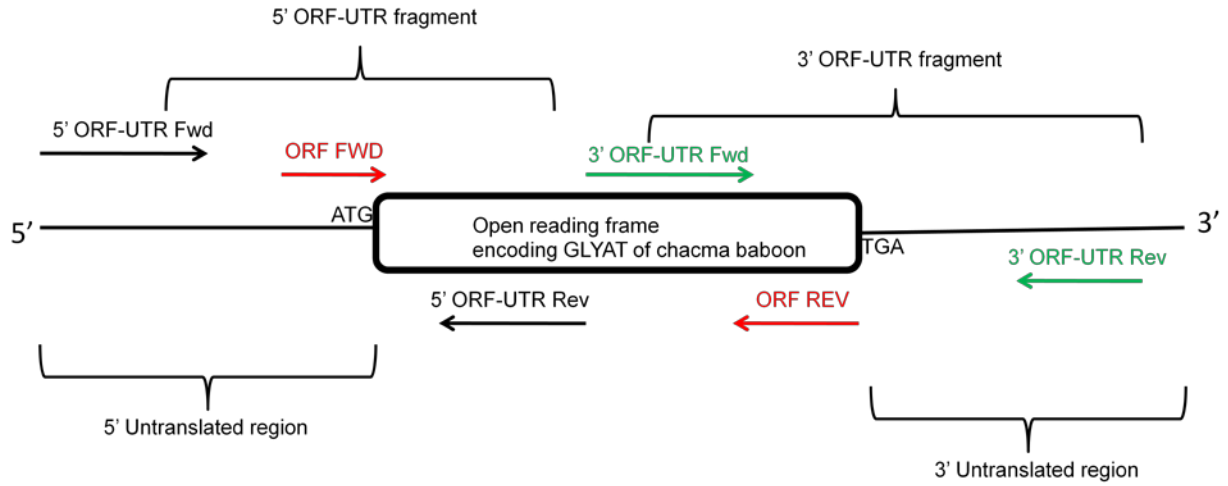


Figure 2.5: A schematic representation of GLYAT cDNA. This shows how primers were used to amplify the ORF encoding GLYAT and the two fragments used to confirm human primer sequence. Primers were paired as follows: ORF FWD and ORF REV(red); 5'ORF-UTR Fwd and 5' ORF-UTR Rev (black); 3' ORF-UTR Fwd and 3' ORF-UTR Rev (green).

A 5 µl aliquot cDNA synthesis reaction product was used in 50 µl PCR to set up an $MgCl_2$ concentration gradient PCR (as described in section 2.2.1.6). The purpose of the $MgCl_2$ gradient PCR was to establish the $MgCl_2$ concentration required for optimum PCR. An *Ex Taq* PCR buffer containing 1.5 mM $MgCl_2$ was used and more $MgCl_2$ was added gradually to vary the final concentration of the $MgCl_2$. Primers for the house keeping gene glyceraldehyde 3-phosphate dehydrogenase (GAPDH) were used as positive control generating a PCR product with 238 base pairs. The house keeping genes are those genes that are compulsory to all organisms and enable existence of an organism. The optimum $MgCl_2$ concentration was found to be 3 mM (results not shown). The products of the PCRs containing $MgCl_2$ final concentrations of 2 mM and 3 mM were gel extracted using the kit described in section 2.2.1.7. Then, a sample of each gel extract was analysed on a 1 % agarose gel (see Figure 2.6).

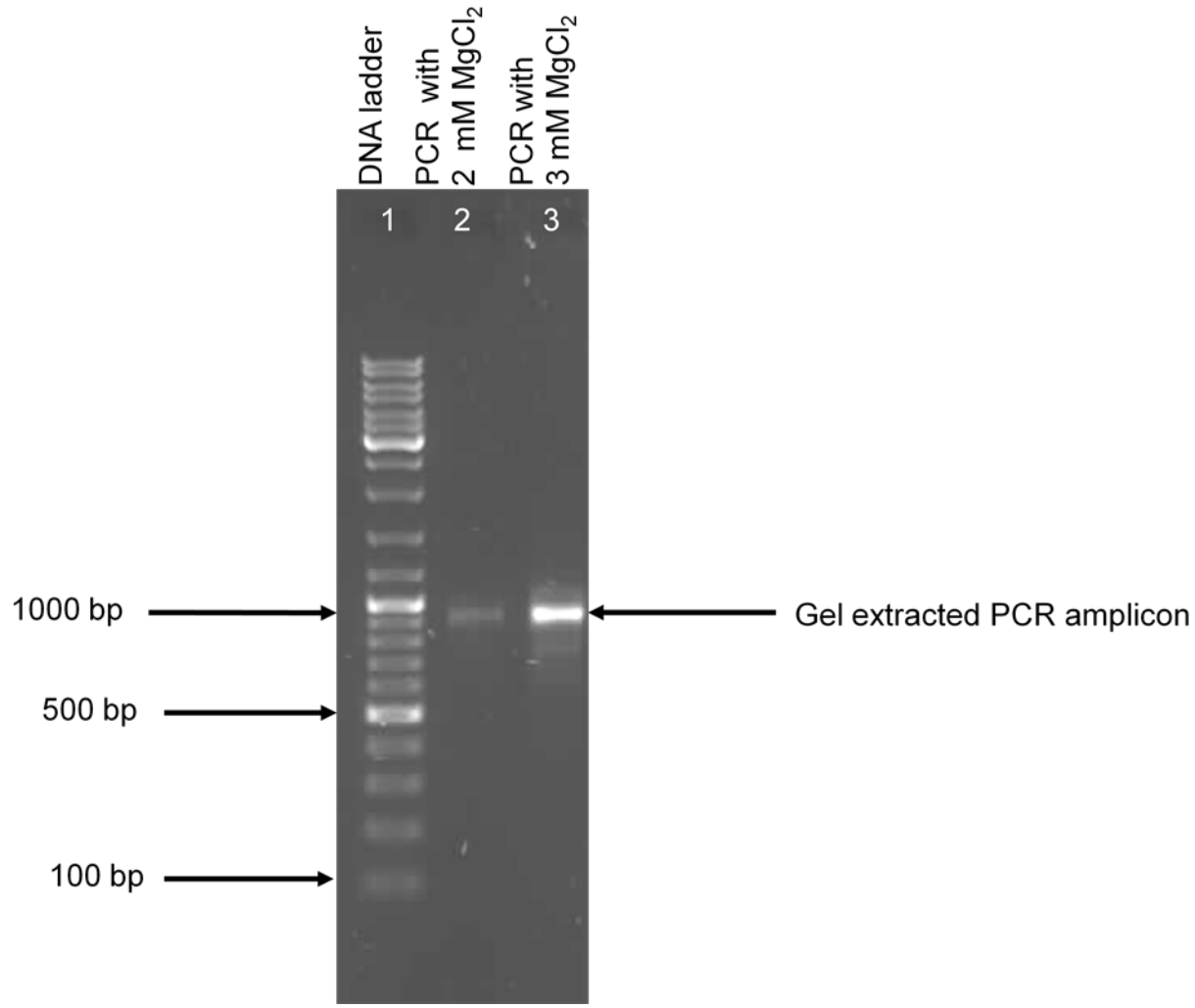


Figure 2.6: 1% Agarose gel analysis of the gel extracted PCR amplicons generated with GLYAT specific primers from total RNA isolated from a chacma baboon liver. Lanes (1) Fermentas O'GeneRuler DNA Ladder cat# SM1173 (2) Purified PCR amplicon prepared with 2 mM $MgCl_2$ (3) Purified PCR amplicon purified with 3 mM $MgCl_2$

The PCR products containing an ORF encoding GLYAT of chacma baboon appeared in lanes 2 and 3 as bands near the 900 bp band of the DNA ladder. DNA concentration of

PCR products after gel extraction of PCR with 2 mM $MgCl_2$ was 17 ng/µl giving a yield of 0.85 µg with an A_{260}/A_{280} ratio of 1.8. The reaction with 3 mM $MgCl_2$ gave DNA concentration of 25 ng/µl giving a yield of 1.25 µg and an A_{260}/A_{280} ratio of 1.8. The gel extracted PCR product in lane 2 appeared less visibly than the one in lane 3 due to the difference in amounts that were loaded. The same volume of 5 µl was loaded on both but due to difference in concentrations the final amount was different. PCR was done with *Ex Taq*™ DNA polymerase which produces PCR products with extra 3'-dA overhanging on each strand of the DNA. The TA vector has a 3'-ddT that is complementary to the PCR product 3'-dA overhang. The next step was to clone the PCR product into the TA vector and expression vector.

2.3.4 Cloning of a PCR amplicon encoding chacma baboon GLYAT open reading frame into a TA vector

Primers used to amplify the PCR product had restriction enzyme sites that were situated too close to the end of each primer for the enzyme to digest efficiently. Cloning the PCR product into a TA vector first enabled a much efficient restriction enzyme digestion. The PCR product was cloned into the TA vector using a commercial TA vector from Fermentas (as described in section 2.2.1.8.1). The resulting recombinant plasmid (pTZ57R/A ligated to the PCR product) was called pBabTAV001 (a confirmation restriction digestion is shown in Figure 2.9).

2.3.5 Transformation of pBabTAV001, pColdIII and pColdIII-EHDV into JM109 cells followed by plasmid midi preparation

The reason for this transformation was to make enough of the plasmid DNA for midi preparation. The pColdIII was obtained commercially in small amount and needed to be transformed to increase its amount. The pColdIII-EHDV was obtained in a small amount from Dr AC Potgieter containing an insert encoding for a viral protein of epizootic haemorrhagic disease virus (EHDV) and it too required to be transformed first. The pBabTAV001 required trasnformation before plasmid midi preparation could be done. The separate transformations were done as illustrated by the flow diagram in Figure 2.7.

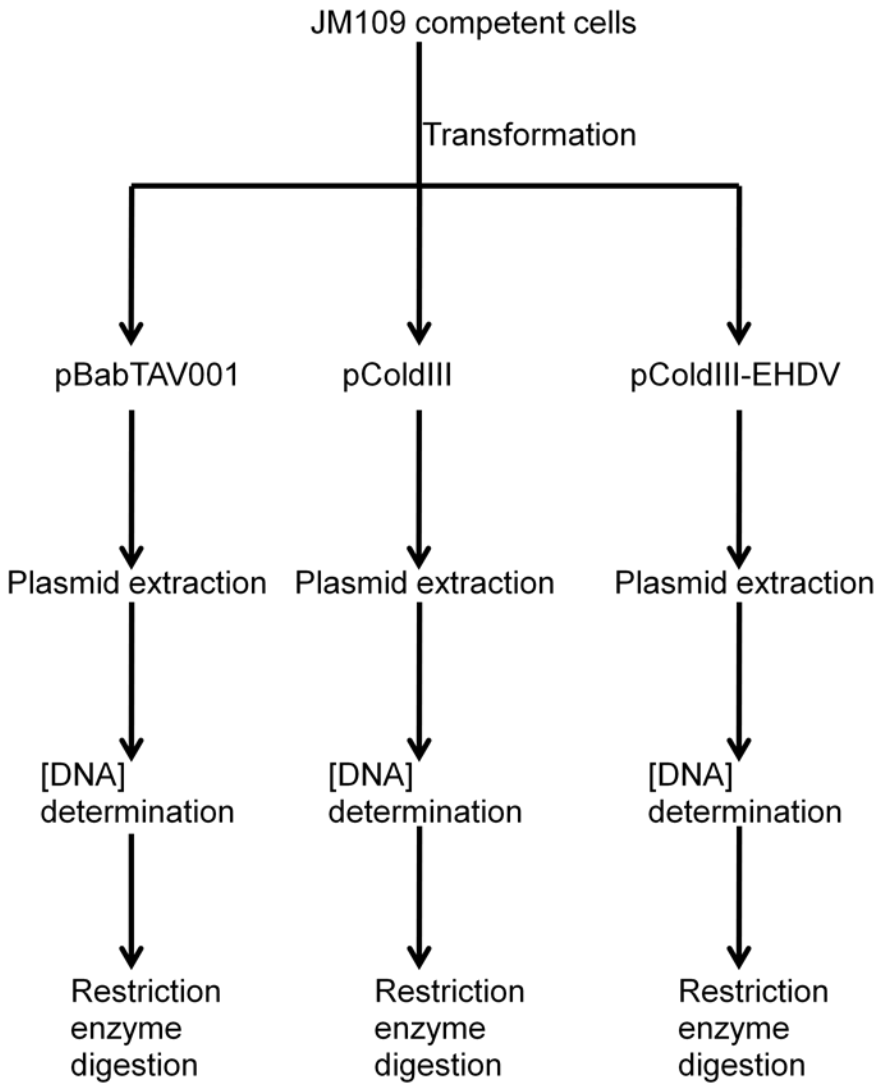


Figure 2.7: A flow diagram illustrating the number of transformation experiments conducted with different plasmids on JM109 cells.

Separate transformations of pBabTAV001, pColdIII and pColdIII-EHDV plasmids into competent JM109 cells, with a transformation efficiency of 1 x 10^7 cfu/ µg (described in section 2.2.1.8.3) was done as described in section 2.2.8.4. Plasmid midi preparation of all plasmids from JM109 cells was done using a kit described in section 2.2.1.8.6. The results of the transformation and DNA extraction experiments are shown in Table 2.2.

Table 2.2: A table illustrating results of the transformation of pBabTAV001, pColdIII and pColdIII-EHDV plasmids into JM109 cells

	Expected results	JM109 cells pBabTAV001	JM109 cells pColdIII	JM109 cells pCold-EHDV
Culture final volume (ml)	100	100	100	100
Culture density	Saturated	Saturated	Saturated	Saturated
DNA volume extracted (µl)	600	600	600	600
$A_{260/280}$ ratio	1.7	1.7	1.7	1.7
[DNA] extracted (ng/µl)	333	147.8	198	191.5
DNA yield (µg)	200	88.68	118.8	114.9

The results shown in Table 2.2 were compared with the expected results. The Promega PureYield™ plasmid midi preparation system is designed to purify 200 µg of plasmid DNA with an A_{260}/A_{280} ratio greater than 1.7 from a 100ml saturated culture. Therefore, all plasmid midi preparations gave a relatively pure DNA when compared with the expected DNA purity stated by the kit manufacture. The plasmid DNA yield for pColdIII and pColdIII-EHDV was relatively average as it was within the range stated by the manufacture of 100 to 200 µg. The next step was the restriction enzyme digestion of the extracted plasmids.

2.3.6 Restriction enzyme digestion of pBabTAV001, pColdIII and pColdIII-EHDV

The double restriction enzyme digestion of pBabTAV001, pColdIII and pColdIII-EHDV with NdeI and XhoI was done in order to: (1) separate the PCR product containing an ORF encoding GLYAT of chacma baboon from pBabTAV001; (2) linearise pColdIII in such a way that it is compatible to ligate the restricted PCR product (3) and separate the

insert (EHDV) from pColdIII-His-tag so that it can be substituted by chacma baboon GLYAT digested PCR amplicon. The additional restriction enzyme digestion of pBabTAV001, pColdIII and pColdIII-EHDV plasmids with the restriction enzymes BamHI, PstI and EcoRI was to confirm their plasmids restriction maps.

The general restriction enzyme digestion reaction set up is described in section 2.2.1.8.7. However, variations from the general set up are shown in Table 2.3.

Table 2.3: Restriction enzyme digestion of the extracted plasmid DNA

Component	**pBabTAV001 NdeI/XhoI**	**pBabTAV001 EcoRI**	**pBabTAV001 BamHI**	**pBabTAV001 PstI**	**pColdIII NdeI/XhoI**	**pCold III-EHDV Nde/XhoI**
H_20	To final volume	To final volume	To final volume	To final volume	To final volume	To final volume
Buffer O 10 X (50mM Tris-Cl; 10 mM MgCl2; 100 mM NaCl; 0.1 mg/ml BSA; pH 7.5)	1 X	None	None	1 X	1 X	1 X
Buffer BamHI 10 X (10 mM Tris-Cl (pH 8); 5 mM $MgCl_2$; 100 mM KCl; 0.02 % Triton X-100; 0.1 mg/ml BSA)	None	None	1X	None	None	None
Buffer EcoRI 10 X 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 100 mM NaCl, 0.02% Triton X-100, 0.1 mg/ml BSA	None	1 X	None	None	None	None
NdeI	60 U	None	None	None	60 U	60 U
XhoI	30 U	None	None	None	30 U	30 U
EcoRI	None	10 U	None	None	None	None
BamHI	None	None	10 U	None	None	None
PstI	None	None	None	10 U	None	None
DNA	5 µg	1 µg	1 µg	1 µg	5 µg	5 µg
Final volume	**84 µl**	**20 µl**	**20 µl**	**20 µl**	**84 µl**	**84 µl**

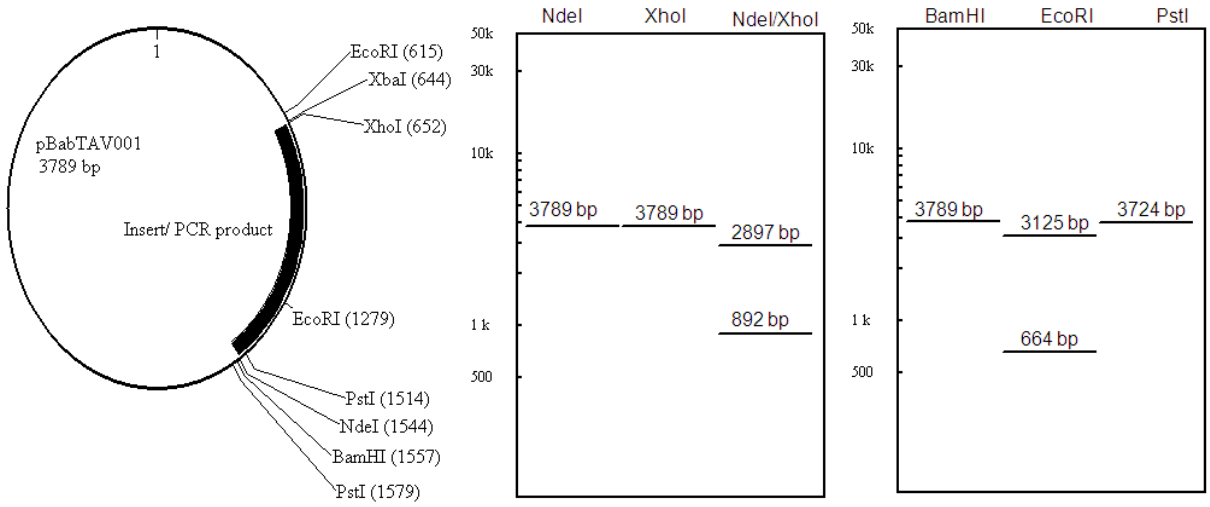


Figure 2.8: A diagramatic representation produced using DNAMAN™ computer software illustrating the pBabTAV001 plasmid map with the expected restriction pattern by the enzymes NdeI, XhoI, BamHI, EcoRI and PstI.

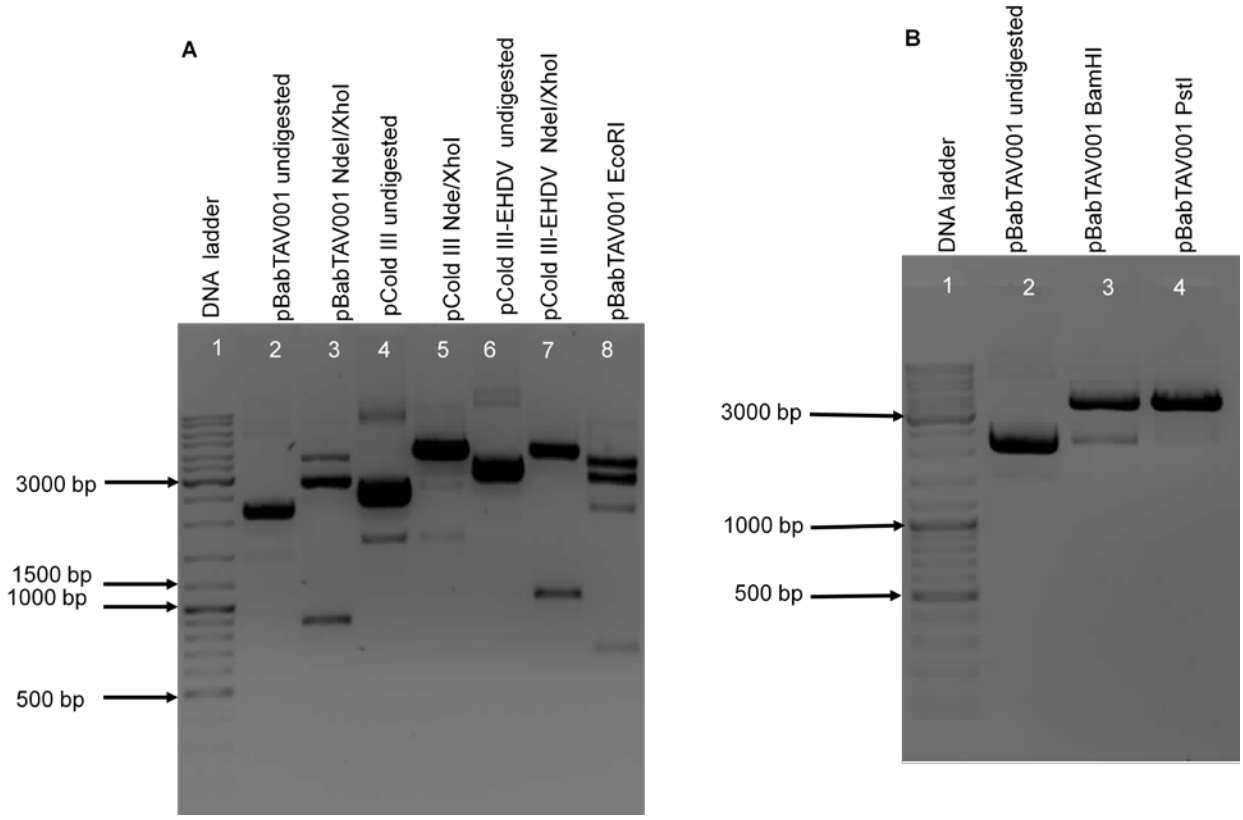


Figure 2.9: 1% Agarose gel electrophoresis of restriction enzyme digestion of pBABTAV001 and pColdIII-EDHV. (A) Lanes: (1) Fermentas DNA ladder cat# SM1173. (2) Undigested pBabTAV001. (3) pBabTAV001 digested with NdeI and XhoI. (4) pCold III undigested. (5) pCold III digested by NdeI and XhoI. (6) undigested pColdIII-EHDV. (7) pCold III–EHDV digested with NdeI and XhoI (8) pBabTAV001 digested with EcoRI. (B) Lanes: (1) Fermentas DNA ladder cat# SM1173. (2) pBabTAV001 undigested. (3) pBabTAV001 linearised by digestion with BamHI (4) pBabTAV001 linearised with PstI.

Figure 2.8 illustrates the expected results after restriction enzyme digestion of pBabTAV001 as shown in Table 2.3. In Figure 2.9(A), the actual gel picture showing the results after restriction enzyme digestion of pBabTAV001, pColdIII and pColdIII-EHDV. Lane 1 shows DNA ladder which was supposed to show 21 bands on the gel instead only 17 appear. The fading of bands on the gel could have been caused by heating up

of the gel due to a prolonged electrophoresis, which led to loss of intercalating agent by the leading bands. Lane 2 was loaded with undigested pBabTAV001 as a negative control for the restriction enzyme digestion experiment. The results of the double restriction enzyme digestion of pBabTAV001 with NdeI and XhoI are shown in Figure 2.9 (A) lane 3 and show consistency with the predicted results shown in Figure 2.8 (B). In addition to the two expected bands (892 bp and 2897 bp) there was a third band which was a result of incomplete digestion [Figure 2.8 (A) shown as size 3789 bp]. Lane 4 had undigested pColdIII serving as a negative control to see any uncompleted digestions. Lane 5 had the product of the restriction enzyme digestion of pColdIII with NdeI and XhoI where the pColdIII was separated from its 20 bp located between NdeI and XhoI restriction sites (see pColdIII multiple cloning site in Figure 2.3). Lanes 6 and 7 had pColdIII-EHDV undigested and digested with NdeI and XhoI respectively. After the pColdIII-EHDV has been digested the EHDV was separated and the free pColdIII with His-tag (was named pColdIII-His and size was estimated using DNAMAN to be 4432 bp) and had 17 amino acid repeats of serine-glycine linking the nucleotide sequence that representing the His-tag. The serine-glycine linker starts with a glutamic acid (E) and ends on a serine (S) (Figure 3.1 A). Lane 8 had the pBabTAV001 digested with EcoRI. The EcoRI restriction enzyme restricts pTZ57R at a single position and the GLYAT open reading frame once producing two fragments (664 bp and 3125 bp) which is comparable with expected results shown in Figure 2.8 (C). It appears that there was some undigested pBabTAV001 DNA and this is shown by two extra bands: one band represents the completely undigested DNA which is comparable to lane 2 and the other

band seems to be linearised pBabTAV001 which could have resulted from the EcoRI enzyme restricting only one site.

The rest of the restriction enzyme digestion products were analysed on a separate 1 % agarose gel shown in Figure 2.9 (B). The DNA fragments of the DNA ladder in lane 1 appear to be very faint and this could be resolved by increasing the amount of the DNA ladder loaded on the gel. Lanes 3 and 4 had pBabTAV001 DNA restricted with the restriction enzymes BamHI and PstI respectively. BamHI only linearised the pBabTAV001 DNA and this is deduced from the undigested fragment which coincides with the undigested plasmid in lane 2. The PstI has two restriction sites on the pBabTAV001 which are 65 bases apart. After the complete digestion only a fragment of 3724 bp was visible and the 65 bp fragments was not visible because the fragment was very small and moved faster to the end of the gel. The next step was to clone the PCR amplicon containing the open reading frame encoding GLYAT of chacma baboon into pColdIII and pColdIII-His. The detailed cloning experiment will be discussed in the next chapter.

However, the position where human primers annealed required confirmation by DNA sequencing. The next step was to sequence that human primer region and then the insert in the multiple cloning site of pBabTAV001 DNA.

2.4. Amplification of a fragment across the position where human primers were bound on the chacma baboon GLYAT cDNA to confirm human primer sequence by DNA sequencing

The expression of a chacma baboon recombinant GLYAT will be reported in the next chapter (Chapter 3). Thus, it was absolute necessary to make certain that every base of the open reading frame was indeed that of chacma baboon. The human primers used to amplify the chacma baboon GLYAT open reading frame were bound at the beginning of the open reading frame. This prevented a few bases from being sequenced (see primer pair ORF FWD and ORF REV illustrated in Figure 2.5). There were 29 bases (ORF REV reverse primer) at the 3'-end of the ORF of chacma baboon GLYAT that were not accounted for during the DNA sequencing of the ORF. It is commonly known that primers can bind to template even when there are one or two base mismatches. The reason for sequencing the chacma baboon transcript, as far as possible, was to sequence across the region were the human primers were bound.

The logic is to get some sequence both before and after the human primer binding sites by so doing it will be possible to gain confidence that the human primer sequence is exactly the same as that of the chacma baboon. The primers used were designed from the human GLYAT transcript (NM_201648) and they are illustrated in Figure 2.5. The 5'-region of the open reading frame encoding GLYAT of chacma baboon was amplified using the primer pair 5'-UTR fwd and 5'-UTR Rev. In addition, the 3'-region was amplified with the primer pair: 3'-UTR fwd and 3'-UTR Rev. GAPDH primers were used to control PCR. The results of the PCR are shown in Figure 2.10.

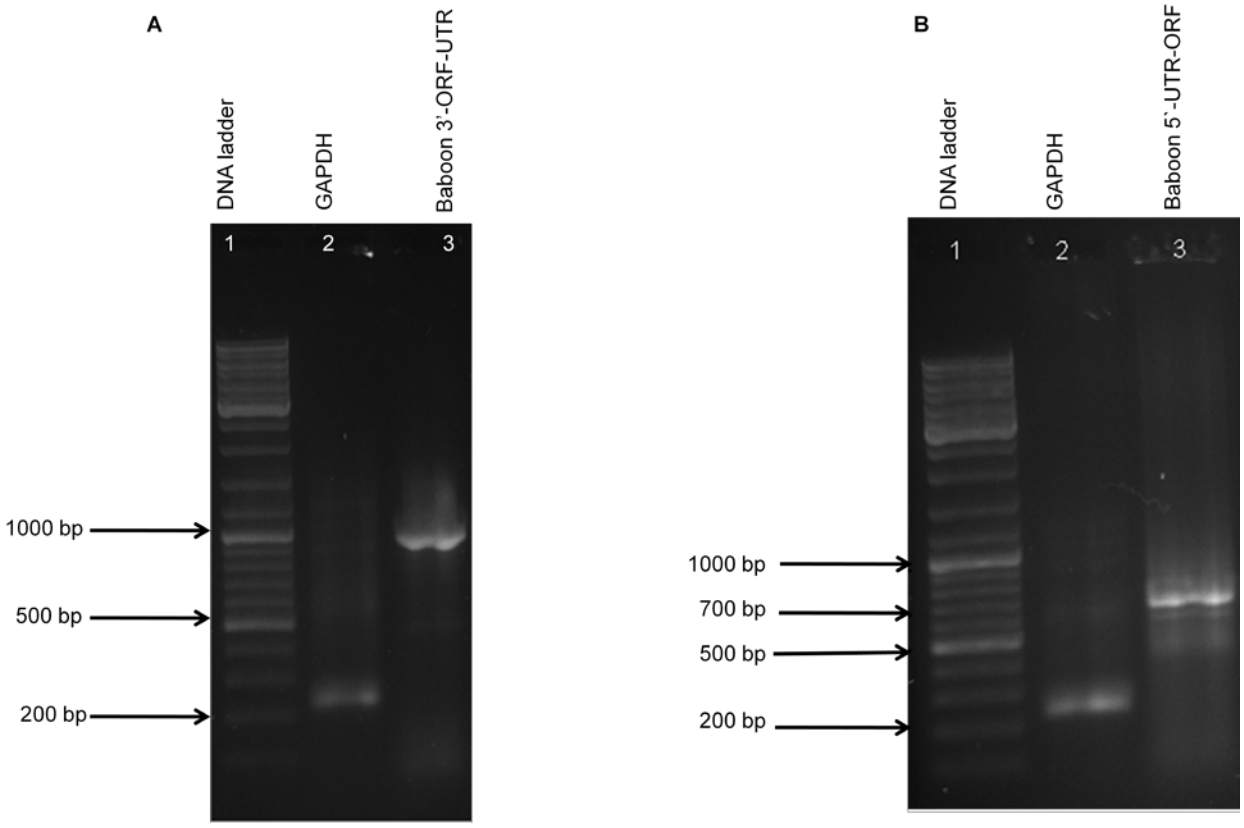


Figure 2.10: 1% Agarose electrophoresis of the resulting fragments after amplification of the 3'-fragment (A) and 5'-fragment (B) which includes part of the open reading frame and an untranslated region (UTR) of the cDNA encoding for GLYAT of chacma baboon. Lanes: (A1) Fermentas™ DNA ladder cat# SM1173 (A2) GAPDH (A3) chacma baboon GLYAT open reading frame-UTR 3'-fragment (B1) Fermentas™ DNA ladder cat# SM1173 (B2) GAPDH positive control (B3) chacma baboon GLYAT open reading frame-UTR 5'-fragment**.**

The 3'-UTR Rev primer bound 682 bp downstream of the stop codon while its forward primer (3'-UTR fwd) bound 225 bp upstream of the stop codon. The two primers amplify a DNA fragment of 907 bp (A3). The 5'-UTR fwd primer bound 137 bp upstream of the starting codon while the 5'-UTR Rev primer bound 528 bp downstream of the start codon. The two primers amplify a DNA fragment of 665 bp (B3). The 665 bp fragment

(B3) appears on the gel near the 700 bp fragment relative to the DNA ladder (lane B1). The DNA sequence of the primers is shown in Table 2.1. GAPDH fragment in both gels shown in gels [Figure 2.10 (A2) and (B2)] amplified successfully. The success in GAPDH amplification meant that the PCR was working. The PCR product in lanes A3 and B3 in Figure 2.10 show a small smear which will not make the results any less reliable since these PCR products were gel extracted (described in section 2.2.1.7) and DNA sequenced. The sequenced produced was combined with the sequence of the open reading frame to reconstruct a transcript encoding GLYAT of chacma baboon.

The next experiment was intended to sequence and reconstruct the vervet monkey GLYAT open reading frame from separate exons.

2.5 The extraction of genomic DNA from leucocytes of vervet monkey

The whole genome of the vervet monkey has not been sequenced, but some individual genes have been sequenced and reported in GenBank. The GLYAT gene of the vervet monkey has also not been sequenced. Blood was obtained from a living vervet monkey. Genomic DNA was extracted from the buffy coat. A final concentration of 93 ng/µl total genomic DNA with a yield of 9300 ng and purity of 1.8 was extracted. The human GLYAT gene was used to design primers that were used to amplify vervet monkey GLYAT exons. After, vervet monkey GLYAT exons were amplified then gel extracted using the gel extraction kit (described in section 2.2.1.7). The gel extracted PCR products containing exons encoding parts of GLYAT of the vervet monkey were separated on a 1% agarose gel and the picture of the gel is shown in Figure 2.11.

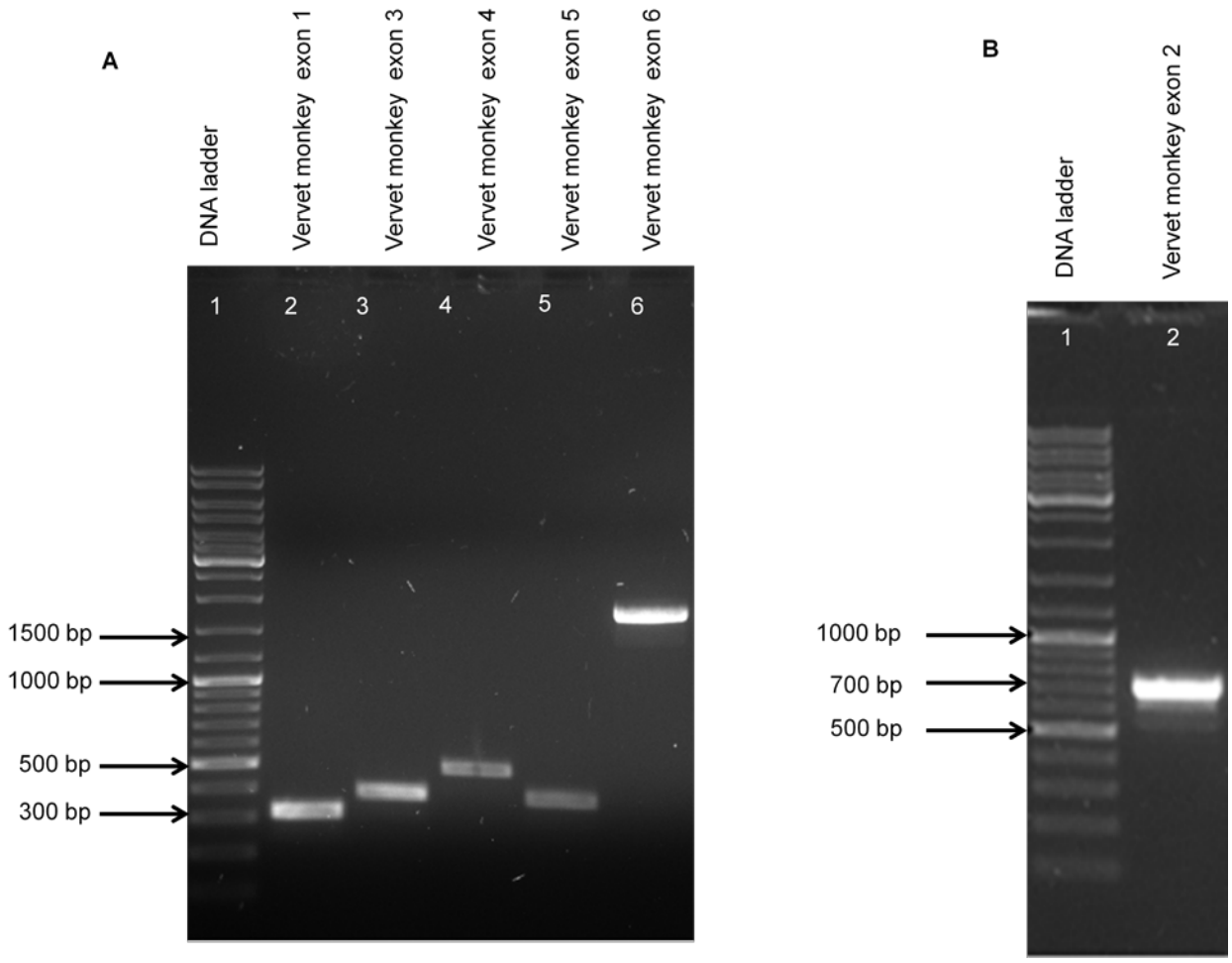


Figure 2.11: Gel extraction of PCR products containing (A) exons 1, 3, 4, 5 and 6 (B) exon 2 coding for GLYAT of vervet monkey. Fragment sizes were predicted using DNAMAN™ software. Lanes: (A1) Fermentas DNA ladder cat# SM1173 (A2) Exon 1 (A3) Exon 3 (A4) Exon 4 (A5) Exon 5 (A6) Exon 6 (B1) Fermentas DNA ladder cat#SM1173 (B2) exon 2.

The PCR primers were designed from the human GLYAT gene to amplify vervet monkey exons (Figure 2.11 A and B) to yield the following fragments: exon1 = 378 bp, exon2 = 700 bp, exon3 = 420 bp, exon4 = 490 bp, exon5 = 420 bp and exon6 = 1625 bp. All primers annealed on the neighbouring introns to the amplified exons.

The next step was to sequence the gel extracted vervet monkey PCR amplicons.

2.6 DNA sequencing

DNA sequencing of the chacma baboon GLYAT open reading frame, chacma baboon GLYAT cDNA UTR and vervet monkey exons was done using the chain termination method (Sanger *et al*., 1977). There has not been any report on the GLYAT gene of the vervet monkey in literature. Total genomic DNA of vervet monkey was extracted from buffy coat obtained from vervet monkey blood. Human GLYAT gene was used as a reference to design primers that must bind on the vervet monkey GLYAT gene introns in order to amplify the exons. All six of the vervet monkey PCR amplified exon-intron fragments were sequenced with the same primers they were amplified with. Human exons were aligned to the vervet monkey exon-intron PCR fragments. Then the exon-intron junction patterns were located on the sequences in order to trim down vervet monkey exons. The exon-intron patterns that were searched for were: AG/GTA; /GTAAGT; R G/GTGAG(R = A or G); AG/GTXXGT (X = A, T, G or C) (Iida and Sasaki, 1983). The vervet monkey GLYAT exons were joined to form a continuous vervet monkey GLYAT open reading frame. Figure 2.12 and Figure 2.13 show the achieved DNA sequences of the chacma baboon and vervet monkey respectively. Known human SNP sites that have been proposed to have a potential effect on GLYAT activity are indicated (Ser17Thr and Arg199Cys).

```
          GC TCT GAG ACG ACT TGG CTG ACG TAG TCC CTC TTT AAA GTA GAG GGT C
1         CG AGA CTC TGC TGA ACC GAC TGC ATC AGG GAG AAA TTT CAT CTC CCA G

          TT CCC ACA CAC GCC GAG TAG TAA AGA AGG ACC TTT GTA GAC GTC TCT G
49        AA GGG TGT GTG CGG CTC ATC ATT TCT TCC TGG AAA CAT CTG CAG AGA C

          AA CGA AAA GTC CGA TTC CAT AGG AGG TAC TAC AAT GGT AAT GTT CCA C
97        TT GCT TTT CAG GCT AAG GTA TCC TCC ATG ATG TTA CCA TTA CAA GGT G
 1                                       MET MET Leu Pro Leu Gln Gly

                                      Ser17Thr

          GG GTC TAC GAC GTC TAC GAC CTC TTT AGG AAC TCC TTC TCG GAG GGT C
145       CC CAG ATG CTG CAG ATG CTG GAG AAA TCC TTG AGG AAG AGC CTC CCA G
 8      Ala Gln MET Leu Gln MET Leu Glu Lys Ser Leu Arg Lys Ser Leu Pro

          GT AGG AAT TTC CAA ATA CCT TGG CAG AAA GTG TAT TTG GTG CCT TTA G
193       CA TCC TTA AAG GTT TAT GGA ACC GTC TTT CAC ATA AAC CAC GGA AAT C
24     Ala Ser Leu Lys Val Tyr Gly Thr Val Phe His Ile Asn His Gly Asn

          GC AAG TTA GAC TTC CGA CAC CAC CTG TTC ACC GGA CTA AAA TTA TGT C
241       CG TTC AAT CTG AAG GCT GTG GTG GAC AAG TGG CCT GAT TTT AAT ACA G
40     Pro Phe Asn Leu Lys Ala Val Val Asp Lys Trp Pro Asp Phe Asn Thr

          AC CAA CAG ACG GGA GTC CTC GTC CTA TAC TGT CTA CTG GAA CTA GTG A
289       TG GTT GTC TGC CCT CAG GAG CAG GAT ATG ACA GAT GAC CTT GAT CAC T
```

```
56       Val Val Val Cys Pro Gln Glu Gln Asp MET Thr Asp Asp Leu Asp His

          TA TGG TTA TGA ATG GTT TAG ATG AGG TTT CTG GAG GTT TTG ACA GTC C
337       AT ACC AAT ACT TAC CAA ATC TAC TCC AAA GAC CTC CAA AAC TGT CAG G
72       Tyr Thr Asn Thr Tyr Gln Ile Tyr Ser Lys Asp Leu Gln Asn Cys Gln

          TT AAG GAA CCA AGT GGT CTT GAA TAG TTA ACC TTT GTC GTA AAT GTC T
385       AA TTC CTT GGT TCA CCA GAA CTT ATC AAT TGG AAA CAG CAT TTA CAG A
88       Glu Phe Leu Gly Ser Pro Glu Leu Ile Asn Trp Lys Gln His Leu Gln

          AA GTT TCA AGT GTC GGA TCG GAC TTA CTC TGA TAT GTT TTA GAA CGT C
433       TT CAA AGT TCA CAG CCT AGC CTG AAT GAG ACT ATA CAA AAT CTT GCA G
104      Ile Gln Ser Ser Gln Pro Ser Leu Asn Glu Thr Ile Gln Asn Leu Ala

          GG TAA TTC AGG AAG TTT CAG TTT GTG TGT GTT ACG TAA GAG ATA TAC C
481       CC ATT AAG TCC TTC AAA GTC AAA CAC ACA CAA TGC ATT CTC TAT ATG G
120      Ala Ile Lys Ser Phe Lys Val Lys His Thr Gln Cys Ile Leu Tyr MET

          GT AGG CTT TGT CGG TTC CTT GAC CGA GGA AAG GAC GAC CTT AGT TTC T
529       CA TCC GAA ACA GCC AAG GAA CTG GCT CCT TTC CTG CTG GAA TCA AAG A
136      Ala Ser Glu Thr Ala Lys Glu Leu Ala Pro Phe Leu Leu Glu Ser Lys

          GT AAT AGA GGG TCA CCA CCG TTT GGG TTC CGG TAG TTG GTT CTC TAC A
577       CA TTA TCT CCC AGT GGT GGC AAA CCC AAG GCC ATC AAC CAA GAG ATG T
152      Thr Leu Ser Pro Ser Gly Gly Lys Pro Lys Ala Ile Asn Gln Glu MET

          AA TTT GAG AGT AGG TAC CTA CAA TTG ATA CGA GTG AAC CAC TTA TTT A
625       TT AAA CTC TCA TCC ATG GAT GTT AAC TAT GCT CAC TTG GTG AAT AAA T
```

168 **Phe Lys Leu Ser Ser MET Asp Val Asn Tyr Ala His Leu Val Asn Lys**

Arg199Cys

AG ACC ATA AAA CCA CCA TTA CTC TCC TCG GTC TCT AAG TAA CTC GCG A

673 TC TGG TAT TTT GGT GGT AAT GAG AGG AGC CAG AGA TTC ATT GAG CGC T

184 **Phe Trp Tyr Phe Gly Gly Asn Glu Arg Ser Gln Arg Phe Ile Glu Arg**

CG TAA GTC TGG AAA GGG TGG TCG ACA GAG AAC CCC GGA CTC CCT TGG G

721 GC ATT CAG ACC TTT CCC ACC AGC TGT CTC TTG GGG CCT GAG GGA ACC C

200 **Cys Ile Gln Thr Phe Pro Thr Ser Cys Leu Leu Gly Pro Glu Gly Thr**

GA CAC ACG ACC TTA GAT TAC CTG GTG TGA CCT CTC TAC TCT TAC CGT C

769 CT GTG TGC TGG AAT CTA ATG GAC CAC ACT GGA GAG ATG AGA ATG GCA G

216 **Pro Val Cys Trp Asn Leu MET Asp His Thr Gly Glu MET Arg M*Et ala***

CG TGG AAC GGA CTT ATG GCC AAG GCC CCA GAA CAC TGA ATA CAG TAG A

817 GC ACC TTG CCT GAA TAC CGG TTC CGG GGT CTT GTG ACT TAT GTC ATC T

232 **Gly Thr Leu Pro Glu Tyr Arg Phe Arg Gly Leu Val Thr Tyr Val Ile**

TA GTG GTG CGG GTC TTA AAC CGG TTT GAA CCG AAA GGA CAG ATA AGA G

865 AT CAC CAC GCC CAG AAT TTG GCC AAA CTT GGC TTT CCT GTC TAT TCT C

248 **Tyr His His Ala Gln Asn Leu Ala Lys Leu Gly Phe Pro Val Tyr Ser**

TA CAT CTG ATG TCG TTA CTT CGA TAC GTT TTT TAC TCA ATG GTT GAC G

913 AT GTA GAC TAC AGC AAT GAA GCT ATG CAA AAA ATG AGT TAC CAA CTG C

264 **His Val Asp Tyr Ser Asn Glu Ala MET Gln Lys MET Ser Tyr Gln Leu**

```
          TT GTA CAA GGG TAA GGG TCT TCG ACC TTG GTC ACC TTG ACA CAC GGA G
961       AA CAT GTT CCC ATT CCC AGA AGC TGG AAC CAG TGG AAC TGT GTG CCT C
280      Gln His Val Pro Ile Pro Arg Ser Trp Asn Gln Trp Asn Cys Val Pro

          AC ACT ACG GTT AGG ACT TGT ATT CTG TCA CAA CCC GTC CAG ACC CGT A
1009      TG TGA TGC CAA TCC TGA ACA TAA GAC AGT GTT GGG CAG GTC TGG GCA T
296      Leu ***

          CA TTA ACC TCC TCA CCT ACC ACC TAC TTT TCC TTC TTA TTT AAC ATT A
1057      GT AAT TGG AGG AGT GGA TGG TGG ATG AAA AGG AAG AAT AAA TTG TAA T

          GT CGT CAT TTC CTC ACC CGT GAC AAA CCC GAG ACC CCT TCG TCA CAC T
1105      CA GCA GTA AAG GAG TGG GCA CTG TTT GGG CTC TGG GGA AGC AGT GTG A

          AC AAG TTG TCC TAG CGG TAC CAG GGA CGT AAG TGT CCA AAG AGT CAC C
1153      TG TTC AAC AGG ATC GCC ATG GTC CCT GCA TTC ACA GGT TTC TCA GTG G

          CT TCC TCC GGG TCC AGG AGA ATG TAA ATT GGA TAC GTT AAA TCG AGG G
1201      GA AGG AGG CCC AGG TCC TCT TAC ATT TAA CCT ATG CAA TTT AGC TCC C

          GT GTA GGT
1249      CA CAT CCA
```

Figure 2.12: Double stranded DNA sequence and the deduced amino acid sequence of the chacma baboon GLYAT transcript.

The cDNA was prepared and amplified from the mRNA encoding GLYAT of the chacma baboon. The chacma baboon GLYAT amplicon was sequenced. A nucleic acid

sequence of 1256 bases was obtained representing the chacma baboon transcript. The chacma baboon transcript contained an open reading frame of 891 bases that translated into 296 amino acids (Figure 2.12). The transcript had a 122 base long DNA region extending the 5'-end of the chacma baboon GLYAT ORF. There was a 243 base long DNA region elongating the 3'-end of the chacma baboon GLYAT ORF. These results are limited because the primers were designed relative to the human GLYAT transcript which means that the length of the entire chacma baboon GLYAT transcript is still unknown.

```
          TAC TAC AAT GGT AAT GTT CCA CGG GTC TAC GAC GTC TAC GAC CTC TTT
1         ATG ATG TTA CCA TTA CAA GGT GCC CAG ATG CTG CAG ATG CTG GAG AAA
1         MET MET Leu Pro Leu Gln Gly Ala Gln MET Leu Gln MET Leu Glu Lys
```

Ser17Thr

```
          AGG AAC TCC TTC TCG GAG GGT CGT AGG AAT TTC CAA ATA CCT TGG TAG
49        TCC TTG AGG AAG AGC CTC CCA GCA TCC TTA AAG GTT TAT GGA ACC ATC
17        Ser Leu Arg Lys Ser Leu Pro Ala Ser Leu Lys Val Tyr Gly Thr Ile

          AAA GTG TAT TTG GTA CCT TTA GGC AAA TTA GAC TTC CGA CAC CAC CTG
97        TTT CAC ATA AAC CAT GGA AAT CCG TTT AAT CTG AAG GCT GTG GTG GAC
33        Phe His Ile Asn His Gly Asn Pro Phe Asn Leu Lys Ala Val Val Asp

          TTC ACC GGA CTA AAA TTA TGT CAC CAA CAG ACG GGA GTC CTC GTC CTA
145       AAG TGG CCT GAT TTT AAT ACA GTG GTT GTC TGC CCT CAG GAG CAG GAT
49        Lys Trp Pro Asp Phe Asn Thr Val Val Val Cys Pro Gln Glu Gln Asp

          TAC TGT CTA CTG GAA CTA GTG ATA TGG TTA TGA ATG GTT TAG ATG AGG
193       ATG ACA GAT GAC CTT GAT CAC TAT ACC AAT ACT TAC CAA ATC TAC TCC
65        MET Thr Asp Asp Leu Asp His Tyr Thr Asn Thr Tyr Gln Ile Tyr Ser



          TTT CTG GAG GTT TTG ACA GTC CTT AAA GAA CCT AGT GGT CTT GAG TAG
```

241 AAA GAC CTC CAA AAC TGT CAG GAA TTT CTT GGA TCA CCA GAA CTC ATC

81 **Lys Asp Leu Gln Asn Cys Gln Glu Phe Leu Gly Ser Pro Glu Leu Ile**

TTA ACC TTT GTC GTA AAT GTC TAA GTT TCA AGT GTC GGA TCG GAC TTA

289 AAT TGG AAA CAG CAT TTA CAG ATT CAA AGT TCA CAG CCT AGC CTG AAT

97 **Asn Trp Lys Gln His Leu Gln Ile Gln Ser Ser Gln Pro Ser Leu Asn**

CTC TGA TAT GTT TTA GAA CGT CGG TAA TTC AGG AAG TTT CAG TTT GTG

337 GAG ACT ATA CAA AAT CTT GCA GCC ATT AAG TCC TTC AAA GTC AAA CAC

113 **Glu Thr Ile Gln Asn Leu Ala Ala Ile Lys Ser Phe Lys Val Lys His**

TGT GTT ACG TAA GAG ATA TAC CGT AGG CTT TGT CGG TTC CTT GAC CGA

385 ACA CAA TGC ATT CTC TAT ATG GCA TCC GAA ACA GCC AAG GAA CTG GCT

129 **Thr Gln Cys Ile Leu Tyr M*Et al*a Ser Glu Thr Ala Lys Glu Leu Ala**

GGA AAG GAC GAC CTT AGT TTC TGT AAT AGA GGG TCA CCA CCG TTT GGG

433 CCT TTC CTG CTG GAA TCA AAG ACA TTA TCT CCC AGT GGT GGC AAA CCC

145 **Pro Phe Leu Leu Glu Ser Lys Thr Leu Ser Pro Ser Gly Gly Lys Pro**

TTC CGG TAG TTG GTT CTC TAC AAA TTT GAG AGT AGG TAC CTA CAA TTG

481 AAG GCC ATC AAC CAA GAG ATG TTT AAA CTC TCA TCC ATG GAT GTT AAC

161 **Lys Ala Ile Asn Gln Glu MET Phe Lys Leu Ser Ser MET Asp Val Asn**

ATA CGA GTG AAC CAC TTA TTT AAG ACC ATA AAA CCA CCA TTA CTC TCC

529 TAT GCT CAC TTG GTG AAT AAA TTC TGG TAT TTT GGT GGT AAT GAG AGG

177 **Tyr Ala His Leu Val Asn Lys Phe Trp Tyr Phe Gly Gly Asn Glu Arg**

Arg199Cys

```
        TCG GTC TCT AAG TAA CTC GCG ACG TAA GTC TGG AAA GGG TGG TCG ACA
577     AGC CAG AGA TTC ATT GAG CGC TGC ATT CAG ACC TTT CCC ACC AGC TGT
193     Ser Gln Arg Phe Ile Glu Arg Cys Ile Gln Thr Phe Pro Thr Ser Cys

        GAG AAC CCC GGA CTC CCT TGG GGA CAC ACG ACC TTA GAT TAC CTG ATG
625     CTC TTG GGG CCT GAG GGA ACC CCT GTG TGC TGG AAT CTA ATG GAC TAC
209     Leu Leu Gly Pro Glu Gly Thr Pro Val Cys Trp Asn Leu MET Asp Tyr

        TGA CCT CTC TAC TCT TAC CGT CCG TGG AAC GGA CTT ATG GCC AAG GTC
673     ACT GGA GAG ATG AGA ATG GCA GGC ACC TTG CCT GAA TAC CGG TTC CAG
225     Thr Gly Glu MET Arg MEt ala Gly Thr Leu Pro Glu Tyr Arg Phe Gln

        CCA GAA CAC TGA ATA CAG TAG ATA GTG GTG CGG GTC TTA AAC CGG TTT
721     GGT CTT GTG ACT TAT GTC ATC TAT CAC CAC GCC CAG AAT TTG GCC AAA
241     Gly Leu Val Thr Tyr Val Ile Tyr His His Ala Gln Asn Leu Ala Lys

        GAA CCG AAA GGA CAG ATA AGA GTA CAT CTG ATG TCG TTA CTT CGA TAC
769     CTT GGC TTT CCT GTC TAT TCT CAT GTA GAC TAC AGC AAT GAA GCT ATG
257     Leu Gly Phe Pro Val Tyr Ser His Val Asp Tyr Ser Asn Glu Ala MET

        GTT TTT TAC TCA ATG GTT GAC GTT GTA CAA GGG TAA GGG TCT TCG ACC
817     CAA AAA ATG AGT TAC CAA CTG CAA CAT GTT CCC ATT CCC AGA AGC TGG
273     Gln Lys MET Ser Tyr Gln Leu Gln His Val Pro Ile Pro Arg Ser Trp
```

```
            TTG GTC ACC TTG ACA CAC GGA GAC ACT
865         AAC CAG TGG AAC TGT GTG CCT CTG TGA
289         Asn Gln Trp Asn Cys Val Pro Leu ***
```

Figure 2.13: Double-stranded DNA sequence and the deduced amino acid sequence of the vervet monkey GLYAT open reading frame

The vervet monkey GLYAT open reading frame was 891 bases long and translated into 296 amino acids. Only an open reading frame for GLYAT of vervet monkey was generated in this experiment because it was sequenced from the genome and not the cDNA. The primers used to amplify the vervet monkey GLYAT exons annealed on the introns. The transcript which is formed by extending the ORF does not contain intron sequences.In order to amplify a transcript; one will have to start from total RNA instead of the genome.

The deduced GLYAT amino acid sequences of both the chacma baboon and the vervet monkey were aligned to the GLYATs from other primates. The objective was to establish how different or similar the chacma baboon and the vervet monkey's GLYAT are to other primate GLYATs. The alignment was done and shown in Figure 2.14 and the phylogenic tree is shown in figure Figure 2.15.

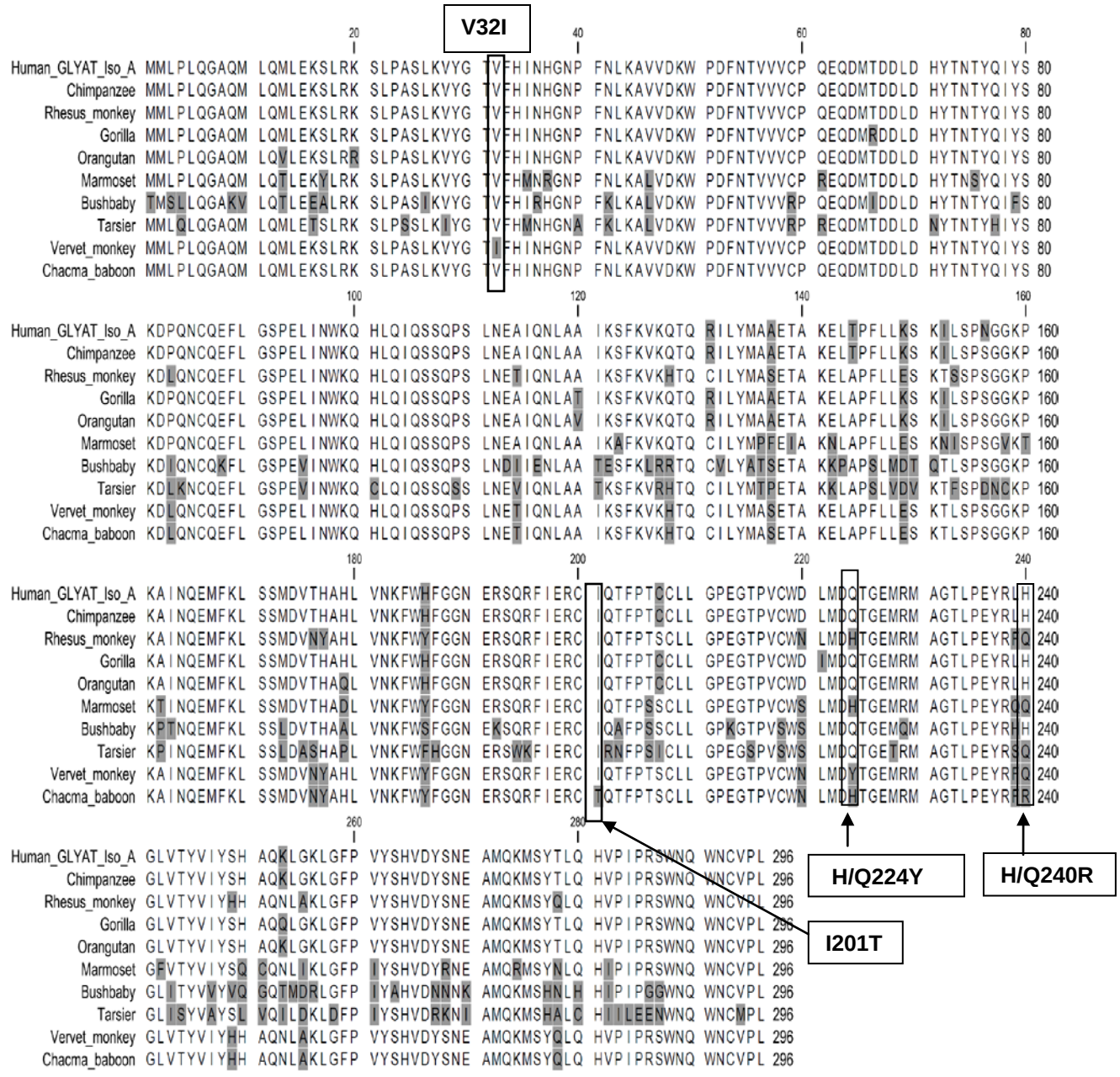


Figure 2.14: The chacma baboon and vervet monkey open reading frames aligned to those from other primates. Amino acids differing from the concensus are shown with a dark background. The following species are shown: Human (ENSP00000340200), gorilla (ENSGGOP00000004041), chimpanzee (ENSPTRP00000006356), rhesus monkey (ENSMMUP00000015380), tarsier (ENSTSYP00000008427), bushbaby (ENSOGAP00000006740), marmoset (ENSCJAP00000015482) and orangutan (ENSPPYP00000003731).

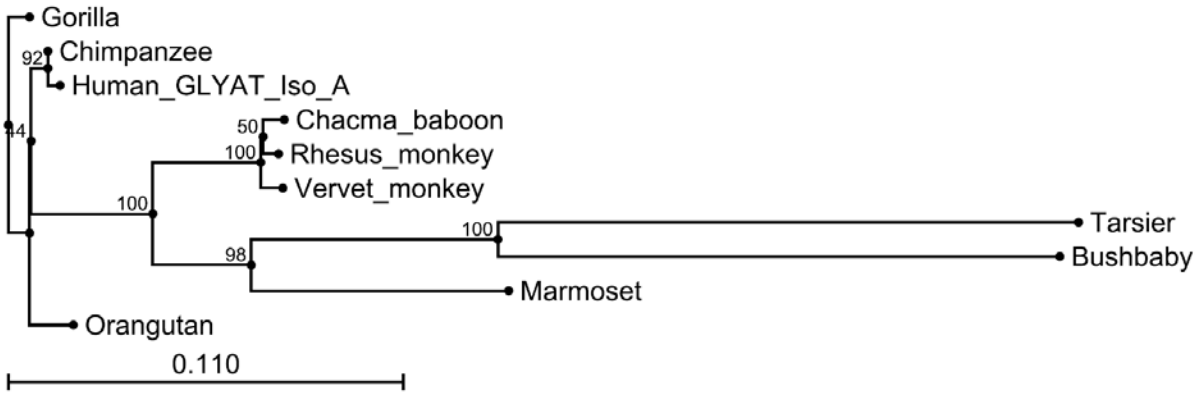


Figure 2.15: A phylogenic tree analysis of GLYAT from primates. The following species are shown: Human (ENSP00000340200), gorilla (ENSGGOP00000004041), chimpanzee (ENSPTRP00000006356), rhesus money (ENSMMUP00000015380), tartier (ENSTSYP00000008427), bushbaby (ENSOGAP00000006740), marmoset (ENSCJAP00000015482) and orangutan (ENSPPYP00000003731).

The GLYAT from the 10 primates aligned were different in 106 amino acid positions (Figure 2.14). From this 106 amino acid positions, There were 4 new variations introduced by GLYAT amino acid sequences from the chacma baboon and the vervet monkey. The vervet monkey introduced an isoleucine in place of a valine at position 32 (I32V) and an arginine in place of a histidine or glutamine at position 224 (H/Q224Y). The chacma baboon introduced a tyrosine in place of isoleucine at position 201 (I201T) and an arginine in place of histidine or glutamine at position 240 (H/Q240R).

The GLYAT proteins from rhesus monkey, chacma baboon and vervet monkey grouped distinctively from those corresponding to human and chimpanzee (Figure 2.15). The chacma baboon and vervet monkey sequences were 99 % identical to rhesus monkey sequence and 92.6 % identical to the human sequence. The human and chimpanzee sequences were 99.7 % identical to each other.

The chacma baboon and vervet monkey GLYAT amino acid sequences were aligned to all 4 human GLYAT family members (Figure 2.16). The objective was to establish how GLYATs from the vervet monkey and the chacma baboon relate to all 4 members of the GLYAT family.

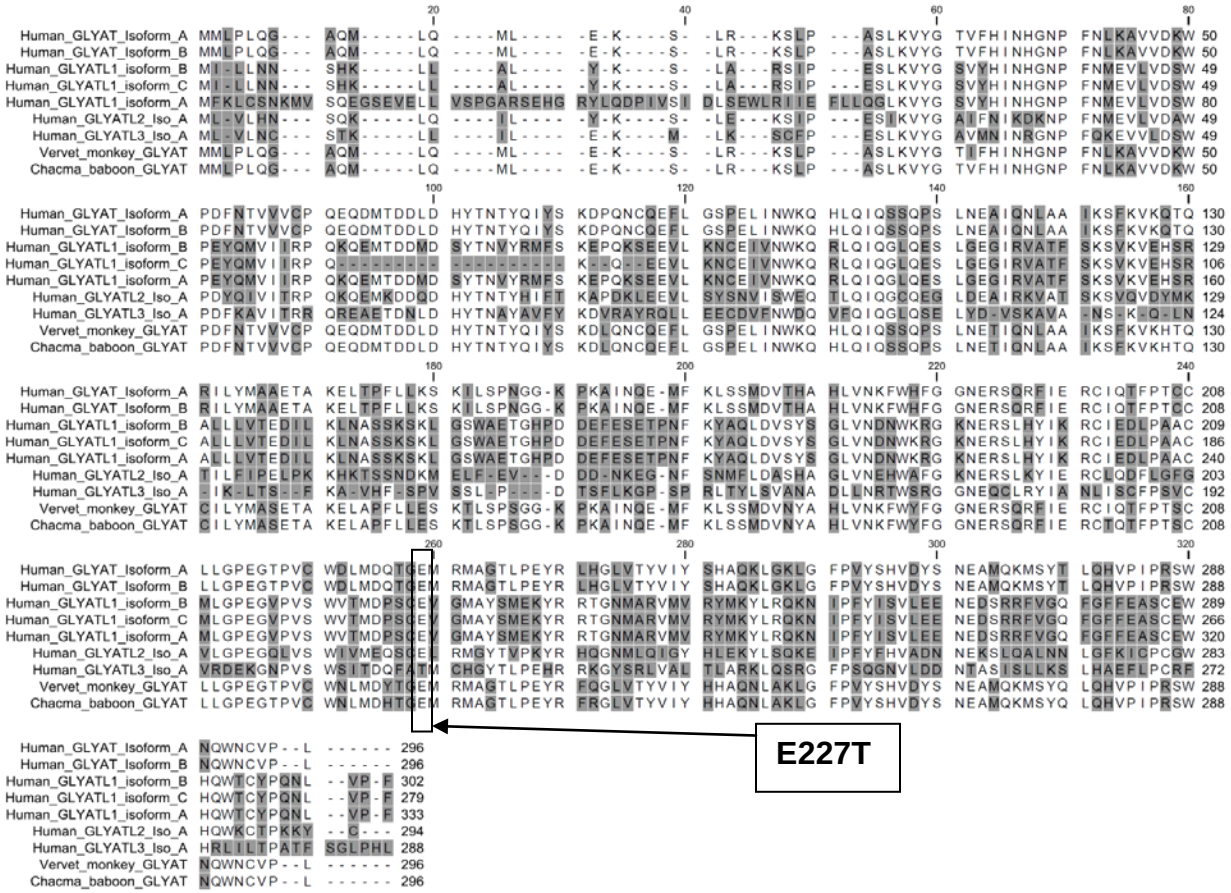


Figure 2.16: The chacma baboon and vervet monkey GLYAT proteins were aligned to all 4 human GLYAT family members (with their isoforms) known to date. Dots were automatically inserted by the computer in order to get a better fit of the alignment. Regions differing from the majority of the sequences are highlited with a dark background. The E227T position is shown.

In Figure 2.16, the GLYATL3 showed a different amino acid at position 227 of the protein (E227T). The possible significance of this position will be discussed later, in section 2.7. The human GLYATL1_isoform_A has the longest polypeptide of all GLYAT family members. It has 333 amino acids whereas GLYAT has 296 amino acids. The last amino acid at the C-terminal of all GLYAT family members is leucine except in the case of GLYATL2_isoform_A and GLYATL3_isoform_A. The human GLYATL2_isoform_A has a unique motif (KKYC) at the C-terminal.

2.7 Summary

Glycine-N-acyltransferase (EC 2.3.1.13, GLYAT) is one of the enzymes from the GCN5-related N-acetyltransferase (GNAT) superfamily responsible for the biotransformation of endogenous and xenobiotic toxic compounds. The Biochemistry Department at the North-West University (Potchefstroom campus) has an initiative to study detoxification profiling using substrate loading tests. The goal is to eventually be able to correlate patients' genotype to their biochemical detoxification profile. Hence, the genes encoding GLYAT enzymes need to be characterised. The aim of this project was to characterize the GLYAT enzyme in two additional non-human primates (the chacma baboon and the vervet monkey) so that we can start to understand its diversity and gain more insight into the mammalian detoxification system.

A cDNA library was prepared from total RNA of a chacma baboon. The cDNA encoding GLYAT was successfully amplified from this cDNA library, cloned and sequenced. The amplicon containing a transcript encoding the GLYAT of chacma baboon was sequenced up to 1256 bases long (Figure 2.12). The amplicon contained an open reading frame of 891 bases which translated into a polypeptide of 296 amino acids. There were 122 bases flanking the 5'-end of the chacma baboon GLYAT ORF. There were 243 bases flanking the 3'-end of the chacma baboon GLYAT ORF. The primers used to amplify the chacma baboon GLYAT transcript were designed from the human GLYAT transcript region spanning its ORF.

The vervet monkey GLYAT ORF DNA was sequenced by first generating PCR amplicons (from genomic DNA) containing DNA fragments that represent six exons encoding the vervet monkey GLYAT. To construct the vervet monkey GLYAT ORF, the separate exons were joined to form an open reading frame of 891 bases long which translated into 296 amino acids (Figure 2.13).

There were 4 new variations introduced by GLYAT amino acid sequences from the chacma baboon and the vervet monkey. The vervet monkey introduced an isoleucine in place of a valine at position 32 and an arginine in place of a histidine or glutamine at position 224. The chacma baboon introduced a tyrosine in place of isoleucine at position 201 and an arginine in place of histidine or glutamine at position 240.The SNPs

that have been proposed to have an effect on enzyme activity (Carnadas *et al*., 2010) are shown (Ser17Thr and Arg199Cys).

The phylogenic analysis after the alignment of GLYAT amino acids showed that the GLYAT proteins from rhesus monkey, chacma baboon and vervet monkey grouped distinctively from those corresponding to human and chimpanzee (Figure 2.15). The chacma baboon and vervet monkey showed 99% identity to rhesus monkey but only 92.6 % identity to the human sequence. All human GLYAT family members were aligned and they revealed that the position 227 on the protein was not conserved (Figure 2.16). The position 227 on the GLYAT protein was suggested to be important for GLYAT catalytic mechanism because it was the most conserved position in the GNAT superfamily members along the amino acids sequence making up the conserved GNAT fold (MSc of Mr CPS Badenhorst, 2009).

Chapter 3
Expression of a chacma baboon's recombinant glycine-N-acyltransferase in *Escherichia coli* and the preparation of a crude cytoplasmic extract containing native glycine-N-acyltransferase from chacma baboon liver

3.1 Introduction

In the previous chapter, the ORF encoding glycine-N-acyltransferase (GLYAT) of the chacma baboon was amplified from cDNA, cloned and sequenced. This chapter reports on the expression of a recombinant GLYAT from the cloned ORF in *Escherichia coli* (Origami™ strain) using the expression vector pColdIII. Mammalian recombinant proteins require an appropriate host to express and fold properly. The Origami™ cells and pColdIII system were selected because they were used successfully to express a soluble bovine recombinant GLYAT which had enzyme activity (MSc of Mr CPS Badenhorst, 2009). The Origami™ cells are easy to work with, can be grown to high density, their genetics is known and they lack the trx/gor gene which results in increased levels of disulphide bond formation in recombinant proteins synthesised in the cytoplasm (reviewed by Sorensen and Mortensen, 2005).

When proteins are expressed at lower temperatures their synthesis and folding is slowed down. This sometimes creates better conditions for the formation of a soluble and functional recombinant enzyme. Growing cultures slowly at lower temperatures has been shown to produce special proteins called the “cold shock” proteins in what has been termed a “cold shock” response (Jones *et al*., 1987). These “cold shock” proteins are expressed as a result of the induction of the cspA promoter in response to cold environmental temperatures to help the cell survive the cold (Goldberg *et al*., 1997). Chaperones (i.e. Dnak, DnaJ, GrpE, GroEL and GroES) are proteins that help with the folding of newly expressed proteins to avoid protein aggregation and facilitate formation of native proteins (Nishihara *et al*., 1998). Thus, combining the cspA promoter with chaperone expression should promote better expression and a correctly folded chacma baboon recombinant GLYAT.

The cold shock expression plasmid used in this experiment was a modified version of pColdIII of Takara™ (GENBANK accession number AB186390). The pColdIII was modified in such a way that there is a histidine tag (His-tag) at the C-terminal of the expressed polypeptide. A nucleotide sequence encoding a serine-glycine repeat (18 amino acids in total) was included as a spacer between the recombinant protein of interest and the His-tag [Figure 3.1 (A)]. The modified pColdIII plasmid was obtained from Dr AC Potgieter with an insert that encodes a protein of epizootic hemorrhage disease virus which will be referred to as EHDV for the purpose of this project. Since this EHDV protein is not important for this project it was replaced with chacma baboon’s GLYAT ORF.

The TAKARA™ chaperone plasmid used in this chapter was pG-TF2 (8.3 kbp) [Figure 3.1 (B)]. It expresses the chaperone proteins: groES (10 kDa), groEL (60 kDa) and Tf (trigger factor) (56 kDa) from the promoter Pzt-1 which is induced by tetracycline. The chaperone plasmid carries an origin of replication from pACYC and a resistance gene to chloramphenicol (Cm^r). Chaperones from pG-TF2 can only be used with *E. coli* systems that utilize other resistance than resistance to chloramphenicol (Cm^r) such as the pColdIII type of plasmids which express resistance to ampicillin.

In addition, a crude cytoplasmic extract containing native chacma baboon GLYAT protein was also prepared from chacma baboon liver and tested for enzymatic activity of the reaction for human GLYAT enzyme.

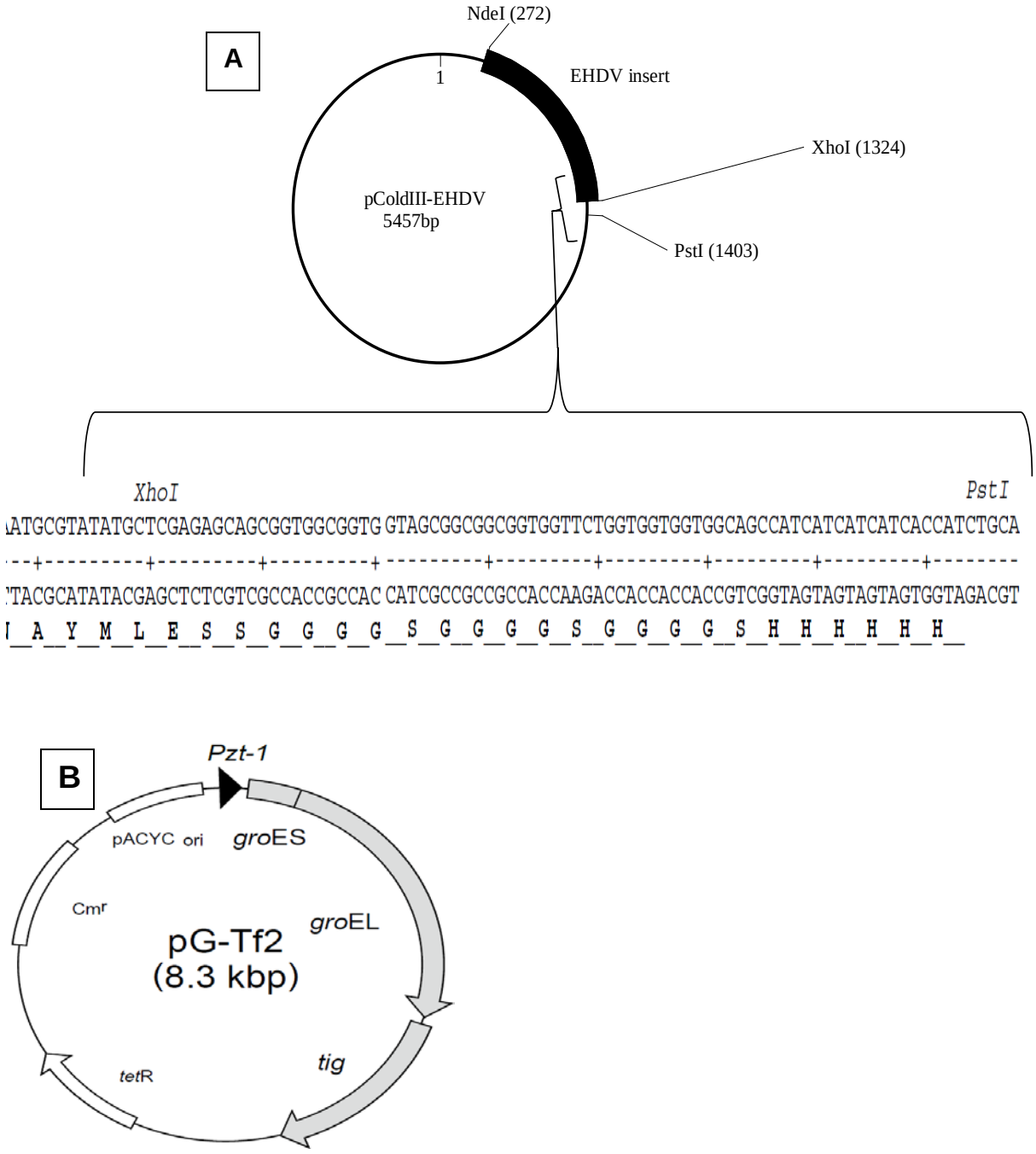


Figure 3.1: A schematic representation of the double stranded plasmid DNAs of: (A) the expression vector containing an insert (pColdIII-EHDV) (B) the chaperone plasmid (pG-Tf2 [8.3 kbp])

The aims of the research presented in this chapter were to:

1) Express a recombinant GLYAT of chacma baboon using pColdIII as an expression vector in Origami™ cells

2) Prepare crude cytoplasmic extract from liver of chacma baboon and test for the presence of GLYAT activity

3) Compare the enzyme activity of the native and recombinant GLYAT of the chacma baboon

3.2 Material and methods

3.2.1 Expression of baboon recombinant GLYAT protein in Origami™ cells

3.2.1.1 Co-expression of chaperones and a recombinant GLYAT of chacma baboon in Origami™ cells

The pColdIII plasmid has the cspA promoter, which is induced by cold temperatures. The cspA promoter is induced optimally at 15°C, where normal expression of cellular protein slows down. Thus, the cold favors protein expression from the cspA promoter. In this project, the PCR product containing a cloned copy of an open reading frame encoding GLYAT of chacma baboon was cloned into a modified pColdIII plasmid so as to exploit the advantage of the cpsA promoter. It is commonly known in literature that mammalian proteins expressed in *E. coli* may experience difficulty in folding because the *E. coli* does not have appropriate mammalian cell machinery. The chaperones

[groES (10 kDa), groEL (60 kDa) and Tf (56 kDa)] were used to help to correctly fold foreign proteins expressed in an expression system.

It is important that cells are transformed with the chaperone plasmid first then followed by any other plasmid of interest. The Origami™ cells were first made competent and then transformed (as described in section 2.2.1.8.3 and 2.2.1.8.4) with the chaperone plasmid. The successful transformants containing the chaperone plasmid were kept in media containing 25 µg/ml chloramphenicol. The chloramphenicol will ensure that only cells that were successfully transformed with the chaperone plasmid will remain viable in the 15 % glycerol stock.

The Origami™ cells containing the chaperone plasmid were made competent, for the second time, then transformed with the pColdIII plasmids (described in section 2.2.1.8.3 and 2.2.1.8.4). The transformed cells were inoculated in 15 ml Luria-Bertani (LB) broth containing 25 µg/ml chloramphenicol to select cells that contain the chaperone plasmid and 100 µg/ml ampicillin to select cells that contain one of the pColdIII plasmids. The cultures were incubated at 37 °C shaking at 200 rpm overnight to reach saturation. The next morning, each culture was centrifuged at 2 000 x g for 10 min and the supernatant was discarded. Each pellet was resuspended in 50 ml LB broth with 25 µg/ml chloramphenicol, 100 µg/ml ampicillin and 0.01 µg/ml tetracycline to induce the chaperone expression from pG-TF2 plasmid. The inoculated culture was incubated at 37 °C with shaking at 200 rpm for an hour. The 50 ml culture was divided into two 25 ml cultures and incubated at 15 °C for 30 min. One 25 ml culture was used as a negative

control by omitting IPTG. The other 25 ml culture had IPTG added to a final concentration of 0.5 mM to induce the *lac* operon which will result in the expression of chacma baboon recombinant GLYAT. After induction, all cultures were incubated at 15 °C for 24 hours.

3.2.1.2 Protein extraction after co-expression of chaperone plasmid and recombinant GLYAT in Origami™ cells

Four empty 50 ml tubes were weighed and the four 25 ml expression cultures were poured into them. They were all centrifuged at 3 000 x g for 30 min and the supernatant was discarded. Pellets were allowed to dry by placing the tubes upside down for five minutes on paper. Then, the following was added to each culture: 5 ml Bugbuster® (Novagen™) / gram wet paste to disrupt the cell wall of *E. coli*, 1 KU rLysozyme solution/ per ml Bugbuster®. The rLysozyme disrupts the cell wall by hydrolysing the N-acetylmuramide linkages in the cell wall and by so doing enhances extraction efficiency. Then 25 U of benzonase® nuclease is added per ml Bugbuster® to reduce the viscosity of the extract due to liberation of chromosomal DNA. The suspension was incubated at room temperature shaking at 40 rpm for 20 min. Then, 100 µl of total fraction was removed for both induced and uninduced cultures and stored on ice. The remaining cultures were centrifuged at 3 000 x g for 40 min. The resulting supernatant was taken to be the soluble fraction and the pellets as insoluble fraction.

3.2.1.3 10 % Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

The objective of the method is to separate proteins according to their sizes. However, proteins expressed in a cell have different charges which could affect their migration speed when an electric current is applied. In order to make all proteins carry the same charge a sample is denatured by boiling in the presence of a reducing agent which disrupts the disulphate bridges formed by the cysteine residues. SDS is an ionic detergent which binds to proteins and causes them to acquire a negative charge that prevents them from refolding. Then, the samples are loaded on the polyacrylamide gel. An electric field is applied to separate denatured proteins according to size (Laemli *et al*., 1970).

10 % SDS-PAGE was used for a variety of samples as indicated in text. Both the stacking and separation gels were prepared following the protocol described in Russel and Sambrook (2001). The separating gel was composed of: 0.1 % SDS, 10 % acrylamide, 0.27 % bisacrylamide, 375 mM Tris-HCl pH 8.8, 0.008 % TEMED and 0.08 % ammonium persulfate. The solution was poured between the two glass plates and placed in a BIO-RAD mini protean gel casting tray (70 mm x 76 mm x 1mm). Saturated isobutanol was poured on top and left to stand for an hour. The isobutanol prevents the oxygen in air from inhibiting the gel polymerisation reaction. The isobutanol was replaced with the stacking gel and Teflon combs were inserted. The stacking gel was prepared with a solution containing: 0.78 % acrylamide, 0.0208 % bisacrylamide

solution, 0.2 X Tris.Cl/SDS pH 6.8, 0.01 % (v/v) ammonium persulfate and the solution made to a final volume of 25 ml with 18 Ω water. Then, 0.0002 % (v/v) TEMED was added and mixed for 5 sec. The addition of TEMED starts the polymerization instantly therefore TEMED was added last. The solution was poured between the two vertical gel casting plates and allowed to polymerise at room temperature for an hour.

Each protein sample was prepared before the loading on gel as follows: 5 µl of protein sample, 2 M DTT (reducing agent), Novagen™ 1 x dual color® protein loading buffer (0.0625 M Tris-Cl, 0.4 mM EDTA (pH 8.5), 2 % (w/v) SDS, 10 % (v/v) glycerol), 0.01 % (w/v) bromophenol blue and 0.006 % (w/v) pyronin Y) and water (autoclaved 18 Ω) to make up the final volume of 20 ul. Proteins were denatured by heating samples at 95 °C for 5 min. The solution was centrifuged briefly (~2 sec). 5 µl of each samples were loaded on the gel. A constant current of 25 mA was applied at 60 Volts for 30 minutes. The SDS polyacrylamide gel was immediately immersed in a staining solution containing Coomasie dye for 30 min. Destaining was conducted by immersing the gel in destain solution overnight. The destain solution was composed of 5 parts ddH_20, 4 parts methanol and 1 part acetic acid (Russel and Sambrook, 2001). The next day, the destain solution was replaced with distilled water.

3.2.2 Preparation of a crude cytoplasmic extract containing native glycine-N acyltransferase from the chacma baboon's liver

3.2.2.1 Preparation of a mitochondrial supernatant fraction

A 25 g piece of fresh liver of chacma baboon was placed in a 50 ml Falcon tube together with 4 volumes of 0.13 M KCl at pH 8 at 4°C. The volume of the KCl was calculated using the volume which the liver tissue reached in the Falcon tube. The tissue was homogenized manually on the bench using a glass homogenizer. The homogenate was centrifuged at 6 000 x g for 10 min and the pellets were discarded. A second centrifugation at 9 000 x g was done for 10 min and the supernatant was discarded. The pellet was resuspended in 10 % of the original volume of 0.13 M KCl pH 8 and stored at -80 °C. The next day, the suspension was thawed and then frozen repeatedly 3 times at -80 °C. After the final thawing, the suspension was centrifuged at 35 000 x g for 2 hours at 4 °C. The pellet was discarded and the supernatant was used for protein determination.

3.2.2.2 The bicinchoninic acid assay for determining protein concentration

Proteins have a property to interact with cupric ion (Cu^{2+}) to release a Cu^{1+} ion. The coprous ion (Cu^{1+}) binds to bicinchoninic acid (BCA) in an alkaline environment to form a complex that displays a deep blue color and has a maximum absorbance at 562 nm. This absorbance is directly proportional to the protein concentration (Smith *et al*., 1985).

A sample from the final supernatant (3.2.2.1) was diluted (1/10). Triplicate dilutions were prepared. The diluted sample was then mixed with 20 X the volume of BCA:

$CuSO_4.5H_2O$ (50:1) in a microlitre plate, in triplicate. A series of samples with increasing BSA concentrations to form a standard curve were prepared where the concentration of bovine serum albumin (BSA) was varied in a gradient of 0 µg/µl, 20 µg/µl, 40 µg/µl, 60 µg/µl, 80 µg/µl, 100 µg/µl. Both the diluted crude cytoplasmic extracts and standard series were incubated at 37 °C for 20 min and absorbance was read at 562 nm in a BioTek™ microplate fluorescence (FL 600) plate reader. The amount of protein in the diluted crude cytoplasmic extract was quantified by using the standard curve obtained after plotting values obtained for the absorbance of BSA standard series.

3.2.2.3 Protein precipitation with ammonium sulfate

The protein present in the crude cytoplasmic extract from chacma baboon's liver was determined as described above (section 3.2.2.3) and then adjusted to 20 mg/ml. Then, ammonium sulfate salt was added to a final concentration of 288 mg/ml in the supernatant and mixed until all the ammonium sulfate was dissolved. The suspension was centrifuged at 10 000 x g for 15 min and the pellet was discarded. More ammonium sulfate was added to a final concentration of 144 mg/ml in the supernatant and dissolved. The suspension was centrifuged at 10 000 x g for 10 min and the supernatant was discarded. The pellet was resuspended in 4 ml of 0.1 M KCl; 20 mM Tris-HCl, pH 8. The suspension was aliquoted into 300 µl volumes and frozen at -80 °C. Protein determination was done as described in 3.2.2.2.

3.2.2.4 Enzyme activity test for GLYAT using the DTNB-based spectrophotometric assay

The DTNB-based spectrophotometric assay for measuring GLYAT activity has been used for a while now (Webster *et al*, 1976). This enzyme activity test was done on the precipitated protein of the crude cytoplasmic extracts containing the native chacma baboon's GLYAT using the substrates glycine and benzoyl Coenzyme A (benzoyl-CoA). The principle here is that the native GLYAT of chacma baboon will transfer glycine to benzoyl and thus free the Coenzyme A. The Coenzyme A contains the thiol group (-SH) which interacts with the disulfide bond on the 5, 5 dithiobis (2-nitrobenzoic acid) molecule (DTNB). The thiol group attacks the disulfide bond causing a cleavage to form 2-nitro-5-thiobenzoate (NTB^-) which ionizes to give a yellow NTB^{2-}. The NTB^{2-} is quantified by measuring the absorbance at 412 nm. Thus, the rate of appearance of NTB^{2-} can be used to estimate the rate of release of Coenzyme A from benzoyl Coenzyme A (Ellman, 1959; Kolvraa and Gregersen, 1986).

The chacma baboon GLYAT reaction was set up as follows: 0.025 g/l DTNB, 1 % (w/v) benzoyl Coenzyme A, 20 % (w/v) glycine, 1 % (w/v) chacma baboon's liver crude cytoplasmic extract sample and water to final volume of 500 µl. The bovine crude cytoplasmic extract was used as a positive control since it had shown enzymatic activity. The crude cytoplasmic extract from bovine liver containing GLYAT was obtained from a separate project (MSc of Ms M. Snyders). Two sets of negative controls were set up. One had all the components of the experiment except for glycine and the

other had all components of the experiment except for the benzoyl Coenzyme A. The omitted components were compensated for with water.

3.3 Results

3.3.1 Cloning of baboon GLYAT ORF into an expression vector

The objective of the experiment was to clone the PCR product containing a cloned copy of an open reading frame encoding GLYAT of chacma baboon into pColdIII and pColdIII-His. The plan was to express a recombinant GLYAT fused to a C-terminal histidine tag from pColdIII-His so that it can be purified using a nickel affinity column. The unmodified commercial pColdIII was also expressed to determine whether there could be a difference in the expression of the chacma baboon's recombinant GLYAT. The pColdIII, pColdIII-EHDV and the PCR product containing a cloned copy of an open reading frame encoding GLYAT of chacma baboon were digested with restriction enzymes NdeI and XhoI (described in section 2.2.1.8.7) to give products: pColdIII (NdeI/XhoI), pColdIII-His (NdeI/XhoI), and a digested amplicon containing chacma baboon's GLYAT ORF respectively. The products shown in Figure 2.9 (A) were gel extracted. These gel extracted products were: a PCR product containing a cloned copy of an open reading frame encoding GLYAT of chacma baboon (lane 2), pColdIII (lane 3) and pColdIII-His (lane 4). These gel extraction products were analysed on a 1 % agarose gel shown in Figure 3.1.

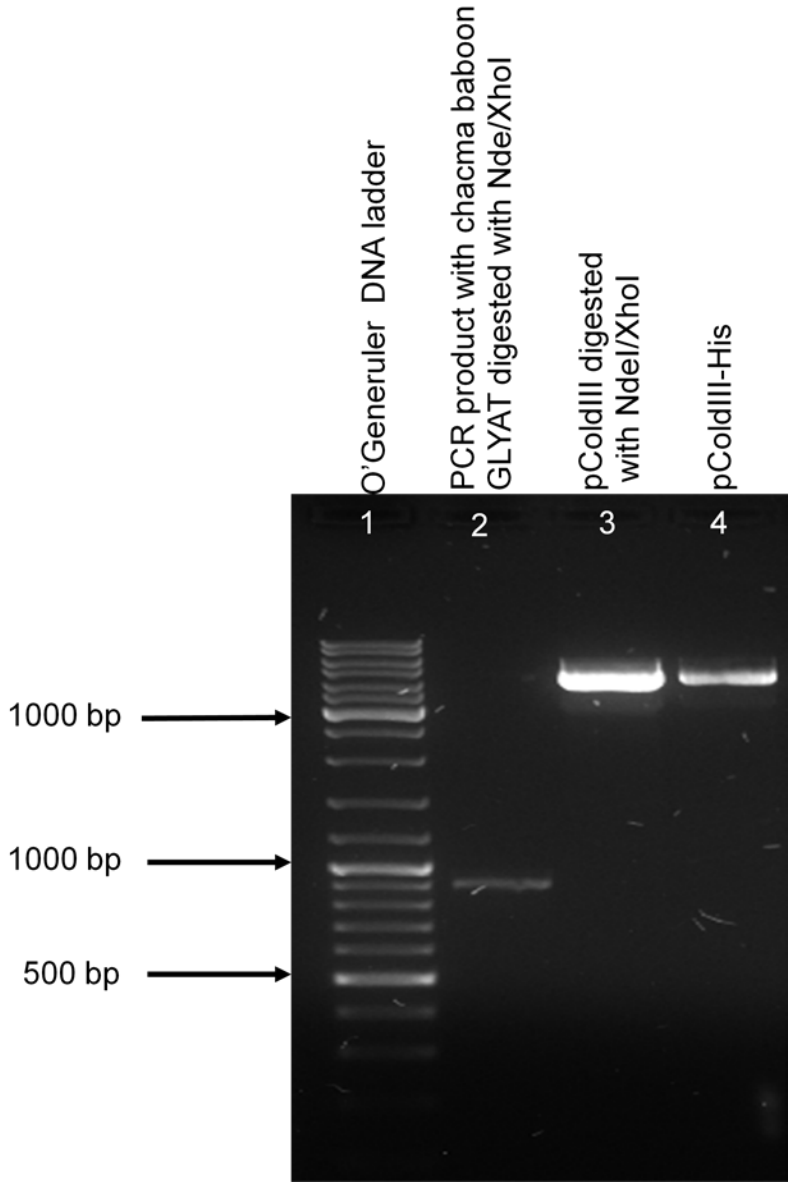


Figure 3.2: A 1 % agarose gel analysis of the gel extraction of the two pColdIII vectors and GLYAT insert after restriction enzyme digestion. Lanes: (1) O'Generuler DNA ladder from Fermentas cat #SM173. (2) PCR product containing an ORF encoding GLYAT of chacma baboon digested with NdeI and XhoI then gel extracted. (3) pColdIII digested with NdeI and XhoI then gel extracted. (4) pColdIII-His derived from the digestion of pColdIII-EHDV with NdeI and XhoI followed by gel extraction.

The DNA concentration and purity of gel extraction products show in Figure 3.2 were determined using the NanoDrop spectrophotometre. The DNA concentration of the gel extracted of PCR product containing a cloned copy of an open reading frame encoding for chacma baboon's GLYAT (Figure 3.2, lane 2) was 90 ng/µl; the DNA yield was 4.5 µg and the $A_{260/280}$ ratio was 1.8. The DNA concentration of the gel extract containing the pColdIII (Figure 3.2, lane 3) was determined to be 161 ng/µl which represents a DNA yield of 8.05 µg and the $A_{260/280}$ ratio was 1.8. The gel extract containing the pColdIII-His-tag (Figure 3.2, lane 4) had a DNA concentration of 120 ng/µl and the DNA yield was 6 µg and the $A_{260/280}$ ratio was 1.8. The Macherey-Nagel gel extraction kit (described in section 2.2.1.7) is capable of DNA recoveries of 15 µg when eluting with TE buffer but expected to drop to only about 1 to 5 µg when final DNA elution is done with water. In this experiment, the DNA was eluted with water because the TE buffer may cause problems in the subsequent cloning steps. Thus, the DNA concentration recovered was good given the fact that the DNA was eluted with water. The next step was to clone the baboon GLYAT into both the pColdIII and pColdIII-His-tag vectors.

The PCR product containing a cloned copy of an open reading frame encoding GLYAT of chacma baboon was ligated into both the pColdIII to form pBabpColdIII DNA and pColdIII-His to form pBabGLYpColdIII-His-tag (described in section 2.2.1.8.8). The pBabpColdIII and pBabGLYpColdIII-His-tag were transformed into the JM109 cells (described in section 2.2.1.8.4). There were no colonies visible on the plates with JM109 cells transformed with pBabpColdIII DNA which indicated that the PCR product containing a cloned copy of an open reading frame encoding GLYAT of chacma baboon

failed to clone into the pColdIII plasmid. Thus, the next step of midi preparation of plasmid DNA was done using the JM109 cells transformed with pBabpColdIII-His-tag DNA since they had shown successful transformation.

A single colony containing JM109 cells transformed with pBabGLYpColdIII-His-tag was picked and inoculated in a 100 ml LB broth that contained 25 µg/ml chloramphenicol (the chaperone plasmid expresses resistance to chloramphenicol) and 100 µg/ml ampicillin. The culture was grown at 37 °C overnight. The pBabGLYpColdIII-His-tag was extracted from the JM109 cells using the PureYield^{TM} plasmid midi preparation system extraction kit (described in section 2.2.1.8.6).

Plasmid DNA concentration for pBabGLYpColdIII-His-tag was determined on the NanoDrop spectrophotometre to be 407 ng/µl which indicated a DNA yield of 40.7 µg with an $A_{260/280}$ ratio of 1.7. This was a 40 % DNA yield when compared to the ideal yield stated by the manufacturer.

Restriction enzyme digestion was done using pBabGLYpColdIII-His-tag DNA amounting to 1915 ng with restriction enzymes NdeI, XhoI, PstI and EcoRI. The purpose was to confirm whether the DNA encoding GLYAT was correctly inserted into pColdIII-His-tag. The expected fragment patterns were predicted using the DNAMAN™ computer software and they are shown in Figure 3.3. The products of the restriction enzyme digestion were separated on a 1 % agarose gel and the gel picture is shown in Figure 3.4.

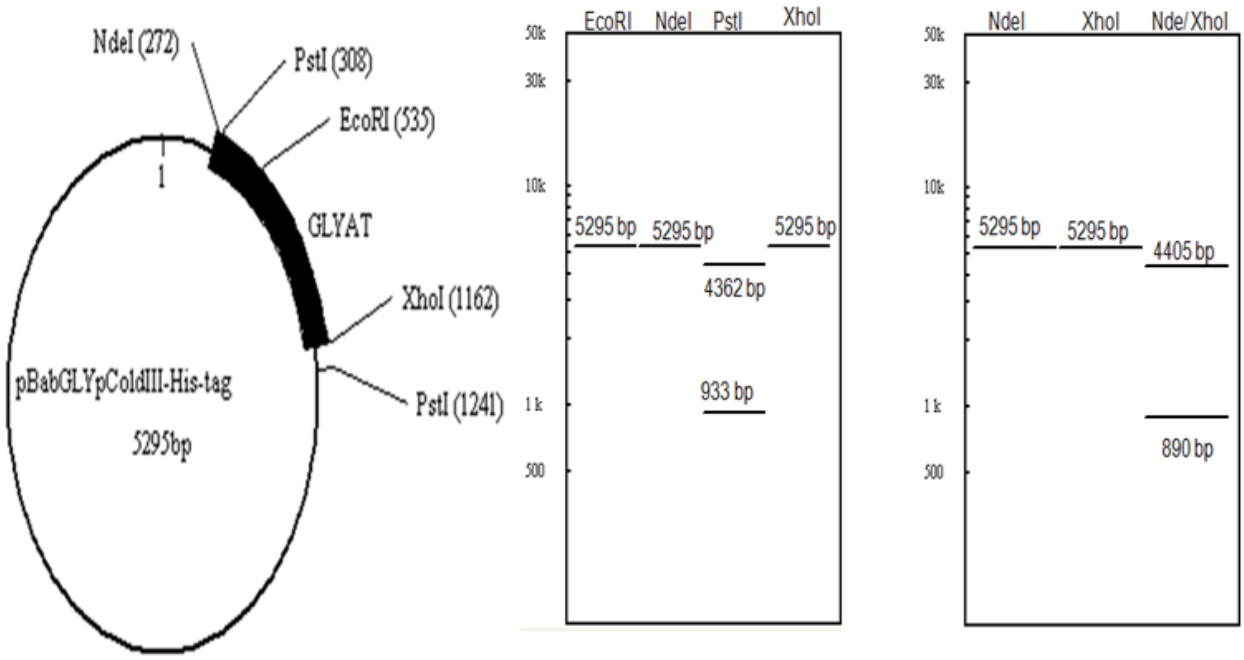


Figure 3.3: A diagrammatic representation of the expected gel patterns of the pBabGLYpColdIII-His-tag digested with the restriction enzymes: EcoRI, NdeI, XhoI and PstI. These gel patterns were predicted using the DNAMAN computer software.

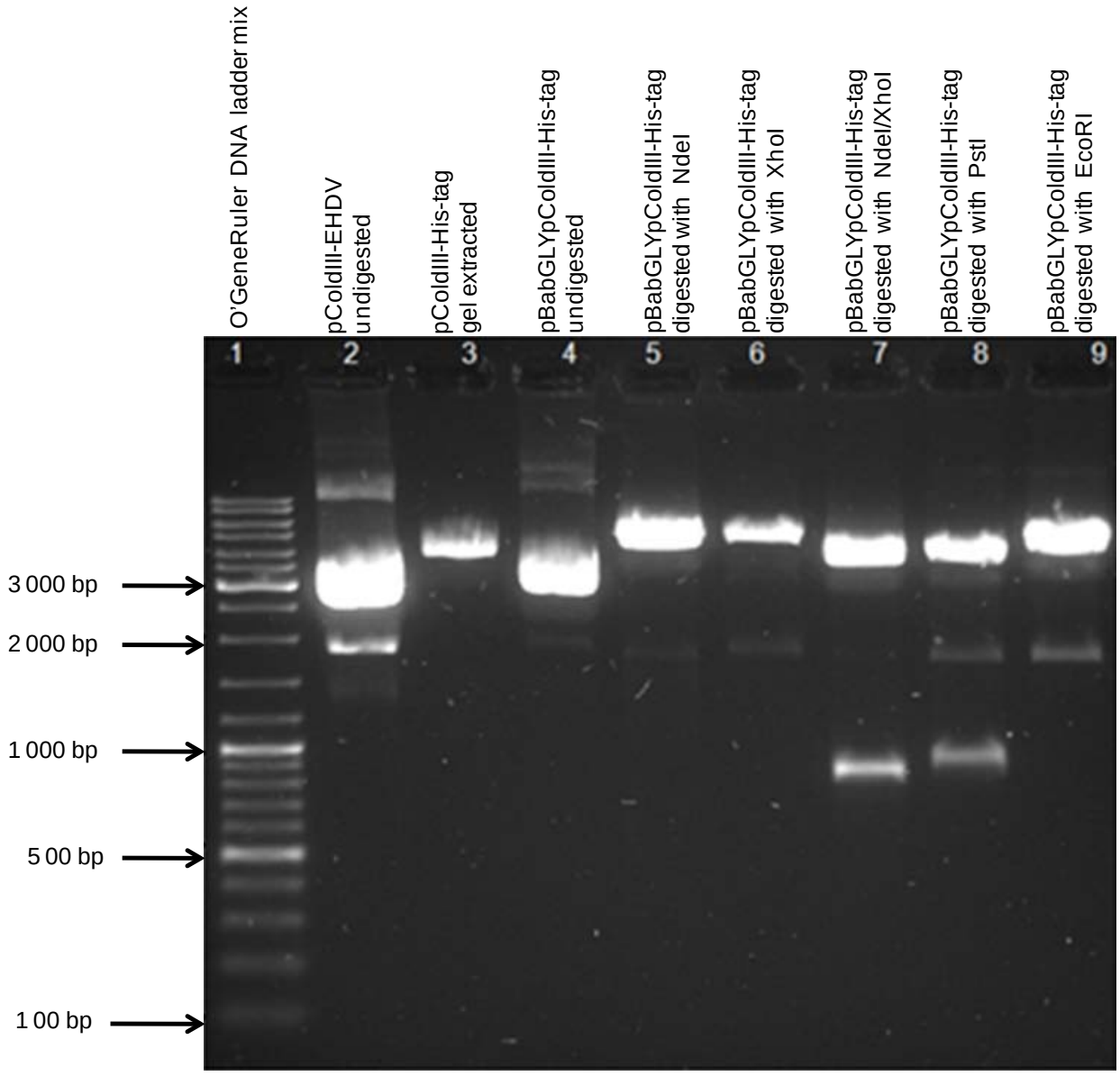


Figure 3.4: Analyses of the cloning of the open reading frame encoding GLYAT of the chacma baboon into pColdIII-His-tag by restriction enzyme digestion. Lanes: 1) O'GeneRuler DNA ladder from Fermentas cat #SM173. 2) Undigested pColdIII-EHDV. 3) Gel extracted pColdIII-His-tag. 4) Undigested pBabGLYpColdIII-His-tag. 5) pBabGLYpColdIII-His-tag digested with NdeI. 6) pBabGLYpColdIII-His-tag digested with XhoI. 7) pBabGLYpColdIII-His-tag double digested with NdeI and XhoI. 8) pBabGLYpColdIII-His-tag digested with PstI. 9) pBabGLYpColdIII-His-tag digested with EcoRI.

Figure 3.4 shows undigested pColdIII-EHDV with the characteristic three bands representing supercoiled and various relaxed circular forms in lane 2. Restriction enzyme digestion of pColdIII-EHDV with NdeI and XhoI to give pColdIII-His-tag (4405 bp) followed by a gel extraction gave a band of 4405 bp (lane 3). This 4405 bp fragment is the linearised vector pColdIII-His-tag into which the DNA fragments encoding GLYAT of the chacma baboon was cloned. pBabGLYpColdIII-His-tag DNA was linearised by separate digestions with NdeI and XhoI in lanes 5 and 6 respectively to give a linearised vector of size 5295 bp. Both NdeI and XhoI restriction enzymes were used to digest pBabGLYpColdIII-His-tag (5295 bp) to give pColdIII-HisTag (4405 bp) and a DNA fragment encoding GLYAT of chacma baboon (890 bp). The restriction enzyme PstI was used to cleave a 933 bp fragment from pBabGLYpColdIII-His-tag (5295 bp) shown in lane 8. However, there appears to be a band at around 2 000 bp in all the lanes except in lane 3 which had the gel extracted linear pColdIII-HisTag. This 2 000 bp band could be due to the undigested pBabGLYpColdIII-His-tag (5295 bp); since it is present as the bottom band on the same position on the gel in the undigested plasmid (lane 2) plasmid.

Therefore, the actual gel pattern of the restriction enzyme digestion was similar to the pattern generated by the DNAMAN computer software. The pBabGLYpColdIII-His-tag was DNA sequenced to confirm the restriction enzyme digestion results. The cloning of the DNA fragment encoding GLYAT of chacma baboon into pColdIII-His-tag was successfully confirmed (Figure 3.4).

Next, the pBabGLYpColdIII-His-tag and empty pColdIII were co-expressed with pG-TF2 in Origami™ cells.

3.3.2 Bacterial co-expression of a recombinant GLYAT with chaperones (groES, groEL and Tf)

The objective was to co-express a recombinant baboon GLYAT protein with chaperones groES, groEL and Tf in Origami™ cells. To achieve this, an empty pColdIII was also expressed so that the total and soluble fractions could be compared to that obtained from the expression of the pBabGLYpColdIII-His-tag. The recombinant chacma baboon GLYAT was expressed fused to a histidine tag so that it could be purified using a nickel affinity column. pBabGLYpColdIII-His-tag and pG-TF2 were transformed into Origami cells (described in section 2.2.1.8.4). The bacterial co-expression of GLYAT of chacma baboon and the chaperones was done (described in section 3.2.1.1). Proteins were extracted (described in section 3.2.1.2). Total and soluble fractions were separated by SDS-PAGE. The expected size of the recombinant GLYAT was predicted by translating the open reading frame encoding GLYAT of chacma baboon together with the His-tag sequence (inserted in the pBabGLYpColdIII-His-tag using) the DNAMAN™ computer software.The gel picture is shown in Figure 3.5.

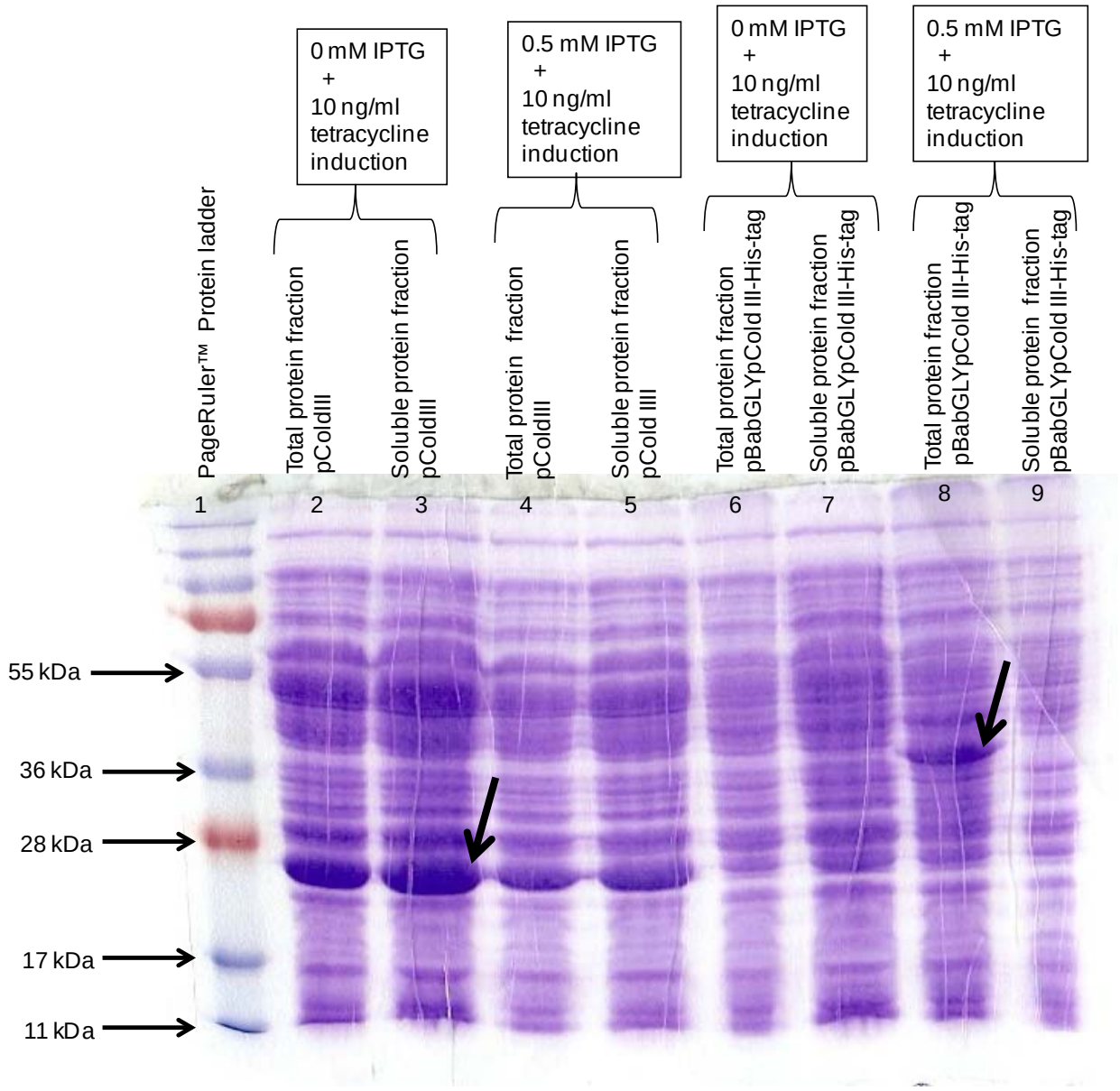


Figure 3.5: SDS-PAGE analysis of total and soluble protein fractions expressed in Origami™ cells expressing chaperones with and without the presence of baboon GLYAT. IPTG was used to induce the lac operator on the pColdIII and pColdIII-His-tag. Tetracycline was used to induce the Pzt-1 promoter on the pG-TF2 plasmid to express the chaperones. Lanes: 1) PageRuler™ prestained protein ladder plus (Fermentas cat #SM1811). Lanes 2 to 5 represent expression from pColdIII. Lanes 6 to 9 represent the expression of a recombinant chacma baboon GLYAT.

In Figure 3.5, the expressed recombinant chacma baboon GLYAT was expected to be visible just above 36 kDa (as shown by the bold arrow in lane 8). Therefore, the expression of a recombinant chacma baboon GLYAT in Origami™ cells was successful. Lane 8 (bold arrow) shows high expression of the recombinant chacma baboon GLYAT in total protein fraction. Lane 9 had soluble protein fraction and does not show the recombinant GLYAT. Lack of a recombinant GLYAT in the soluble fraction could mean that the protein was insoluble. However, the fact that no band seemed to be present in lane 9 suggested failure to express soluble recombinant protein.

Lanes 2, 3, 4 and 5 have a thick band just below the 28 kDa band (as shown by the bold arrow in lane 3) which could be the protein encoded by the multiple cloning site of the pColdIII. The multiple cloning site of the pColdIII encodes for 25 amino acids (as shown in Figure 2.3). In contrast, lanes 6, 7, 8 and 9 do not have such a band which could be due to cloning the chacma baboon's GLYAT ORF into the multiple cloning site.

Chaperones groEL, tig and groES were expected to be visible at 60 kDa and 56 kDa respectively. Somehow, chaperone bands are not visible at their expected positions on the gel. Thus, there is no evidence that the chaperones were expressed. It is possible that the chaperones either expressed very poorly or did not express at all.

The next step was to attempt to express the recombinant GLYAT at the same concentration of IPTG (0.5 mM) but increasing the concentration of tetracycline, which

induces the chaperones, from a 10 ng/ml to 100 ng/ml which amounts to a ten fold increase.

3.3.3 Co-expression of chacma baboon recombinant GLYAT with chaperones (groES, groEL and Tf) using a higher concentration of tetracycline for induction of chaperone expression

Chaperones were not expressed in the previous expression experiment (3.3.2). Successful chaperone expression could help prevent the recombinant protein from forming insoluble bodies. The exact expression conditions were followed as in 3.2.1 except that the concentration of the chaperone inducer (tetracycline) was increased ten fold to a final concentration of 100 ng/ml. Total and soluble fractions were separated by SDS-PAGE. The gel picture is shown in Figure 3.6.

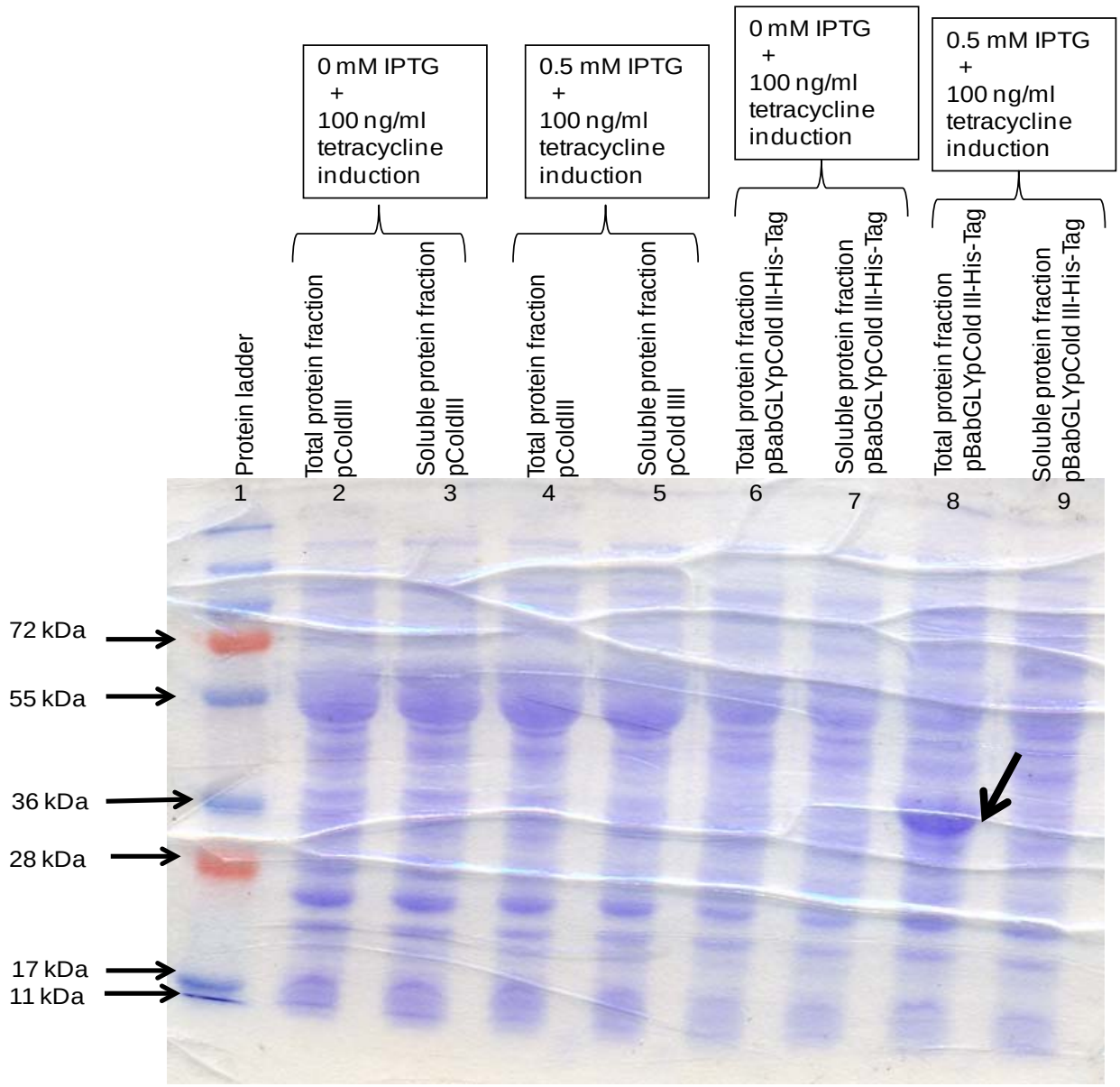


Figure 3.6: SDS-PAGE analysis of protein crude after expression of the chacma baboon recombinant GLYAT in Origami™ cells with the chaperones induced with 0.1 µg/L tetracycline. IPTG was used to induce the lac operator on the pColdIII and pColdIII-His-tag. Tetracycline was used to induce the Pzt-1 promoter on the pG-TF2 plasmid to express the chaperones. Lanes: 1) PageRuler™ prestained protein ladder plus (Fermentas cat #SM1811). Lanes 2 to 5 represent expression of pColdIII. Lanes 6 to 9 represent the expression of a chacma baboon recombinant GLYAT.

High expression of the recombinant chacma baboon GLYAT was visible (lane 8: bold arrow). This can be deduced from the presence of the ~37 kDa band in the total fraction in lane 8 but absent on the soluble fraction in lane 9. The recombinant chacma baboon GLYAT protein (~37 kDa) was again expressed in an insoluble form regardless of the 10 fold increased of the chaperone inducer (tetracycline) concentration.

There was no sign of chaperone expression at 0.1 µg/L tetracycline. Thus, even expression of chaperone with a ten fold increased inducer concentration did not help produce soluble recombinant GLYAT. The expected chaperones were GroEL (60 kDa), Tf (56 kDa) and GroES (10 kDa). These three chaperone protein bands are not visible in any of the lanes on the gel.

Lanes 2, 3, 4 and 5 contain total and soluble protein fractions from cultures that contained both pG-TF2 and pColdIII. These lanes served as negative control to make sure that the 37 kDa band (shown by bold arrow in lane 8) is not present even in the absence of the pBabGLYpColdIII-His-tag DNA in cells. In short, only Origami cells that contained pBabGLYpColdIII-His-tag and induced with IPTG are expected to express GLYAT.

Lanes 6 and 7 had total and soluble fractions of cultures containing pG-TF2 and pBabGLYpColdIII-His-tag plasmid induced with tetracyline but uninduced with IPTG.

This was done to make sure that overexpression of baboon GLYAT was, indeed, under the control of *lac operon*.

Next, a cytoplasmic extract containing GLYAT from liver of chacma baboon was prepared.

3.3.4 Testing for GLYAT enzyme activity in a cytoplasmic extract from liver of chacma baboon

In an event that it would be possible to generate a soluble chacma baboon recombinant GLYAT, having a native GLYAT extracted from baboon liver would be an ideal positive control. Therefore, a crude cytoplasmic extract was prepared from a liver of a chacma baboon as described in 3.2.2. The GLYAT enzymatic activity in this preparation was assayed (described in section 3.2.2.4) and compared with that of a similarly prepared cytoplasmic extract from bovine liver (Provided by Ms. M. Snyders). Both the chacma baboon and bovine liver cytoplasmic extract crudes were diluted to a final protein concentration of 29 µg/µl. The general reaction reaction contained 1 % (w/v) benzoyl-CoA and saturated amounts of glycine (100 mM). The reaction was conducted at 30° C for 12 minutes. In this experiment the 1% benzoyl-CoA was the limiting reagent while the glycine was used in excess amounts (100 mM). The results are shown in Figure 3.7

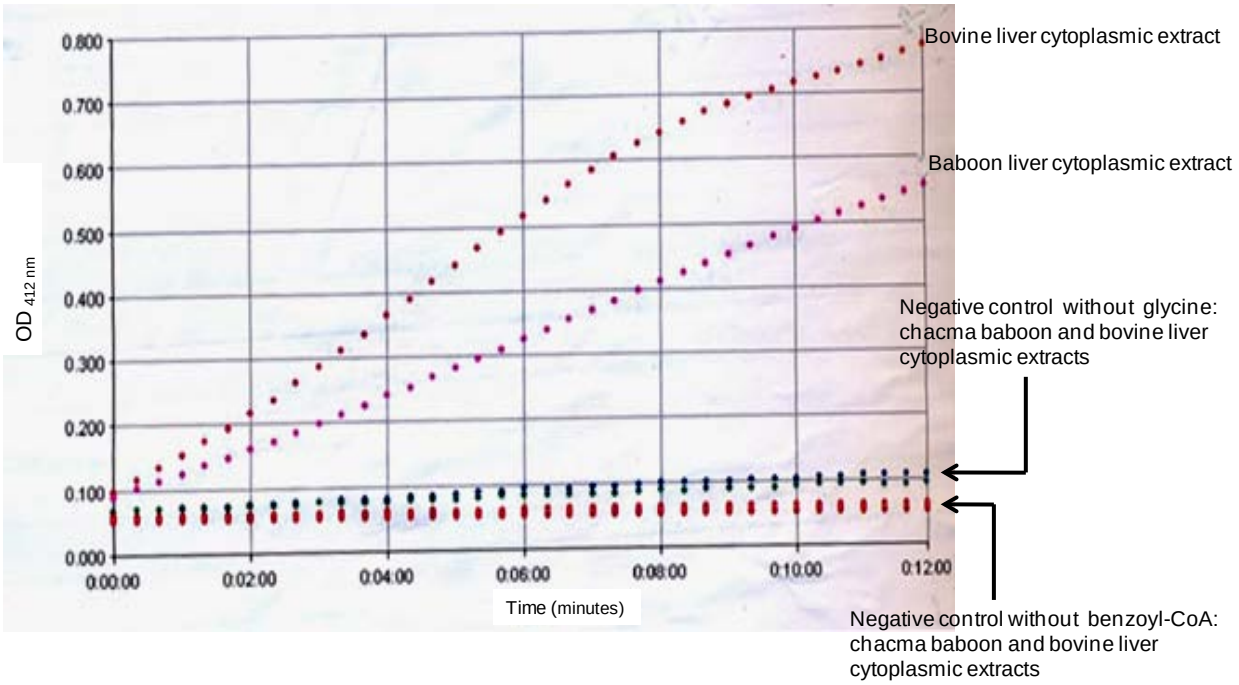


Figure 3.7: Comparison of GLYAT enzymatic activity in cytoplasmic extracts from livers of a chacma baboon and a bovine.

The cytoplasmic extract from bovine liver showed significantly more enzyme activity than that from chacma baboon liver (Figure 3.7). The negative controls (reaction containing no glycine or benzoyl-CoA) did not show any enzymatic activity. The negative controls are not supposed to show any enzymatic activity. It can not be concluded whether the bovine GLYAT had more activity than the chacma baboon GLYAT because it is not known how much GLYAT was in each extract.

3.4 Summary

The original expression vector pColdIII does not normaly contain a base sequence encoding for a His-tag which can be used to selectively purify the recombinant proteins. The base sequence encoding for His-tag was incorporated into pColdIII, by Dr AC Potgieter, in such a way that the His-tag will be linked onto the C-terminal of the recombinant protein. A serine-glycine repeat of 18 bases was also inserted to serve as a linker between the His-tag and the recombinant protein. The open reading frame encoding GLYAT of chacma baboon was successfully cloned into this modified pColdIII expression vector (pColdIII-His-tag). The success in cloning was confirmed by restriction enzyme digestion (Figure 3.4) and DNA sequencing (results not shown). A recombinant GLYAT from the chacma baboon was expressed in an insoluble form (Figure 3.5). The chaperones did not appear to be expressed even at the highest concentration of their inducer (100 ng/ml tetracycline) before tetracycline becomes toxic to all the cells including those containing the chaperone plasmid (Figure 3.6).

In addition, a cytoplasmic crude extract from liver of the chacma baboon was prepared and compared with a separately prepared cytoplasmic crude extract from bovine (provided by Ms M Snyders) (Figure 3.7). The objective was to compare the native protein purified from the chacma baboon liver with the recombinant protein. Since it was not possible to express a soluble recombinant GLYAT it did not seem useful to purify GLYAT from the crude cytoplasmic extract.

Chapter 4
Concluding discussion

4.1 Introduction

In this chapter the knowledge gathered from all the experiments in this project is summerised and discussed.

The Biochemistry Department of the North-West University is investigating the mammalian detoxification system. One of the goals is to understand how the mammalian body detoxifies the biochemical metabolites that accumulate due to inborn errors of metabolism. To achieve our goal, we have chosen to focus on phase II of detoxification with the main focus on the GLYAT enzyme (EC 2.3.1.13). The reason for chosing GLYAT is that it catalyses the removal of the potentially toxic acyl-CoAs (endotoxins and xenobiotics) from mammalian circulation. The product N-acylglycine is produced by GLYAT in an effort to remove the accumulated acyl-CoAs from circulation. Furthermore, xenobiotics are compounds from outside the body that the body can not metabolise any further and need to remove from circulation due to their potentially harmful nature (Caldwell, 1986).

Neither the genomes nor the open reading frames of genes encoding GLYAT enzymes in both chacma baboon and vervet monkey have been sequenced. Characterising GLYAT from other species will give insight to the diversity of GLYAT.

Thus, the aim for the project was to compare the molecular characteristics of GLYAT of the vervet monkey and the chacma baboon with that currently known in literature. In this report the nucleic acid and deduced amino acid sequences of novel open reading frames encoding for GLYAT of the chacma baboon (*Papio ursinus*) and vervet monkey (*Chlorocebus pygerythrus)* are reported (Figure 2.13). We envisage using information on diversity of GLYAT in the rational design of a recombinant GLYAT that can catalyse a larger spectrum of substrates.

4.2 DNA sequencing of the chacma baboon GLYAT transcript and the vervet monkey GLYAT open reading frames

A liver of chacma baboon and blood sample from a vervet monkey were obtained to investigate GLYAT molecular characteristics. PCR of GLYAT ORFs of the chacma baboon and the vervet monkey was done using GLYAT primers designed from the human GLYAT gene. Human genes have been used before to design primers that worked on baboons and vervet monkeys (Jasiska *et al*., 2007). Amplicons containing the open reading frames encoding GLYAT of both chacma baboon and vervet monkey

were compared to that available in the scientific public domain encoding GLYAT of humans, rhesus monkeys and chimpanzees.

A nucleic acid sequence of 1256 bases of the chacma baboon transcript was obtained. The open reading frames encoding GLYAT of chacma baboon and vervet monkey were translated into an amino acid sequence. The chacma baboon transcript contained an open reading frame of 891 bases that translated into 296 amino acids (Figure 2.12). The transcript had 122 bases of untranslated region flanking the 5'-end of the ORF and 243 bases of untranslated region at the 3'-end of the coding strand. The vervet monkey GLYAT open reading frame was 891 bases long and translated into 296 amino acids (Figure 2.13). The translated amino acids representing GLYAT of chacma baboon and vervet monkey were aligned to those corresponding to the human, chimpanzee and rhesus monkey, which were obtained from the Ensembl sequence database. Their translated amino acid sequences (Figure 2.13) suggested 8 variations relative to the corresponding enzyme of humans in the region encoding the core GNAT fold. There were 4 new variations introduced by GLYAT amino acid sequences from the chacma baboon and the vervet monkey. The vervet monkey introduced an isoleucine in place of a valine at position 32 and an arginine in place of a histidine or glutamine at position 224. The chacma baboon introduced a tyrosine in place of isoleucine at position 201 and an arginine in place of histidine or glutamine at position 240.The implications of these substitutions are unknown at the moment.

The GLYAT family belongs to GNAT superfamily that has the characteristic of containing a conserved GNAT fold (Figure 1.2). It is known in literature that the conserved core GNAT fold is made up of amino acids from the C-terminal which runs through amino acids at positions 207 to 296 of the protein. The N-terminal is composed of positions 1 to 206 of the protein. The conserved core GNAT fold is described as having an active function in binding and catalysis of the substrate. The proposed importance of the GNAT fold does not prohibit any other amino acids from interacting with substrate or interfering with the reaction mechanism at any stage. The C-terminal of the GLYAT protein from chacma baboon was compared to templates on Swiss-Model online server and it was observed that position 201 of GLYAT is located on the beta sheet region on the N-terminal of the GNAT fold.

A phylogenic tree analysis of GLYAT from primates (Figure 2.15) showed that GLYAT of the vervet monkey, chacma baboon and rhesus monkey formed a distinct group separate from that from humans and chimpanzee. The chacma baboon and vervet monkey GLYAT amino acid sequences appeared to be 99% identical to the rhesus monkey sequence and 92.6% identical to the human sequence, while the human and chimpanzee GLYAT sequences appeared to be 99.7% identical to each other. The chacma baboon and vervet monkey novel GLYAT nucleic acid sequences showed to be 99% similar to the rhesus monkey nucleic acid and 92.6 % similar to the human nucleic acid sequence.

The chimpanzee has always been the non-human primate that is genetically closer to the human than the rest (Caccone and Powell, 1989). In fact, some workers found that single base substitutions lead to only 1.2 % difference between the human and the chimpanzee when they analysed the genes they were working on (Chen *et al*, 2005). The difference of 1.2 % between the chimpanzee and the human GLYAT sequences coincides with the 99 % similarity found in this project. Other workers found that the human GLYAT was 92 % identical to that of the rhesus monkey (Cardenas *et al*., 2010). Thus, the results found in this study are consistant with current knowledge in the literature. Both the chacma baboon and the vervet monkey GLYAT have Ser at position 17 and Arg at position 199 of the protein. The two positions (Ser17Thr and Arg199Cys) were suggested to have a potential role in GLYAT activity (Cardens *et al*., 2010).

Glutamate at position 226 (E226) of bovine GLYAT corresponding to position 227 (E227) in primates has been pointed out as having a possible role in the catalytic mechanism of GLYAT (Badenhorst *et al*., 2010). This position was selected because it was the most conserved position in GNAT superfamily members along the amino acids sequence making up the conserved GNAT fold of the GNAT superfamily. Badenhorst and co-workers (MSc of Mr CPS Badenhorst, 2009) used site-directed mutagenesis to illustrate the fact that E227 is of catalytic importance. Badenhorst and co-authors concluded that any GLYAT member without the glutamate residue at that position may not have any catalytic activity. The glutamate (E227) of GLYATL3 appeared to be substituted with threonine (T227). However, it is possible that the threonine residue is

characteristic of either the GLYATL3 enzyme or the specific individual animal that contributed the sequence.

4.3 Bacterial expression of a recombinant chacma baboon GLYAT

A partially soluble and enzymatically active bovine recombinant GLYAT protein has been expressed in *E. coli* in our laboratory [by Ms M Snyders (MSc in preparation) and Badenhorst *et al*. 2010]. One of the objectives of this report was to express a soluble enzymatically active chacma baboon recombinant GLYAT in *E. coli*.

An experiment to express a soluble recombinant chacma baboon GLYAT together with the chaperones groEL, groES and Tf was conducted. The objective was to compare enzymatic activity of the native GLYAT purified from the chacma baboon liver with that of the recombinant chacma baboon GLYAT. High expression of the recombinant chacma baboon GLYAT (with a His-tag at the C-terminal) was obtained (Figure 3.5). However, it was completely insoluble. Chaperone expression was unsuccessful and attempts to express chaperones at maximum chaperone inducer concentration (tetracycline) prescribed by the chaperone manufacturer also failed. The recombinant chacma baboon GLYAT protein was not purified further because it did not seem like it would be a fruitful effort after visualizing no soluble recombinant chacma baboon GLYAT on SDS-PAGE (Figure 3.5 and 3.6).

An enzymatically active recombinant human GLYATL2 has been expressed in *E. Coli* cells, BL21(DE3)pLysS strain, (Waluk *et al*, 2010). GLYATL1 has been expressed and showed activity in mammalian kidney cells, HEK 293 cells, (Zhang *et al*, 2007). However, there has not been a report about a primate recombinant GLYAT. There has been a failed attempt to express a recombinant human GLYAT in *E. coli,* BL21 strain, (van der Westhuizen, 1998). To solve the problem of insoluble recombinant protein, a chacma baboon recombinant GLYAT protein could be expressed in a mammalian expression system or in a bacterial expression system with molecular chaperones.

4.4 Preparation of a crude cytoplasmic extract containing native glycine-N-acyltransferase from chacma baboon liver

Another objective for my MSc project was to purify a native GLYAT protein from a chacma baboon liver. The purpose being to compare the characteristics of a GLYAT protein purified from liver with those of a recombinant GLYAT. A crude cytoplasmic extract was successfully prepared from chacma baboon liver as described by van der Westhuizen (1998). An enzyme activity test was done using the crude cytoplasmic extract from liver to determine if it had any GLYAT activity as described before (Ellman, 1954; Gron *et al*, 1976; Kolvraa and Gregersen, 1986). Benzoyl-CoA and glycine have been shown to be the best substrates for the GLYAT reaction (Schachter and Taggard, 1954). The GLYAT activity in the crude cytoplasmic extract from liver of chacma baboon was compared with that from a crude cytoplasmic extract from a bovine liver (prepared by Ms M Snyders). Both extracts had GLYAT activity. The amount of GLYAT in the two preparations could not be quantified and therefore the activity of the two enzymes could not be compared. The enzyme phenylacetyltransferase also favours the substrates benzoyl-CoA and glycine but this enzyme may not have participated in this case because it has been shown that this enzyme requires addition of monovalent cations such as K^+, Rb^+, Na^+, Li^+, Cs^+ or NH_4^+ (Nandi *et al*., 1979).

A native GLYAT protein has been purified from both bovine and human liver by workers from our Department before (van der Westhuizen *et al*., 2000; van der Westhuizen, 1998).

4.5 Future prospects

The E227T variation brought by the new GLYATL3 (Waluk *et al*., 2010) may have significance in terms of enzyme activity. Thus, the E227 position needs to be revaluated and its importance investigated in more detail. In the future, it may be fruitful to sequence the complete GLYAT gene from both the vervet monkey and chacma baboon in order to analyse these genes for SNPs.

It was not possible to produce a soluble recombinant GLYAT. A soluble recombinant GLYAT could be pursued by using different expression systems and vectors. The comparison in substrate specificity between the native and recombinant GLYAT may help get us closer to characterising the GLYAT substrate binding and active sites.

In the last 20 years, GLYAT has been implicated in a variety of pathways in addition to detoxification. GLYAT was one of the enzymes that were implicated in the metabolic excretion of nitrogen in monkeys (Kazuo, 1991). GLYAT is implicated in signaling pathways of the fatty acid primary amides (Merkler *et al*., 1996; Farrel and Merkler, 2008). Thus, GLYAT is no longer confined to be just a detoxification enzyme. This indicates that there could be many more unknown functions of GLYAT in the mammalian body. The recent discovery of additional members of the GLYAT family (GLYATL1 [Zhang *et al*, 2007], GLYATL2 [Zhang *et al*, 2007] and GLYATL3 [Waluk *et al*, 2010]) suggests that GLYAT could be more important and have more diverse

functions than it is thought in literature. The fact that two recombinant members of the GLYAT family (human GLYATL1 and GLYATL2) have been expressed and that GLYATL2 was enzymatically active (Waluk *et al*., 2010) suggests that we should continue our efforts to generate a recombinant set of GLYATs from different species.

List of references

Badenhorst, C. P. S. 2009. Molecular characterisation of a recombinant bovine glycine N-acyltransferase. MSc dissertation. North West University, Potchefstroom.

Bandmann, O., Vaughan, J., Holmans, P., Marsden, C. D. and Wood, N. W. 1997. Association of slow acetylator genotype for N-acetyltransferase 2 with familial Parkinson's disease. The Lancet 350:1136-1139.

Baneyx, F. 1999. Recombinant protein expression in Escherichia coli. Current Opinion in Biotechnology 10:411-421.

Bartlett, K. and. Gompertz., D. 1974. The specificity of glycine-N-acylase and acylglycine excretion in the organicacidaemias. Biochemical Medicine 10:15-23.

Boom, R., Sol, C. J.A., Salimans, M. M.M., Jansen, C. L., Wertheim-van Dillen, P. M.E. and van der Noordaa, J. 1990. Rapid and simple method for purification of nucleic acids. Journal of Clinical Microbiology 28:495-503.

Bowcock, A. M., Ray, A.E., Henry, S., and Pravinkumar B. 1989. Rapid detection and sequencing of alleles in the 3flanking region of the Interleukin -6 gene. Nucleic Acids Research 17:6855-6864.

Bowery, N. G., and Smart, T. G. 2006. GABA and glycine as neurotransmitters: a brief history. British Journal of Pharmacology 147:S109-S119.

Caccone, A., and Powell, J. R. 1989. DNA Divergence Among Hominoids. Evolution 43:925-942.

Caldwell, J. 1986. Conjugation Mechanisms of Xenobiotic Metabolism: Mammalian Aspects, p. 2-28, Xenobiotic Conjugation Chemistry, vol. 299. American Chemical Society.

Cardenas, C. L. L., Bourgine, J., Cauffiez, C., Allorge, D., Lo-Guidice, J.M., Broly, F., and Chevalier, D. 2010. Genetic polymorphisms of Glycine N-acyltransferase (GLYAT) in a French Caucasian population. Xenobiotica 0:1-9.

Chakravrati. A. 2001. Single base differences between human genomes underlie differences in susceptibility to, or protection from, a host of diseases. Hence the great potential of such information in medicine. Nature 409:822-823.

Chandra, T. S., Robin, K., Vincent, J. W., and Savio L. C. 1983. Isolation and Sequence Characterization of a cDNA Clone of Human Antithrombin III. Proceedings of the National Academy of Sciences of the United States of America 80:1845-1848.

Chantrenne, H. 1951. The requirement for coenzyme A in the enzymatic synthesis of hippuric acid. Journal of Biological Chemistry 189:227-233.

Cheng, Z., Ventura, M. S., Xinwei, K., Philipp, Graves., Tina, O., Kazutoyo, C, Deanna D., Pieter, W, Richard, K. P., Svante, R., Mariano, E., and Evan, E. 2005. A genome-wide comparison of recent chimpanzee and human segmental duplications. Nature 437:88-93.

Chirgwin, J. M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. 1979. Isolation of Biologically Active Ribonucleic Acid from Sources Enriched in Ribonucleaset. American Chemical Society 18:5294-5299.

Coles, B. F., and Kadlubar, F.F. 2003. Detoxification of electrophilic compounds by glutathione S-transferase catalysis: Determinants of individual response to chemical carcinogens and chemotherapeutic drugs? BioFactors 17:115-130.

International Human Genome Sequencing Consortium Eric S. Lander, L. M. L., Bruce Birren, Chad Nusbaum, Michael C. Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William FitzHugh, Roel Funke, Diane Gage, Katrina Harris, Andrew Heaford, John Howland, Lisa Kann, Jessica Lehoczky, Rosie LeVine, Paul McEwan, Kevin McKernan, James Meldrim, Jill P. Mesirov, Cher Miranda, William Morris, Jerome Naylor, Christina Raymond, Mark Rosetti, Ralph Santos, Andrew Sheridan, Carrie Sougnez, Nicole Stange-Thomann, Nikola Stojanovic, Aravind Subramanian & Dudley

Wyman for Whitehead Institute for Biomedical Research, Center for Genome Research:, Jane Rogers, John Sulston, Rachael Ainscough, Stephan Beck, David Bentley, John Burton, Christopher Clee, Nigel Carter, Alan Coulson, Rebecca Deadman, Panos Deloukas, Andrew Dunham, Ian Dunham, Richard Durbin, Lisa French, Darren Grafham, Simon Gregory, Tim Hubbard, Sean Humphray, Adrienne Hunt, Matthew Jones, Christine Lloyd, Amanda McMurray, Lucy Matthews, Simon Mercer, Sarah Milne, James C. Mullikin, Andrew Mungall, Robert Plumb, Mark Ross, Ratna Shownkeen & Sarah Sims for The Sanger Centre:, Robert H. Waterston, Richard K. Wilson, LaDeana W. Hillier, John D. McPherson, Marco A. Marra, Elaine R. Mardis, Lucinda A. Fulton, Asif T. Chinwalla, Kymberlie H. Pepin, Warren R. Gish, Stephanie L. Chissoe, Michael C. Wendl, Kim D. Delehaunty, Tracie L. Miner, Andrew Delehaunty, Jason B. Kramer, Lisa L. Cook, Robert S. Fulton, Douglas L. Johnson, Patrick J. Minx & Sandra W. Clifton for Washington University Genome Sequencing Center, Trevor Hawkins, Elbert Branscomb, Paul Predki, Paul Richardson, Sarah Wenning, Tom Slezak, Norman Doggett, Jan-Fang Cheng, Anne Olsen, Susan Lucas, Christopher Elkin, Edward Uberbacher & Marvin Frazier for US DOE Joint Genome Institute:, Richard A. Gibbs, Donna M. Muzny, Steven E. Scherer, John B. Bouck, Erica J. Sodergren, Kim C. Worley, Catherine M. Rives, James H. Gorrell, Michael L. Metzker, Susan L. Naylor, Raju S. Kucherlapati, David L. Nelson & George M. Weinstock for Baylor College of Medicine Human Genome Sequencing Center:, Yoshiyuki Sakaki, Asao Fujiyama, Masahira Hattori, Tetsushi Yada, Atsushi Toyoda, Takehiko Itoh, Chiharu Kawagoe, Hidemi Watanabe, Yasushi Totoki & Todd Taylor for RIKEN Genomic Sciences Center:, Jean Weissenbach, Roland Heilig, William Saurin, Francois

Artiguenave, Philippe Brottier, Thomas Bruls, Eric Pelletier, Catherine Robert & Patrick Wincker for Genoscope and CNRS UMR-8030:, André Rosenthal, Matthias Platzer, Gerald Nyakatura, Stefan Taudien & Andreas Rump for Department of Genome Analysis, Institute of Molecular Biotechnology:, Douglas R. Smith, Lynn Doucette-Stamm, Marc Rubenfield, Keith Weinstock, Hong Mei Lee & JoAnn Dubois for GTC Sequencing Center:, Huanming Yang, Jun Yu, Jian Wang, Guyang Huang & Jun Gu for Beijing Genomics Institute/Human Genome Center:, Leroy Hood, Lee Rowen, Anup Madan & Shizen Qin for Multimegabase Sequencing Center, The Institute for Systems Biology:, Ronald W. Davis, Nancy A. Federspiel, A. Pia Abola & Michael J. Proctor for Stanford Genome Technology Center:, Bruce A. Roe, Feng Chen & Huaqin Pan for University of Oklahoma's Advanced Center for Genome Technology:, Juliane Ramser, Hans Lehrach & Richard Reinhardt for Max Planck Institute for Molecular Genetics:, W. Richard McCombie, Melissa de la Bastide & Neilay Dedhia for Cold Spring Harbor Laboratory, Lita Annenberg Hazen Genome Center:, Helmut Blöcker, Klaus Hornischer & Gabriele Nordsiek for GBF—German Research Centre for Biotechnology:, Richa Agarwala, L. Aravind, Jeffrey A. Bailey, Alex Bateman, Serafim Batzoglou, Ewan Birney, Peer Bork, Daniel G. Brown, Christopher B. Burge, Lorenzo Cerutti, Hsiu-Chuan Chen, Deanna Church, Michele Clamp, Richard R. Copley, Tobias Doerks, Sean R. Eddy, Evan E. Eichler, Terrence S. Furey, James Galagan, James G. R. Gilbert, Cyrus Harmon, Yoshihide Hayashizaki, David Haussler, Henning Hermjakob, Karsten Hokamp, Wonhee Jang, L. Steven Johnson, Thomas A. Jones, Simon Kasif, Arek Kaspryzk, Scot Kennedy, W. James Kent, Paul Kitts, Eugene V. Koonin, Ian Korf, David Kulp, Doron Lancet, Todd M. Lowe, Aoife McLysaght, Tarjei Mikkelsen, John V. Moran,

Nicola Mulder, Victor J. Pollara, Chris P. Ponting, Greg Schuler, Jörg Schultz, Guy Slater, Arian F. A. Smit, Elia Stupka, Joseph Szustakowki, Danielle Thierry-Mieg, Jean Thierry-Mieg, Lukas Wagner, John Wallis, Raymond Wheeler, Alan Williams, Yuri I. Wolf, Kenneth H. Wolfe, Shiaw-Pyng Yang & Ru-Fang Yeh for *Genome Analysis Group (listed in alphabetical order, also includes individuals listed under other headings):, Francis Collins, Mark S. Guyer, Jane Peterson, Adam Felsenfeld & Kris A. Wetterstrand for Scientific management: National Human Genome Research Institute, US National Institutes of Health:, Richard M. Myers, Jeremy Schmutz, Mark Dickson, Jane Grimwood & David R. Cox for Stanford Human Genome Center:, Maynard V. Olson, Rajinder Kaul & Christopher Raymond for University of Washington Genome Center:, Nobuyoshi Shimizu, Kazuhiko Kawasaki & Shinsei Minoshima for Department of Molecular Biology, Keio University School of Medicine:, Glen A. Evans, Maria Athanasiou & Roger Schultz for University of Texas Southwestern Medical Center at Dallas:, Aristides Patrinos for Office of Science, US Department of Energy: & Michael J. Morgan for The Wellcome Trust:. 2001. Initial sequencing and analysis of the human genome. Nature 409:860-921.

Below are those in the International Human Genome Sequencing Consortium that participated in the finishing of the human reference sequence beyond the publication of the draft human genome (Nature 409 861-921 (2001)). They are grouped into sequencing centers, analysis groups, and scientific management. Some center lists are further subdivided.

1. The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1RQ, United Kingdom

Zahra Abdellah, Alireza Ahmadi, Shahana Ahmed, Matthew Aimable, Rachael Ainscough, Jeff Almeida, Claire Almond, Andrew Ambler, Karen Ambrose, Kerrie Ambrose, Robert Andrew, Daniel Andrews, Neil Andrews, Dan Andrews, Eva Apweiler, Hazel Arbery, Beth Archer, Gareth Ash, Kevin Ashcroft, Jennifer Ashurst, Robert Ashwell, Deborah Atkin, Andrea Atkinson, Barry Atkinson, John Attwood, Keith Aubin, Katherine Auger, Terry Avis, Anne Babbage, Sarah Babbage, Joanne Bacon, Claire Bagguley, Jonathan Bailey, Andrew Baker, Ruby Banerjee, Simon Bardill, Darren Barker, Gary Barker, Daniel Barker, Karen Barlow, Laurent Baron, Anika Barrett, Rebecca Bartlett, David Basham, Victoria Basham, Alex Bateman, Karen Bates, Caroline Baynes, Lisa Beard, Susan Beard, David Beare, Alastair Beasley, Helen Beasley, Oliver Beasley, Stephan Beck, Emma Bell, Damian Bellerby, Tristram Bellerby, Richard Bemrose, James Bennett, David Bentley, Andrew Bentley, Mary Berks, Michael Berks, Graeme Bethel, Christine Bird, Ewan Birney, Helen Bissell, Suzanne Blackburne-Maze, Sarah Blakey, Christel Bolton, James Bonfield, Ralph Bonnett, Richard Border, Amanda Bradley, Nicola Brady, Jason Bray, Sarah Bray-Allen, Anne Bridgeman, Jonathan Brook, Shane Brooking, Andrew Brown, Clive Brown, Jacqui Brown, Margaret Brown, Mary Brown, Richard Bruskiewich, Jackie Bryant, David Buck, Veronica Buckle, Claire Budd, Sarah Buller, Jill Burberry, Deborah Burford, Joanne Burgess, Wayne Burrill, Christine Burrows, John Burton, Christine Burton, Phil Butcher, Adam Butler, Murray Cairns, Nick Camm, Christopher Campbell, Bruno

Canning, Carol Carder, Paul Carder, Nigel Carter, Tamara Cavanna, Shani Chalk, Ka Chan, Joanna Chapman, Rachel Charles, Neil Chillingworth, Tom Chothia, Connie Chui, Rob Clack, Michele Clamp, Anthea Clark, Graham Clark, Kevin Clark, Sarah Clark, Sue Clark, Richard Clark, Betty Clarke, Eddie Clarke, Kay Clarke, Adrian Clarke, Laura Clarke, Chris Clee, Sheila Clegg, Karen Clifford, Julia Coates, Victoria Cobley, Alison Coffey, Penelope Coggill, Lotte Cole, Rachael Collier, Simon Collings, John Collins, Philip Collins, Louise Colman, Anthony Connolly, Richard Connor, Jennie Conquer, Donald Conroy, Doug Constance, Leanna Cook, Jonathan Cooper, Rachel Cooper, Robert Cooper, Maria Coppola, Teresa Copsey, Nicole Corby, Linda Cornell, Ruth Cornell, Christine Cornell, Amanda Cottage, Alan Coulson, Gez Coville, Anthony Cox, Tony Cox, Robert Coxhill, Matthew Craig, Tom Crane, Matt Crawley, Victor Crew, James Cuff, Karl Culley, Auli Cummings, Kirsti Cummings, Paul Cummings, Adam Curran, Valery Curwen, Jeffrey Cutts, Rachael Daniels, Lucy Davidson, Jonathon Davies, Joy Davies, Nicholas Davies, Robert Davies, John Davis, Jayne Davis, Matthew Davis, Elisabeth Dawson, Rebecca Deadman, Peter Dean, Simon Dear, Frances Dearden, Marcos Delgado, Panos Deloukas, Janet Dennis, Pawandeep Dhami, Catherine Dibling, Ruth Dobbs, Richard Dobson, Catherine Dockree, Daniel Doddington, Steven Dodsworth, Norman Doggett, Thomas Down, Andrew Dunham, Ian Dunham, Anne Dunn, Matthew Dunn, Richard Durbin, Jillian Durham, Ireena Dutta, Ruth Dwyer, Lauren Dyer, Mark Earthrowl, Timothy Eastham, Emma Eastham, Carol Edwards, Karen Edwards, Andrew Ellington, David Elliott, Matthew Ellwood, Becky Emberson, Helen Errington, Gareth Evans, John Evans, Katie Evans, Richard Evans, Eduardo Eyras, Louisa Faulkner, Charlotte Fellingham, Theresa Feltwell, Stephen

Fennell, Robert Finn, Tina Flack, Claire Felming, Kerry Fleming, Jonathan Flint, Mark Flint, Yvonne Floyd, Simon Footman, John Fowler, Deborah Frame, Matthew Francis, Stephen Francis, Adam Frankish, John Frankland, Audrey Fraser, David Fraser, Lisa French, David Fricker, Daniel Frost, Jackie Frost, Lorna Frost, Carole Frost, Liam Fuller, Kathryn Fullerton, Alison Gardner, Patrick Garner, Jane Garnett, Leigh Gatland, Lindsay Gatland, Jilur Ghori, Ben Gibbs, Diane Gibson, Elizabeth Gibson, James Gilbert, Lisa Gilby, Christopher Gillson, Rebecca Glithero, April Gooderham, Matthew Gorton, Darren Grafham, Michael Grant, Susan Grant, Iain Gray, Emma Gray, Lisa Green, James Greenhalgh, Joe Greenhill, Sam Griffiths-Jones, Philippa Gregg, Simon Gregory, Susan Gribble, Coline Griffiths, Ed Griffiths, Mark Griffiths, Russell Grocock, Ian Guthrie, Rhian Gwilliam, Rebekah Hall, Karen Halls, Gretta Hall-Tamly, John Hamlett, John Hamlett, Sian Hammond, Julie Hancock, Adam Harding, Joanne Harley, David Harper, Georgina Harper, Patrik Harper, Grant Harradence, Charlene-Lou Harrison, Elliott Harrison, Ruth Harrison, Elizabeth Hart, Daniel Hassan, Natalie Hawkins, Kellie Hawley, Kerry Hayes, Paul Heath, Rosemary Heathcott, Cathy Hembry, Carl Henderson, Tim Herd, Stephen Hewitt, Douglas Higgs, Guy Hillyard, Russell Hinkins, Sara-Jane Ho, David Hodgson, Michael Hoffs, Jane Holden, Janet Holdgate, Ele Holloway, Ian Holmes, Sarah Holmes, Simon Holroyd, Alison Hooper, Lucy Hopewell, Ben Hopkins, Gary Hornett, Geoff Hornsby, Tony Hornsby, Sharon Horsley, Roger Horton, Philip Howard, Philip Howden, Kevin Howe, Gareth Howell, Timothy Hubbard, Elizabeth Huckle, Jaime Hughes, Jennifer Hughes, Louisa Hull, Holger Hummeric, Sean Humphray, Matthew Humphries, Adrienne Hunt, Paul Hunt, Sarah Hunt, Giselle Hunter, David Hyde, Michael Ince, Judith Isherwood, Vivek Iyer, Janet Izatt, Monica

Izmajlowicz, Niclas Jareborg, Bijay Jassal, Grant Jeffery, Kim Jeffery, Colin Jeffrey, Kerstin Jekosch, Lee Jenkins, Tina Johansen, Cheryl Johnson, Christopher Johnson, David Johnson, Keith Jolley, Abigail Jones, Claire Jones, Juliet Jones, Matthew Jones, Michael Jones, Steven Jones, Shirin Joseph, Ann Joy, Linsey Joy, Victoria Joy, Gillian Joyce, Mark Jubb, Kanchi Karunaratne, Michael Kay, Danielle Kaye, Lyndal Kearney, Stephen Keenan, Simon Kelley, Joanna Kershaw, Ross Kettleborough, Cathy Kidd, Peter Kierstan, Andrew Kimberley, Andrew King, Simon Kingsley, Colin Kingswood, Gillian Klingle, Andrew Knights, Ian Korf, Anders Krogh, Heena Lad, Philip Laidlaw, Michael Laing, Gavin Laird, Christine Lambart, Ralph Lamble, Cordelia Langford, Ben Larke, Timun Lau, Stephanie Lawlor, Sampsa Leather, Minna Lehvaslaiho, Johannes Lemke, Steven Leonard, Daniel Leongamornlert, Margaret Leversha, Julia Lightning, Sarah Lindsay, Matthew Line, Sally Linsdell, Peter Little, Christine Lloyd, David Lloyd, Victoria Lock, William Lock, Anne Lodziak, Ian Longden, Howard Loraine, Rachel Lord, Jane Loveland, Jamie Lovell, Georgina Lye, Neil Marriott, Anna Marrone, Paul Marsden, Victoria Marsh, Matthew Martin, Sancha Martin, Gareth Maslen, Debbie Mason, Lucy Matthews, Paul Matthews, Nick Matthews, Madalynne Maynard, Owen McCann, Joseph McClay, Craig McCollum, Louise McConnachie, Bill McDonald, Louise McDonald, Jennifer McDowall, Sarah McGuire, Carole McKeown, Stuart McLaren, Kirsten McLay, James McLean, John McMurdo, Amanda McMurray, Des McMurray, Natalie McWilliams, Nalini Mehta, Patrick Meidl, Noel Menuge, Simon Mercer, Asab Miah, Gos Micklem, Simon Miles, Sarah Milne, Dippica Mistry, Shailesh Mistry, Jake Mitchell, Jeff Mitchell, Maryam Mohammadi, Christophe Molina, Paul Mooney, Madeline Moore, Andrea Moreland, Beverley Mortimore, Richard Mott, Ian Mullenger, Jim

Mullikin, Brian Munday, Elaine Munday, Andy Mungall, Clare Murnane, Kerry Murrell, Alison Myers, David Negus, Bee Ng, David Niblett, Jonathan Nicholson, Tim Nickerson, Sukhjit Nijjar, Zemin Ning, James Nisbet, Karen Novik, Christopher Odell, Daniel O'Donovan, Francess Ogbighele, Tom Oinn, Hayley Oliver, Karen Oliver, Helena Orbell, Anthony Osborn, Joan Osborne, Emma Overton-Larty, Sophie Palmer, Richard Pandian, Adrian Parker, Christopher Parkin, Kim Parkin, Ginny Parry-Brown, Dina Patel, Ritesh Patel, Alexandra Pearce, Danita Pearson, Anna Peck, Richard Peck, John Peden, Sarah Pelan, Chantal Percy, Andrew Perito, Isabelle Perrault, Anna Peters, Roger Pettett, Ben Phillimore, Kim Phillips, Samantha Phillips, Darren Platt, Emma Playford, Bob Plumb, Matthew Pocock, Keith Porter, Tarryn Porter, Christopher Potter, Simon Potter, Don Powell, Radhika Prathalingham, Elena Prigmore, Michael Quail, Hanna Quarrie, Chris Quince, Matloob Qureshi, Helen Ramsay, Yvonne Ramsey, Sally Ranby, Richard Rance, Vikki Rand, Joanne Ratford, Lewis Ratford, Daniel Read, Donald Redhead, Richard Redon, Christine Rees, Mary Reid, Astrid Reinhardt, Alex Rice, Catherine Rice, Peter Rice, Suzzanne Richard, Susan Richardson, Kerry Ridler, Lyn Riethoven, Rachel Rigby, Melanie Robinson, Rebecca Rochford, Jane Rogers, Lisa Rogers, Hugh Ross, Mark Ross, Angela Rule, James Rule, Ben Russell, Jayne Rutter, Kamal Safdar, Natalie Salter, Javier Santoyo-Lopez, David Saunders, Carol Scott, Deborah Scott, Ian Scott, Fiona Seager, Margaret Searle, Paul Searle, Stephen Searle, Harminder Sehra, Joe Shakespeare, Jason Shardelow, Greg Sharp, Teresa Shaw, Charles Shaw-Smith, Jennifer Shearing, Karen Sheppard, Richard Sheppard, Elizabeth Sheridan, Ratna Shownkeen, Richard Silk, Matthew Sims, Sarah Sims, Shanthi Sivadasan, Carl Skuce, Luc Smink, Andrew Smith, Laura Smith, Lorraine Smith,

Michelle Smith, Russell Smith, Stephanie Smith, James Smith, Hannah Sneath, Cari Soderlund, Victor Solovyev, Erik Sonnhammer, Elizabeth Sotheran, William Spooner, Lee Spraggon, Janet Squares, Suzanna Squares, Michael Stables, James Stalker, Steve Stamford, Melanie Stammers, Helen Steingruber, Yvonne Stephens, Charles Steward, Aengus Stewart, Michael Stewart, Ian Still, Mo Stock, Lisa Stoppard, Philip Storey, Roy Storey, Sarah Stowe, Carol Strachan, Greg Strachan, Claire Stribling, John Sturdy, John Sulston, Chris Swainson, Mark Swann, David Swarbreck, Neil Sycamore, Matthew Tagney, Steven Tan, Elizabeth Tarling, Amy Taylor, Gillian Taylor, Kate Taylor, Ruth Taylor, Sam Taylor, Susan Taylor, Louise Tee, Julieanne Tester, Andrew Theaker, David Thexton, Craig Thomas, Daniel Thomas, Karen Thomas, Ruth Thomas, Roselin Thommai, Andrea Thorpe, Karen Thorpe, Glen Threadgold, Scott Thurston, Emma Tinsley, Alan Tracey, Jonathan Travers, Anthony Tromans, Ben Tubby, Cristina Tufarelli, Kathryn Turney, Darren Upson, Zoe Van Helmond, Mark Vaudin, Ramya Viknaraja, Wendy Vine, Paul Voak, David Vollenhoven, Sarah Walker, Melanie Wall, Justine Wallis, Michelle Wallis, Graham Warren, Georgina Warry, Andy Watson, Nicola Watt, Anthony Webb, Jeannette Webb, John Weir, Alan Wells, Sarah Wells, Robert Welton, Paul West, Tony West, Angela Wheatley, Carl Wheatley, Gideon Wheeler, Hayley Whitaker, Adam White, Amelia White, Brian White, Johnathon White, Simon White, Sally Whitehead, Matthew Whiteley, Pamela Whittaker, Adam Whittaker, Sara Widaa, Anna Wild, Jane Wilkinson, Paul Wilkinson, David Willey, Andy Williams, Bill Williams, Leanne Williams, Helen Williamson, Tamsin Wilmer, Laurens Wilming, Brian Wilson, Gareth Wilson, Margaret Wilson, Nyree Wilson, Siobhan Wilson, Wendy Wilson, Piers Wilson, Philip Window, Jenny Winster, Claire Winzar, James Witt, Fred Wobus,

Emma Wood, Joe Wood, Sharon Woodeson, Kathryn Woodfine, Rebecca Woodhouse, Rebecca Woodmansey, Richard Wooster, Matthew Wray, Paul Wray, Charmain Wright, Kathrine Wright, Debbie Wright, Julia Wyatt, Jane Xie, Louise Young, Sheila Young, Ruth Younger, Shenru Zhao

2. Washington University Genome Sequencing Center, Box 8501, 4444 Forest Park Avenue, St. Louis, Missouri 63108 USA

Amanda Abbott-Ozersky, Amber Isak, Amy Berghoff, Amy D. Reily, Andrea Holmes, Andrew Levy, Andrew Van Brunt, Aniko Sabo, Anthony Harris, Anu Desai, Asif Chinwalla, Aye M. Tin-Wollam, Betty Lamar, Brian Mailey, C. Richard Harkins, Caryn Wagner-McPherson, Catherine Marquis-Homeyer, Catrina Fronick, Chad Tomlinson, Charlene Pearman, Christine Nguyen, Chunyan Wang, Colin Kremitzki, Craig Pohl, Cynthia Strong, Dan Layman, Dan Bentley, Darin Blasiar, David Dooling, Delali Buatsi, Douglas Johnson, Edward A. Belter, Jr., Edward Paulson, Elaine R. Mardis, Elizabeth Boatright, Ernest Goyea, Feiyu Du, George Hu, Glendoria Elliott, Holly Bradshaw-Cordum, Hui Du, Hui Sun, James Randolph, Jason Carter, Jason E. Waligorski, Jason Maas, Jeffrey Woessner, Jennifer Edwards, Jennifer Murray, Jennifer Randall-Maher, Joanne Nelson, Joelle Kalicki-Veizer, Johar Ali, John McPherson, John Spieth, John Wallace, Jon Armstrong, Joshua Heyen, Kelly Carpenter, Kelie Kang, Kelly Mead, Kelsi Scott, Kimberley Delhaunty, Kyriena L. Schatzkamer, Kris Wylie, Krista Haglund, Kym Hallsworth-Pepin, Kyung Kim, Lachlan Oddy, LaDeana Hillier, Laura P. Courtney, Lauren Caruso, Lee Trani, Li Ding, Lucinda A. Fulton, Maria Cedroni, Marco Marra, Mark Johnson, Martin Yoakum, Matt Cordes, Maxim Radionenko, Michael Becker,

Michael D. McLellan, Mike Nhan, Mike Wendl, Mundeep Sekhon, Nancy Miller, Neenu Grewal, Neha Shah, Nicolas Berkowicz, Patrick J. Minx, Patty Wohldmann, Phil Latreille, Philip Ozersky, Prashant Sinha, Rachel Maupin, Rekha Meyer, Rick Meyer, Richard K. Wilson, Robert S. Fulton, Robert H. Waterston, Sandra W. Clifton, Sara Kohlberg, Sara Jaeger, Shiaw-Pyng Yang, Scott Abbott, Scott Martinka, Scott S. Kruchowski, Sharhonda Swearengen-Shahid, Shawn Leonard, Shunfang Hou, Stacie Gattung, Stephanie Andrews, Stephanie Chissoe, Susan M. Rock, Tamberlyn Bieri , Theresa Rohlfing, Tina A. Graves, Tony Gaige, Tracie Miner, William Nash, Yoram Shotland

3. Whitehead Institute for Biomedical Research, Center for Genome Research, Nine Cambridge Center, Cambridge, Massachusetts 02142 USA.

Current address: Eli and Edythe Broad Institute, 320 Charles Street, Cambridge, MA 02141-2023 (§Center Director; ‡Project Leadership; *Major Contributor)

Eric S Lander§, Chad Nusbaum‡, Bruce Birren‡, Kerstin Lindblad-Toh‡, Jill P Mesirov‡, Robert Nicol‡, Michael C Zody‡, Jean L Chang‡, Christina A Cuomo‡, Ken Dewar‡, Mike Fitzgerald‡, David B Jaffe‡, Xiaoping Yang‡, Amr Abouelleil*, Nicole R Allen*, Harindra Arachchi*, Jennifer Baldwin*, Toby Bloom*, Mark Borowsky*, Boris Boukhgalter*, Jon Butler*, Bruno Chazaro*, Mieke Citroen*, April Cook*, Ben Corum*, Kurt DeArellano*, Kathy Dooley*, Lester Dorris*, Matthew Endrizzi*, Jeffery Erickson*, Gary Gearin*, Sante Gnerre*, Andreas Gnirke*, Nabil Hafez*, Daniel Hagopian*, Jennifer Hall*, Catherine Hosage-Norman*, Charlien Jones*, Michael Kamal*, Asha

Kamat*, Chinnappa Kodira*, Patrick Kuharic*, Kurt LaButti*, Teresa Lai*, Jessica Lehoczky*, Rosie Levine*, Xiaohong Liu*, Tashi Lokyitsang*, Annie Lui*, Richard Mabbitt*, Pendexter Macdonald*, John Major*, Jonathan Manning*, Lisa Marinelli*, Charles D Matthews*, Evan Mauceli*, Atanas Mihalev*, Glen Munson*, Jerome Naylor*, Cindy Nguyen*, Paula O'Donnell*, Sinead O'Leary*, Keith O'Neill*, Stephen Parker*, Bruno Piqani*, Anthony Rachupka*, Umadevi Ramasami*, Christina Raymond*, Joseph Rodriguez*, Rebecca Schupbach*, Christopher Seaman*, Ted Sharpe*, Andrew Sheridan*, Cherylyn Smith*, Carrie Sougnez*, Sabrina Stone*, Matthew Stubbs*, Jessica Talamas*, Pema Tenzin*, Kerri Topham*, Sarah Towey*, Vijay Venkataranan*, Charles Whittaker*, Jane Wilkinson*, Andrew Zimmer*, Emmanuel Adekoya, Mostafa Ait-Zahra, Alla Ali, Thaddeus Allen, Mechele Anderson, Scott Anderson, Jeff Armbruster, Pasang Bachatsang, Tashi Bayul, Berta Blitshsteyn, Jason Blye, Leonid Boguslavskiy, Jody Camarata, Kevin Campo, Yama Cheshatsang, Al Collymore, Tony Considine, Patrick Cooke, Robert David, Tenzin Dawoe, Stewart DeGray, Chemey Dhongsar, Passang Dorje, Kunsang Dorjee, Noah Duffy, Alan Dupes, Abderrahim Farina, Susan Faro, Pat Ferriera, Heather Fischer, Sheila Fisher, Karen Foley, Diane Gage, Steph Gardyna, Seth Gordon, Audra Goyette, Joe Graham, Edward Grandbois, Kunsang Gyaltsen, Birhane Hagos, Beah Hatcher, Andrew Heaford, Andrew Heller, Tracy Honan, Nathan Houde, Lee Hughes, Bill Hulme, Ilian Iliev, Cristyn Kells, Alix Kieu, Peter Kisner, Dawa Lama, Tom Landers, Jean Pierre Leger, Doreen Lewis, Tammy Lewis, Yeshi Lokyitsang, Christine MacLean, Dick Marabella, Kebede Maru, Megan McCarthy, Tina McGhee, James Meldrim, Louis Meneus, Tanya Mihova, Val Mlenga, Leon Mulrain, Christian Newes, Na Nguyen, Thu Nguyen, Chou Dolma Norbu,

Nyima Norbu, Ose Okoawo, Jennifer O'Loughlin, Bamidele Omotosho, Sahal Osman, Tenzin Phulchung, Pema Phunkang, Nadia Pierre, Rayale Rameau, Vernada Ray, Cecil Rise, Peter Rogov, Steven Seaman, Ngawang Sherpa, Brian Spencer, John Stalker, Nicole Stange-Thomann, Sharon Stavropoulos, Kevin Stetson, Casey Stone, Pierre Tchuinga, Senait Tesfaye, Joumathe Theodore, Yama Thoulutsang, Tsamla Tsamla, Nawang Tsomo, Duane Valle, Helen Vassiliev, Rose Veil, Andy Vo, Tsering Wangchuck, Tsering Wangdi, Dudley Wyman, Shailendra Yadav, Shane Yeager, Rahel Retta Yeshitela, Geneva Young, Joanne Zainoun, Lisa Zembeck

4. US Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, California 94598 USA; comprising affiliations to the following institutes: *Los Alamos National Laboratory (LANL); P.O. Box 1663 Los Alamos, NM 87545; §Lawrence Berkeley National Laboratory (LBNL) One Cyclotron Road Berkeley, CA. 94720; ‡Lawrence Livermore National Laboratory (LLNL) 7000 East Ave. Livermore, CA 94550-9234; †Stanford Human Genome Center, Stanford University School of Medicine, Palo Alto, California 94305 USA

Anne Abrajano§, Catherine Adam‡, Aaron Adamson‡, Andrea Aerts§, Arun Aggarwal§, Dana Alcivare§, Michelle Alegria‡, Michelle Allegria-hartmann‡, Jennifer Alleman‡, Susan Allen‡, Michael R. Altherr *, Chris Amemiya‡, Gina Amico-keller‡, Janice Andora‡, Carol Andredesz‡, Tim Andriese‡, Sinoula Apostolou *, Terisita Arcaina§, Marlon Arcaina§, Andre Arellano‡, Perry Arellano-Jones§, Linda K. Ashworth‡, Cara Aslanidis‡, Lawreen Asuncion‡, Cristina Attix‡, Aaron Avila‡, Julie Avila‡, Phil Bach§, Hummy Badri‡, Eva Bajorek†, M. Baker§, Richard Baker§, Dan

Baker‡, Michele Bakis§, Michael Banda§, Suba Basu‡, Mark Batzer‡, Jason Baumohl§, Keith Beall§, Leslie Beaucham *, George I. Bell *, J Benke§, William Benner§, Douda Bensasson§, John Bercovitz§, Anne Bergmann‡, Tony J. Beugelsdijk *, Rita Bhakta‡, Aleksandr Bituin§, Stacey Black†, Robert Blazej§, Heather Blumer *, John Boehm§, Juanan Boen‡, Marnel Bondoc§, Eric Bowen§, Brigitte F. Brandriff‡, Wade Brannon§, Elbert Branscomb‡, Thomas Brettin *, Michael Bridgers *, Peter Brokstein§, Nancy C. Brown *, Robert Bruce‡, David C. Bruce *, William J. Bruno *, Jennifer E. Bryant *, Judith M. Buckingham *, Kerem Bulbul§, Nathan Bunker‡, Matt Burgin‡, Karolyn Burkhart-Schultz‡, Paul Butler‡, Sean Caenepeel§, David F. Callen *, Connie S. Campbell *, Evelyn W. Campbell *, Mary l. Campbell *, Mary L. Campbell *, Chenier Caoile†, Lolo Cardenas§, Faviola Cardenas§, Anthony V. Carrano‡, Jason Carriere†, Mario Cepeda§, Patrick Chain‡, Jean F. Challacombe *, Yee Man Chan†, Jarrod Chapman§, Jeffrey Chapple‡, Leslie a. Chasteen *, Simeon Chavarria§, Jin Chen *, Ray Chen§, Chira Chen‡, Jan-Fang Cheng§, Olga Chertkov *, Han C. Chi *, Sylvia Chin‡, Corey Chinn§, Sally Chiu§, Mari Christensen‡, A Chung§, Michael J. Cinkowsky *, Lynn Clark‡, Lynn M. Clark *, David Cleveland§, Jackie Cofield‡, Judith D. Cohn *, Rick Colayco‡, Jessie Combs‡, Robin Comstock§, Karen Connolly§, Alex Copeland‡, Rebecca Cordray§, Earl Cornell§, Sara Cotton§, Olivier Couronne§, Terrance Critchlow‡, Paul Critz§, Stephanie Cummings§, Eileen Dalin§, Linda Danganan‡, Christopher Daum‡, Stuart Davidson§, Cheryl Davis§, Larry l. Deaven *, Larry L. Deaven *, Kerry Deere‡, David Degusta§, Maria Deguzman‡, Paramvir Dehal‡, Pieter De Jong‡, Mirian Denys†, John C. Detter‡, Laurie Devlin‡, Jennifer Dias‡, Victoria Dias‡, Genevieve

Dibartolo‡, Mark Dickson†, Jeanne Dietz-Band *, Richard Digennaro§, Karen Dilts *, Mira Dimitrijevic-Bussod *, Kami Dixon†, Long Do‡, Norman A. Doggett *, Kimberly Dong§, Victor Dorsett§, Suzanne Duarte‡, Inna Dubchak§, Erin Dunwell§, Evan Eichler‡, Chris Elkin‡, Jeffrey Elliott‡, Ger van den Engh‡, David Engle§, Anne Marie Erler‡, Julio Escobar†, Gerald Eveleth‡, Joseph J. Fawcett *, Alicia Ferguson§, Anne Fertitta‡, Jim Fey‡, Patrick Fidel§, Elizabeth Fields‡, Marie Fink§, Laurice Fischer†, Pat Fitch‡, J. Flanagan§, Dave Flowers†, Peg Folta‡, Jonathan Fong§, Amanda F. Ford *, Dea Fotopulos†, Matthew Fourcade§, Jessica Fox†, Ken Frankel§, Marvin Frazier, Jane Fridlyand§, Sharin Fuller‡, Celsa Gallegos *, Michael Galvez§, Stuart Gammon‡, Emilio Garcia‡, Carmen Garcia†, Antonio Garcia-Martinez§, Alison Gardner *, Consuelo Garduno *, Jeffrey A. Garnes‡, Joseph M. Gatewood *, Binyam Gebreyesus§, Cynthia Gelein‡, Maartin Gelpke§, Anca Georgescu‡, Amy Geotina§, Mary Gifford§, Isaias Gil§, Paul Gilna *, Jeff Gingrich‡, Kira Giovanelli *, Tijana Glavina‡, Darren Gold†, Kristen Golinveaux‡, Maria Gomez†, Eidelyn Gonzales†, David Goodstein§, Lynne A. Goodwin *, Laurie A. Gordon‡, Catherine Gordon‡, Ann Gorvad‡, Christine Gould‡, Deborah I. Grady *, Jennifer Grant§, Bruce Gray§, Jeffrey K. Griffith *, Igor Grigoriev§, Jane Grimwood†, Matthew Groza‡, Hannibal Guarin§, Metzalli Guel§, Kate Gunning§, Jane Guo§, Chi Ha‡, Allen Haim§, Patrick Hajek§, Matt Hamilton§, Nancy Hammon§, Sha Hammond‡, Khalid Hamza§, Shunsheng Han *, Trevor Hawkins§, Lauren Haydu†, Keven Helfenbein§, Uffe Hellsten§, Nina Henderson§, David Hendrix§, Karla Henning‡, Victor Hepa‡, Charles Herman§, Carl E. Hildebrand *, Issac Ho§, Susan M.G. Hoffman‡, Ann Holtz‡, Caleb Holtzer†, Roya Hosseini§, Katherine Huang§,

Wayne Huang‡, Zhengping Huang‡, Heather Hue *, Hillary Hughes-Hull‡, David Humphries§, Rov Hunter II§, Matt Hupman§, Sandy Huynh‡, Paula Imbro‡, Nicole Inacio§, Drew Ingram‡, Sanjay Israni‡, Mark Jacintho§, Terri Jackson§, Joseph Jaklevic§, Michael Jaklevic§, Jamie Jett§, Phillip E. Jewett *, Jian Jin§, Lori Johnson‡, Stephanie Johnson‡, Genevieve Johnson‡, Myrna D. Jones *, Micheil Jones§, Yun Jordan§, Eugine Jung§, Kristen Kadner§, Orsalem Kahsai§, Orsalem Kahsai§, Pat Kale‡, Hitesh Kapur‡, Gifford Keen *, Laura Kegelmeyer‡, Lisa Kegg‡, David Keys§, Arsi Khan‡, Susu Khine‡, Jennifer Kim§, Joomyeong Kim‡, Heather Kimball§, William Kimmerly§, Emanuel H. Knill *, D Knorr‡, Arthur Kobayashi‡, William Kolbe§, Marie-Claude Krawczyk *, Gabriel Kremmidiotis *, Brent Kronmiller‡, Duane Kubischta§, Jennifer Kuehl§, Carol Kuhn‡, Chinnie Kwan‡, Amy Kyle‡, Anne Lacombe§, Melanie Lafrades§, Marissa Lam§, Edward Lam§, Jane E. Lamerdin‡, Christine Lansang‡, Victoria Lao‡, Alla Lapidus§, Bernadette Lato‡, Lisa Lee§, Peter Lee§, Joon ho Lee§, Denise Lee‡, Byung-in Lee‡, Karl Lehman§, Gregory Lennon‡, Kendra Lesar *, Keith Lewis§, Tina Leyba *, Kim Lieuallen‡, Ken Lindo§, K. Lindquist§, Kirsten Lindstrom§, Albert Linkowski§, Jing Mei Liu *, Stephanie Liu‡, Crystal Llewellyn-Silva§, Rebecca Lobb *, Jonathan l. Longmire *, Jonathan L. Longmire *, Frederick Lopez†, Jose Lopez†, Yunian Lou§, Michael G. Lowenstein *, Stephen Lowry§, Robert Lucas§, Susan Lucas‡, Thom Ludeman *, Lyndsey Lundgren§, Daniel Lynch§, Gary Ma§, Migdad Machrus§, Madison Macht‡, Catherine A. Macken *, Ramki Madabhushi‡, Magner‡, Venus Mahmoodi§, Ryan Mahnke‡, Stephanie Malfatti‡, Mary l. Maltbie *, Chitra Manohar‡, Marisa Mariano‡, Graham A. Mark *, Thomas G. Marr *, Christian Marroquin§, Christopher Martin§,

Joel Martin§, Michele Martinez§, Diego Martinez‡, Hadi Masood‡, Cynthia Maunes‡, Mary Kay McCormick *, Paula McCready‡, Phil McGurn‡, Kimberly McMurray *, Kimberly L McMurray *, Jenny McNinch‡, Catherine Medina†, Linda J. Meincke *, Linda J. Meincke *, George Mercado‡, Carl Mieczkowski§, Maggie Mier *, Trini Miguel§, Danielle Mihalkanin‡, Christie Miller§, Donald Miller‡, Sheri Miner‡, Monica Misra *, Sophia Mitina§, Harvey Mohrenweiser‡, Joseph Monforte§, Mishelle Montgomery‡, Jenna Morgan§, Jerry Morton‡, Robert K. Moyzis *, Shibani Mukherjee§, Mark O. Mundt *, A. Christine Munk *, Christie A. Munk *, Michael Murphy§, Christine Murray‡, Richard Myers†, Richard Nandkeshwar‡, Cleo M. Naranjo *, Deborah A. Nelson *, Catherine Nelson§, Ward Nelson‡, David O. Nelson‡, Kathryn Nelson§, Jennifer Neunkirch§, April Newman‡, Terrence Ng§, Le-Thu Dinh Nguyen§, Hoa Nguyen§, Lisa Nguyen§, Amber Nivens§, Matt Nolan‡, Matt Nolan‡, Larry Nowlin§, Pierre Oddone§, Jason Olivas‡, Anne S. Olsen‡, Ivan Ovcharenko§, David Ow‡, Paola Pace§, Elizabeth Pack *, Anuradha Padki§, Krishnaveni Palaniappan§, Michael Palazzolo§, Beverly Parson-Quintana *, Shripa Patel§, Robert M. Pecherer *, Vivian Peng§, Yi Peng§, Ze Peng§, Len A. Pennacchio§, Rene Perrier§, Brad Pesavento‡, Karl Petermann§, Linda Peters§, Ellen T. Peterson *, Erin Peterson§, Angela Peterson‡, Joyce Pfeiffer§, Hoan Phan‡, C. Piscia§, Maurice Pitesky‡, Sam Pitluck§, Ingrid Plajzer-Frick§, Alex Poliakov§, Martin Pollard§, Jos Polman§, Aaron Porter§, Martha Posada§, Pat Poundstone‡, Jason Powell *, Daisy Prado‡, Christa Prange‡, Paul Predki§, Jennifer Primus‡, Lyle Probst‡, Nicholas Putnam§, Osama Qasem§, Glenda Quan‡, Mohan Rajagopal§, Melissa Ramirez‡, Lucia Ramirez†, Nicole Ramos‡, David Randolph§, Irma Rapier§,

Barry Rappaport *, Sam Rash‡, Robert l. Ratliff *, Aditya Ray§, Warren Regala‡, Charlie Reiter§, James Retterer†, Paul Richardson§, Kristi Richardson‡, Paul Richardson§, Darryl o. Ricke *, B.E. Riley§, Angel Rivera‡, Simon Roberts§, Donna l. Robinson *, Dezere Robinson§, Chris Robinson‡, Donna L. Robinson *, Juan Rodriguez§, Alex Rodriguez†, Stephanine Rogers†, Daniel Rokhsar§, Stephanie Ross‡, Edward M. Rubin§, Jeffrey Rule§, Andrea Ryan§, George Sakaldasis‡, Azaf Salamov§, Angelica Salazar†, Brent Sanders§, Christina Sanders‡, R. Sarmiento§, Suraj Satpathy§, Elizabeth H. Saunders *, William W. Sawhill *, Eric Saxman *, Jeremy Schmutz†, Duncan Scott§, Paul Scott§, Damian Scott‡, Burt Seymour‡, Saima Shams‡, Harris Shapiro§, Maria Shin§, Timothy Shought *, Jeff Shreve§, Linda Sindelar§, Evan Skowronski‡, Nathaniel Slater§, Tom Slezak‡, Karen Smith§, Doug Smith§, Joel Smith‡, Troy Smith‡, Cari Soderlund *, Corey Spainhower§, Raymond l. Stallings *, Malinda Stalvey *, Justin T. Stege *, Shaela Stephens-Bolds§, Victor Stevko§, Evan Stone§, Jonathan Strasser§, Lisa Stubbs‡, Janet Stultz§, Sandhya Subramanian§, Leslie J. Summers‡, Robert D. Sutherland *, Grant R. Sutherland *, Kristina Tacey§, Beckie Tagett *, Tori Takaoka§, Tracy Takenaka§, Serena Tam§, Olivia Tan†, Roxanne Tapia *, Owatha l. Tatum *, Astrid Terry‡, Judith G. Tesmer *, Nina Thayer *, James Theil§, Paulette Thomas‡, Linda S. Thompson *, Hope Tice‡, Damon Tighe§, David C. Torney *, Mary Tran-Gyamfi‡, Margie Trankiem‡, Barbara J. Trask‡, Marianne Treuhaft‡, Stephan Trong‡, Ming Tsai†, Susan Tsujimoto‡, Heidi Turner‡, Katherine Tynan‡, Stephanie Ueland‡, Samantha Ueng *, Levy E. Ulanovsky *, Chun Un§, Quyen Ung§, Anna Ustaszewska‡, Michelle Valdez *, Yvonne Valles§, Marvin Van Dilla‡, Ryan Van

Luchene§, Zachary Vardanian§, Michelle Vargas‡, Steffan Vartanian†, Eugene Veklerov§, Nelson Velasco‡, Lisa Velasquez‡, Lisa Vergez‡, Vijay Viswanathan‡, Nu Vo†, Mark Wagner‡, Mark Wagner‡, Keya Walker§, Mei Wang§, Joann Wang§, Edward Wehri‡, Richard Weidenbach§, Li Weng§, Sarah Wenning‡, Sara Wentz‡, Jeremy Wheeler†, P. Scott White *, Catherine White§, Jennifer White‡, Scott P. White *, Scott Whitmore *, Sarah Wiegel†, Albert l. Williams *, Albert L. Williams *, Patricia l. Wills *, Richard Wills‡, Patricia L. Wills *, Steve Wilson§, Jessica Wollard‡, Andrew W. Womack *, Benjamin Wong‡, Diane Wood *, Eric Woods§, Tanja Woyke§, Kristin Wright‡, Jung-Rung Wu *, Kevin Wu†, Melissa Wycoff-Montenegro§, Stacia Wyman§, Gary Xie *, Alex Xu§, Lynn Yamamoto§, Joan Yang†, Candace Yano§, Mimi Yeh‡, Kathy Yokobata‡, Jennifer Yokoyama§, Charles Yu§, Brian Yumae§, Rui Yun§, Qing Zhang§, Carol Zhou‡, Deborah Zierten†

5. Baylor College Of Medicine Human Genome Sequencing Center, Baylor College Of Medicine, One Baylor Plaza, Houston Texas 77030 USA

Charles Q. Adams, Babajide Adio-Oduola, Carlana C. Allen, Heather Allen, Stephen L. Alsbrooks, Harshinie C. Amaratunge, Amita G. Amin, Vivian Anyalebechi, Heather Hope Arredondo, Chinwe O. Asomugha, Shannon Michelle Austin, Mulu Ayele, Dilrukshi P. Bandaranaike, Pertrice Baptiste, Joseph A. Barbaria, Arthur Beaudet, Terell M. Bellard, Michael Anthony Benard, Venita Bhuchar, Kesha E. Bimage, Kaneza Biswalo, Jessica Anne Blair, Tia Blanks, David B. Blomstrom, Peter R. Blyth, David Bonnin, Angela K. Brimhall, Mary J. Brown, Michael Brown, Nathaniel P. Bryant, Christian J. Buhay, Paula Ellen Burch, Carrie E. Burkett, Kevin L. Burrell, Eliana Calderon, Veronica Cardenas,

Kelvin Carter, Kristal Casias, Iracema Cavazos, Sandra R. Cavazos, Heather Ceasar, Joseph Chacko, Sheryl N. Chan, Dean Chavez, Guan Chen, Mike Z. Chen, Rui Chen, Zhijian Chen, Constantine Christopoulos, Joseph Chu, Kerstin Petra Clerc Blankenburg, Raynard Cockrell, Caroline D. Cox, Marcus D. Coyle, Amanda L. Crawford, Andrew Cree, Michelle Dang, Stephanie R. Dathorne, Mary Louise Davila, Candi Mon'et Davis, Clay Davis, Latarsha Davy-Carroll, Claudia De Anda, Andrea De Latorre, Oliver Delgado, Shawn Denson, Christine Deramo, Yan Ding, Huyen H. Dinh, Jeremy E. Donlin, Karen J. Douthwaite, Heather Draper, Lisa Marie D'Souza, Shannon Dugan-Rocha, Adam M. Dunn, Kenneth J. Durbin, Darius Dziuda, Christopher D. Earnhardt, Kristy A. Eaves, Darren Edgar, Susan Omwanghe Osayanro Edionwe, Amy Egan, Natalia Emelianenko, Alexandra J. Emery-Cohen, Michael Escotto, Craig S. Eugene, Sonia Fernandez, Pushpa Ranjani Fernando, Maria C. Finley, Nicole Flagg, Lisa D. Forbes, Myre J. Foster, Abdul M. Gabisi, Marvin Garcia, Ricardo M Garcia III, Toni T. Garner, Melissa Garza, Laura Gay, Stephanie L Gench, Sayantan Ghose, Richard A. Gibbs, Rachel Gill, Guillermo B. Gonzalez, Michelle L. Grady, William Guerra, Whitney V. Guevara, Preethi Gunaratne, Brett Lynn Hall, Brittany Hall, Cerissa Hamilton, Keelan A. Hamilton, Vincent A. Hanak, Brandy A. Harbes, Paul Havlak, Alicia C. Hawes, Ebere Sylvia Onyirioha Hawkins, Nicholas Henderson, Judith Hernandez, Omar Hernandez, Allyson Herseim, Sandra Hines, Matthew Edward Hitchens, Anne V. Hodgson, Marilyn E. Hogues, Mara Hollaway, Barbara Hollins, Deana K. Homsi, Farah J. Homsi, Amber Hornsby, Stephanie R. Howard, Latoya Lynette Howell, Sally Lucinda Howells, Mei Huang, James W. Huber, Steven Hulyk, Jennifer Hume, Michael Hunt, Devincent G. Idlebird, Kelvin C. Imo, Laronda R. Jackson, Leni Susan Jacob, Alberto Jimenez,

Bennie Johnson, Rudy Johnson, Angela Jolivet, Ryan Kagan, Shayna P. Kelly, Susan H. Kelly, Umer F. Khan, Ziad Mohid Khan, Divya Khurana, Laquisha Monique King, Haika Kisamo, Christie L. Kovar, Larissa Kulish, Belita Leal, Heather Lebow, Cherie Jenae Lee, Kenneth Lee, Fitzherbert Henderson Legall III, Jaclyn M. Levan, Ernest K. Lewis, Lakeshia C. Lewis, Lora R. Lewis, Jane Li, Zhangwan Li, Jing Liu, Jing Liu, Wen Liu, Yih-Shin Liu, Dhammika Liyanage, Pamela London, John Lopez, Lora Lopez, Lorna M. Lorensuhewa, Hermela Loulseged, Demetria A. Lovett, Ryan J. Lozado, Xiuhua Lu, Yue Lu, Alice Lucier, Raymond L Lucier, Jie Ma, Xiaohui Ma, Renita C. Madu, Manjula Maheshwari, Mathew Mahindartne, Kevin Malloy, Brian Mangum, Patricia Mapua, David Mareth, Kirt Martin, Ashley D. Martindale, Evangelina Martinez, Elizabeth Massey, Sterling D. Matthews, Samantha Mawhiney, Michael P. McLeod, John D. McPherson, Michael G. Meador, Sylvia Mendez, Christian Mercado, Iracema C. Mercado, Christina E. Merritt, Michael Lee Metzker, Mason W. Mileur, George R. Miner, Emmanuel Minja, Teresa Mitchell, Farida Mohabbat, Khatera Mohabbat, Julie Montemayor, Baize Montgomery, Niki Moore, Margaret B. Morgan, Sidney Morris, Mala Munidasa, Matthew B. Murphy, Debra Dianne Murray, Donna Marie Muzny, Latha Nair, Seema R. Nair, Amit N. Nanavati, Colin E. Nankervis, Lynne V. Nazareth, Dearl D. Neal, Robin Ngoc Ngo, Bao-Viet Nguyen, Natalie M. Nguyen, Ngoc Bich Nguyen, Elizabeth Nickerson, Sarah Louise Norris, Ogechi O. Nwaokelemeh, Stanley Nwokenkwo, Melissa Obregon, Maryann Oguh, Geoffrey O. Okwuonu, Kenneth C. Okwuonu, Alisa Olarnpunsagoon, Njideka Oragunye, Jerry Orantes, Rodolfo J. Oviedo, Araceli C. Pace, Shane Pahlavan, Vladimir Panasenko, David Neil Parker, Kenya L. Parks, Shiran Pasternak, Bella Mayurkumar Patel, Jalpa N. Patel, Vishal V. Patel, James Patrick, Heidie A. Paul, Brett

A. Payton, Agapito Perez, Lesette M. Perez, William Perrin, Leonard G. Peters, Adam Pickens, Natasha Pieper, Farah Jo Homsi Plopper, Alan Poindexter, Nima M. Pourrajabi, Eltrick L. Primus, Ling-Ling Pu, Maria Puazo, Xiang Qin, Miyo M. Quiles, Juana B. Quiroz, Dina Rabata, Eric K. Rachlin, Kacy Reeves, Richard A. Reigh, Yanru Ren, Stephen Richards, Catharine M. Rives, Travis L. Rodkey, Alberto Rojas, Erica M. Ruiz, San Juana Ruiz, Moazzam Mohammad Sana, Wendi Sanders, Jireh Santibanez, Glenford G. Savery, Steven Edward Scherer, Matthew Schnabl, Graham Scott, Hua Shen, Ida Sisson, Erica Sodergren, Titilola Sonaike, Xing-Zhi Song, Richard P. Sorelle, Anastasia L. Sparks, David Steffen, Angelica E. Sutton, Amanda F. Svatek, Jessica Svetz, Leah Anne Svetz, Paul E. Tabor, Kavitha S. Tamerisa, Aparna Tamirisa, Hongli Tang, Jennifer Tansey, Christina C. Taylor, Tineace R. Taylor, Ethan Brett Telford, Nicole Thomas, Shereen Thomas, Rachel Diane Thorn, Ly Thanh Tran, Zulma Y. Trejos, Brenda Kaye Trevino, Ogechi Nkemdirim Ukegbu, Jeremy B. Urban, Kamran Usmani, Ruben E. Valas, Charles E. Vargo, Lydia I. Vasquez, Virginia A. Vera, Daniel Verduzco, Donna M. Villasana, Lenee Waldron, Davian L. Walker, Randy E. Wall, Natalie Walsham, Jie Wang, Ling Wang, Qiaoyan Wang, Suzhen Wang, Stephanie Ward-Moore, James T. Warren, Xuehong Wei, George Weinstock, Mary Melissa Wesley, David A. Wheeler, Flower White, Manosha I. Wickremasinghe, Gabrielle A. Williams, Angela L. Williamson, Regina Wleczyk, Hailey S. Wooden, Steven H. Wooden, Kim Carlyle Worley, Rita A. Wright, Jiaqian Wu, Syed Yakub, Jennifer Yen, Lillienne Yoon, Vivienne Yoon, Fuli Yu, Ye Yuan, Jingkun Zhang, Shaojie Zhang, Zhengdong Zhang, Jianling Zhou, Sara E. Zorrilla

6. RIKEN Genomic Sciences Center, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanegawa, 230-0045, Japan (*Technical Staff; ‡Technical Support)

Yoshiyuki Sakaki, Masahira Hattori, Asao Fujiyama, Todd Duane Taylor, Atsushi Toyoda, Yasushi Totoki, Hidemi Watanabe, Tetsushi Yada, Yoko Kuroki, Hideki Noguchi, Katsuhiko Murakami, Tomoyuki Aizu*, Rie Arai*, Fumio Ejima*, Tomoko Hasegawa*, Etsuko Hashimoto*, Wakako Hashimoto*, Naomi Inagaki*, Shuuji Inoue*, Hinako Ishizaki*, Emi Isozaki*, Noriko Ito*, Mariko Iwatate*, Mikiko Iwatsu*, Kaoru Kaida*, Takako Kato*, Moe Kimura*, Miho Kiyooka*, Kayo Maruyama*, Ayuko Motoyama*, Hiroko Morita*, Maho Naka*, Saori Nakagawa*, Aki Nishida*, Etsuko Ogawara*, Reina Okumura*, Nao Ohta*, Ritsuko Ozawa*, Ryoko Sakai*, Noriko Shiino*, Minami Shoji*, Keiko Takahashi*, Nobuhiro Takano*, Yoriko Terada*, Miwako Tochigi*, Iori Tomizawa*, Yumi Tsukamoto*, Rina Tsuzuki*, Nozomi Uyama*, Noriko Yamamoto*, Yuko Yamamoto*, Yasue Yamashita*, Miho Yonezawa*, Satoru Yoshida*, Hiromi Wada*, Michiru Fujioka‡, Rintaro Fukawa‡, Takehiko Itoh‡, Takujiro Katayama‡, Chiharu Kawagoe‡, Syunsuke Nagao‡, Nobuhiro Oomori‡, Kenshiro Ooshima‡, Hidetsugu Shimizu‡, Jiu Qin Sun‡, Takashi Tahara‡

7. University of Washington Genome Center, Fluke Hall on Mason Road, Box 352145 Seattle WA 98195 USA

Rajinder Kaul, Maynard V. Olson, Will Gillet, Eric Haugen, Chris Raymond, Yang Zhou, Anthony Palmeiri, Channakhone Saenphimmachak, Sandhya Subramanian, Zaining Wu, Karen Phelps

8. Genoscope and CNRS UMR-8030, 2 Rue Gaston Cremieux, CP 5706, 91057 Evry Cedex, France

Roland Heilig, Ralph Eckenberg, Jean-Louis Petit, Núria Fonknechten, Corinne Da Silva, Laurence Cattolico, Michaël Levy, Valérie Barbe, Véronique de Berardinis, Abel Ureta-Vidal, Eric Pelletier, Virginie Vico, Véronique Anthouard, François Artiguenave, Catherine Robert, Corinne Cruaud, Thomas Brüls, Philippe Brottier, Béatrice Ségurens, Sylvie Samain, Hervé Crespeau, Claude Scarpelli, Gábor Gyapay, Patrick Wincker, William Saurin, Jean Weissenbach, Fumihiko Matsuda

9. Department of Molecular Biology, Keio University School of Medicine 35 Shinanomachi Shinjuku-ku, Tokyo 160-8582, Japan

Nobuyoshi Shimizu, Shinsei Minoshima, Jun Kudoh, Shuichi Asakawa, Kazuhiko Kawasaki, Kazunori Shibuya, Takashi Sasaki, Atsushi Takayanagi, Atsushi Shimizu, Ai Shintani, Eriko Nakato, Susumu Mitsuyama, Satoru Yamazaki, Izumi Obayashi, James Schmeits, Masafumi Ohtsubo, Michiyo Okui, Yuji Murayama, Meng K. Lim, Saho Ohno, Kentaro Nagamine, Yimin Wang, Jun Wang, Aiko Shiohama, Norie Aoki, Nobuaki Shindoh, Hideto Maeda, Katsuhiro Hosono, Hao Yang, Tomohiro Izumiyama, Miho Tatsuyama, Nori Sawada, Yoshitaka Yamaguchi, Masataka Suzuki, Yosuke Ohsawa, Michi Asahina, Yuko Matsushita, Sabine K. Ishikawa, Mio Mejima, Tamami Adachi, Teruyo Asakawa, Mie Furuhashi, Misuzu Osada, Kyoko Ohmori, Etsuko Nakamura, Naoko Sanuki, Masayo Nakagawa, Naoko Togashi, Junko Hamada, Meishan Shen, Mai Kobayashi, Junko Onoda, Aki Mitsui, Saki Watanabe, Akemi Miyazawa, Shinobu

Iwakawa, Akiko Ishibashi, Kei Hosoda, Shuko Matsumoto, Kazuyasu Kojima, Hiromi Tsuruoka, Yali Xuan, Fumiaki Ito*, Yoshiko Shimizu§, Stylianos E. Antonarakis‡

*Setsunan University, Hirakata, Osaka, Japan

§Kyorin University School of Health Sciences, Hachioji, Tokyo, Japan

‡University of Geneva Medical School & University Hospitals, Switzerland

10. Genome Therapeutics Corporation, 100 Beaver St. Waltham, Massachusetts 02453 USA. Current address: Agencourt Bioscience Corp 100 Cummings Center, Beverly MA 01915 USA

John Battles, Patrick Cahill, David Camire, Lynn Doucette-Stamm, Kim Fechtel, Erik Gustafson, Jennifer Horne, Christopher Hynds, Arnold Kana, Hong-Mei Lee, Minh Nguyen, Andrea Rogosin, Marc Rubenfield, Douglas R. Smith, Lisa Standring, Bill Torcaso, Jason Tsolas, Jennifer Walsh, Ying Wang, Keith Weinstock

11. Institute for Systems Biology, 1441 N. 34th Street, Seattle WA 98103

Brian Birditt, Scott Bloom, Leroy Hood, Amardeep Kaur, Ryan Nesbitt, Lee Rowen

12. Department of Genome Analysis, Institute of Molecular Biotechnology, Beutenbergstrasse, 11, D-07745, Jena, Germany

Matthias Platzer, Cornelia Baumgart, Silke Förste, Petra Galgoczy, Gernot Glöckner, Yvonne Görlich, Ivonne Heinze, Niels Jahn, Dorothee Lagemann, Rüdiger Lehmann, Elke Meier, Evelyn Michaelis, Uta Petz, Kathrin Reichwald, Markus Schilhabel, Kathleen Seitz, Roman Siddiqui, Stefan Taudien, Gaiping Wen, Daniela Werler

13. Beijing Genomics Institute/James D. Watson Institute of Genome Sciences (

Beijing Genomics Institute, Chinese Academy of Sciences/James D. Watson Institute of Genome Sciences, China; 2 Institute of Human Genomics Aarhus University, Aarhus, Denmark; 3 Northern National Genome Center, Beijing, China; 4 Southern National Genome Center, Shanghai, China; 5 School of Medicine, Southeast University, Nanjing, China; 6 College of Life Sciences, Peking University, Beijing, China; 7 Center of Bio-X Life Sciences, University of Communication, Shanghai, China; 8 Department of Medical Genetics, University of Washington, Seattle, USA; 9 Institute of Biophysics, Chinese Academy of Sciences, Beijing, China; 10 Genome Sequence Center, BC Cancer Research Center, Vancouver, Canada; 11 Institute of Microbiology, Chinese Academy of Sciences, Beijing, China; 12 Medical college, Xi'an Jiaotong University, Xi'an, China; 13 Hebei Medical University, Shijiazhuang, China; 14 The High School Affiliated to Renmin University Of China, Beijing, China; 15 Wenzhou Medical College, Wenzhou, China)

Jingyue Bao1,15, Qiyu Bao1,15, Weidong Bao, Shihua Bi, Xuemeng Bian, Lars Bolund , , Tianjing Cai , Zhen Cai4, Ting Cao, Yuzhu Cao, Baoxian Chen, Chong Chen, Cong Chen, Jianlong Chen, Jie Chen , Junbao Chen, Lixia Chen3, Shu Chen14, Tong Chen4, , Xiaoyun Chen4, Xinyi Chen14, Yanjiong Chen1,12, Yiyu Chen, Zhu Chen7, Zhihua Cheng7, Cheng Chi14, Hongjuan Cui, Jinhui Cui, Peng Cui1,13, Bin Cong13, Li Dai, Yajun Deng1,12, Hao Ding, Hui Dong7, Wei Dong, Xiaojia Dong3, Yutao Du1,13, Hongyuan Fan, Keke Fan14, Jianqiu Fang, Guodong Feng, Haiyan Feng, Jie Feng, Xiaoli Feng, Gang Fu4, Qiushi Fu14, Anwu Gao, Jimei Gao, Quan Gao3, Yang Gao,

Yimei Gao, Jianing Geng1,13, Guanghui Gong, Jinying Gong3, Jingli Gu4, Jun Gu3, Wenyi Gu4, Xiaocheng Gu , Qiaoning Guan, Qi Gui, Daorong Guo, Lili Guo13, Guiqing Han, Xurong Han, Wei Huang4, Fengying He3, Jiaying He, Lin He , Jie Hu, Songnian Hu, Fang Huang, Guyang Huang4,7, Xiaodi Hou14, Jia Jia4, Nan Jia, Xingchang Jia, Lu Jiang, Yanrui Jiang3, Ziyou Jiang, Yetao Jin, Yongsan Jin, Hui Kang4, Ning Kang3,, Mary-Clare King4, , Yi Kong, Zhaohui Kou3, Jianghua Lai1,12, Meng Lei, Changfeng Li, Chenji Li, Eryao Li3, Feifei Li, Gang Li, Jiayang Li, Jihong Li, Jing Li, Jingfei Li, Jingxiang Li, Li Li, Lili Li, Lin Li, Ming Li7, Nan Li, Ran Li, Shengbin Li12, Shuping Li, Shuxia Li, Shuangding Li12, Shuangli Li1,13, Songgang Li, Tao Li1,12, Wei Li , Wenjie Li, Yan Li, Yanni Li, Yi Li, Zhijie Li, Jinsong Liao, Wei Lin, Wei Ling7, Boyong Liu, Cong Liu14, Dianwu Liu13, Haili Liu, Jing Liu, Jun Liu, Kai Liu, Ning Liu5, Siqi Liu, Wei Liu3, Xinshe Liu12, Yanhua Liu, Yannan Liu, Yilin Liu, Ying Liu, Yu Liu, Yu Liu, Zhanwei Liu, Lingfeng Lu4, Tao Lu3, Yongxiang Lu, Chunqing Luo, Gang Lv9, Cheng Ma, Jiao Ma, Qingmei Ma3, Shanshan Meng, Feng Mu, Lin Ni4, Yuxin Niu, Jiaofeng Pan, Sisu Pan14, Qiuhui Qi, Xiaohua Qi, Xufang Qian7, Zengmin Qian, Boqin Qiang6, Zhenyong Qiao7, Shuangxi Ren7, Li Rong6, Chunyan Shao4, Yufen Shao, Fengye Shen7, Yan Shen6, Guibin Shi3, Hongfang Shi1,13, Michael Smith4, , Liping Song, Shuping Song, Xiaohu Su14, Jiajia Sun, Jiandong Sun, Min Sun, Tao Sun, Wei Sun, Yongqiao Sun1,13, Yu Sun, Yue Sun, Xinyu Tan6, Xiangjun Tang, Ran Tao, Chaoguang Tian, Wei Tian, Yan Tian, Yong Tian3, Yuqing Tian5, Jingli Tong, Yuefeng Tu4, Ma Wan7, Dong Wang, Feng Wang, Guangxin Wang, Guihai Wang, Hongjuan Wang7, Hongwei Wang6, Huanjie Wang14, Huifeng Wang, Jian Wang, Jingqiang Wang, Juan Wang, Jun Wang, Li Wang, Lijie Wang, Lijuan Wang, Lijun Wang4, Liqun Wang7, Shifen Wang,

Shu Wang3,Wenjun Wang, Xiaolei Wang, Xiaoling Wang, Xiaoning Wang5, Xuefei Wang, Xuegang Wang, Yan Wang, Ying Wang, Yuanyuan Wang, Shulin Wei, Gane K-S Wong, Chungen Wu7, Dongying Wu, Qingfa Wu, Ruiming Wu, Xiao Wu14, Xiaojing Wu, Yan Xi, Yingying Xi3, Fei Xie, Yuqing Xiong, Jing Xu, Ruqin Xu, Shuhua Xu7, Wei Xu, Yan Xu14, Yuning Xu3, Zuyuan Xu, Zhenyu Xuan12, Rui Xue, Yali Xue, Chunxia Yan, Fei Yan8, Guangmei Yan4, , Huanming Yang, Rui Yang14, Shudong Yang, Xiaonan Yang, Yongkui Yang, Zhijian Yao6, Chen Ye, Haifeng, Yin4, Bo You, Bing Yu1,12, Jun Yu, Kaiwen Yuan, Yida Yuan, Shujie Yue, Yixin Zeng, Dong Zhai, Bo Zhang, Caimin Zhang, Fengmei Zhang, Guangyu Zhang, Guohua Zhang, Haiqing Zhang, Hongbo Zhang, Jingjing Zhang, Jiuchun Zhang, Lanzhi Zhang, Li Zhang, Meihua Zhang, Meng Zhang, Ming Zhang7, Ruhua Zhang, Tiezhen Zhang, Wei Zhang7, Xianglin Zhang4, Xiaoliang Zhang, Xiuqing Zhang, Wei Zhang, Yan Zhang3, Yi Zhang4, Yilin Zhang, Ying Zhang, Yuansen Zhang1,13, Yuzhi Zhang, Hongmei Zhao, Lijian Zhao1,13, Zhijing Zhao, Zhicheng Zhen6, Haibo Zheng1,12, Ming Zhong7, Haixia Zhou, Nannan Zhou, Xinfeng Zhou3, Yan Zhou7, Yi Zhou3, Bingying Zhu7, Bofeng Zhu1,12, Fenghua Zhu4, Genfeng Zhu4, Miao Zhu, Ning Zhu6, Yongge Zhu, Zhen Zhu, and Shitao Zhuang3

Finishing Group:

Weidong Bao, Bin Chen4, Chong Chen, Lihong Chen3, Zhu Chen4, Peng Cui, Keyue Ding3, Wei Dong, Gang Fu4, Wenyi Gu4, Wei Huang4, Jia Jia4, Wei Li, Boyong Liu, Chunqing Luo, Gang Lv4, Yuxin Niu, Xiaohua Qi, Shuangxi Ren4, Li Rong3, Yan Shen3, Haihui Sheng4, Liping Song, Xinyu Tan3, Chaoguang Tian, Yuefeng Tu4, Huifeng Wang, Jian Wang, Jing Wang6, Jingqiang Wang, Jinhua Wang3, Jun Wang,

Qingfa Wu, Hui Xiong4, Jing Xu, Huanming Yang, Zhijian Yao3, Xiaonan Yang, Jun Yu, Haiqing Zhang, Li Zhang, Xianglin Zhang4, Yilin Zhang, and Genfeng Zhu4

Administrative Group:

Zhu Chen, Gang Fu, Jun Gu, Wei Huang, Jiayang Li, Zaiping Li, Qian Liu, Boqin Qiang, Yan Shen, Guihai Wang, Jian Wang, Xinlai Xu, Zhihong Xu, Huanming Yang, Jun Yu, Guoping Zhan

14. Max Planck Institute for Molecular Genetics Ihnestrasse 73, 14195, Berlin, Germany

Ralf Sudbrak, Alfred Beck, Alia BenKahla, Katja Borzym, Donald Buczek, Verena Gimmel, Katja Heitmann, Steffen Hennig, Sven Klages, Anna Kosiura, Michael Kube, Ines Müller, Hans Lehrach, Silvia Schmoger, Marie-Laure Yaspo, Richard Reinhardt

15. Department of Genome Analysis, GBF – German Research Centre for Biotechnology. Mascheroder Weg 1, D-38124 Braunschweig, Germany

www.genome.gbf.de

Helmut Blöcker, Ansgar Conrad, Martin Czubayko, Klaus Hornischer, Doris Järke, Michael Jarek, Gerhard Kauer, Tschong-Hun Löhnert, Gabriele Nordsiek, Maren Scharfe, Oliver Schön, Stephanie Thies

16. Advanced Center for Genetic Technology, Applied Biosystems Division of Perkin-Elmer Corp., Foster City, CA 94404, USA

Ellson Chen (current address: Vita Genomics, Inc. 7FL, NO. 6, SEC. 1, JungShing Road, Wugu Shiang, Taipei 248, Taiwan), Primo Babayan, Chun-Nan Chen, Cheryl Heiner, Peter Ma, Massimo Zollo, Lin Zuo

Mapping Group (Washington University in St. Louis):

R. Nagaraja, David Schlessinger

17. Analysis group. Individuals listed below in alphabetical order performed analyses discussed in the manuscript, and/or were responsible for assembling and maintaining the finished sequence in the public database. Includes individuals listed under other headings. (1Department of Computer Science, University of California at Santa Cruz, Santa Cruz CA 95064 USA; 2National Center for Biotechnology Information National Institutes of Health Bldg 38A 8600 Rockville Pike, Bethesda MS 20894 USA; 3MRC Functional Genetics Unit, University of Oxford, Department of Human Anatomy and Genetics, South Parks Road, Oxford, OX1 3QX, UK; 4Department of Genetics, Case Western Reserve University, 2109 Adelbert Road Cleveland, OH 44106 USA; 5Wistar Institute, 36th & Spruce Streets, Philadelphia, PA 19104 USA)

Robert Baertsch 1, Ewan Birney 6, Hsiu-Chuan Chen 2, Hiram Clawson 1, Mark Diekhans 1, Evan Eichler 4, Xose Fernandez-Suarez 6, Terrence S. Furey 1, Leo Goodstadt 3, Martin Hammond 6, David Haussler 1, Angela S. Hinrichs 1, Fan Hsu 1, Philip Johnson 2, Donna Karolchik 1, Arek Kasprzyk 6, Damian Keefe 6, W. James Kent 1, Paul Kitts 2, Darin London 6, Yontao T. Lu 1, Donna Maglott 2, Graham McVicker 6, Craig Melsopp 6, Sheila Paul 5, Chris P. Ponting 3, Glenn Proctor 6, Kim Pruitt 2, Stephen Rice 3, Harold Riethman 5, Kate R. Rosenbloom 1, Krishna M. Roskin

1, Barbara Ruef 2, Esther Schmidt 6, Gregory Schuler 2, Michael Schuster 6, Jessica Severin 6, Xinwei She 4, Adam C. Siepel 1, Damian Smedley 6, Arne Stabenau 6, Charles W. Sugnet 1, Daryl J. Thomas 1, Heather Trumbower 1, Abel Ureta-Vidal 6, Imre Vastrik 6, Caleb Webber 3, Ryan J. Weber 1, Cara Woodwark 6

18. Scientific management National Human Genome Research Institute, National Institutes of Health, Bethesda MD 20892 USA
Francis S. Collins, Mark Guyer, Jane Peterson, Adam Felsenfeld, Kris Wetterstrand, Sandra Kamholz. 2004. Finishing the euchromatic sequence of the human genome. Nature 421:931-945.

Rhesus Macaque Genome Sequencing and Analysis Consortium, R. A. Gibbs, J. Rogers, M. G. Katze, R. Bumgarner, G. M. Weinstock, E. R. Mardis, K. A. Remington, R. L. Strausberg, J. C. Venter, R. K. Wilson, M. A. Batzer, C. D. Bustamante, E. E. Eichler, M. W. Hahn, R. C. Hardison, K. D. Makova, W. Miller, A. Milosavljevic, R. E. Palermo, A. Siepel, J. M. Sikela, T. Attaway, S. Bell, K. E. Bernard, C. J. Buhay, M. N. Chandrabose, M. Dao, C. Davis, K. D. Delehaunty, Y. Ding, H. H. Dinh, S. Dugan-Rocha, L. A. Fulton, R. A. Gabisi, T. T. Garner, J. Godfrey, A. C. Hawes, J. Hernandez, S. Hines, M. Holder, J. Hume, S. N. Jhangiani, V. Joshi, Z. M. Khan, E. F. Kirkness, A. Cree, R. G. Fowler, S. Lee, L. R. Lewis, Z. Li, Y.-s. Liu, S. M. Moore, D. Muzny, L. V. Nazareth, D. N. Ngo, G. O. Okwuonu, G. Pai, D. Parker, H. A. Paul, C. Pfannkoch, C. S.

Pohl, Y.-H. Rogers, S. J. Ruiz, A. Sabo, J. Santibanez, B. W. Schneider, S. M. Smith, E. Sodergren, A. F. Svatek, T. R. Utterback, S. Vattathil, W. Warren, C. S. White, A. T. Chinwalla, Y. Feng, A. L. Halpern, L. W. Hillier, X. Huang, P. Minx, J. O. Nelson, K. H. Pepin, X. Qin, G. G. Sutton, E. Venter, B. P. Walenz, J. W. Wallis, K. C. Worley, S.-P. Yang, S. M. Jones, M. A. Marra, M. Rocchi, J. E. Schein, R. Baertsch, L. Clarke, M. Csürös, J. Glasscock, R. A. Harris, P. Havlak, et al. 2007. Evolutionary and Biomedical Insights from the Rhesus Macaque Genome. Science 316:222-234.

Dagert, M., and S. D. Ehrlich. 1979. Prolonged incubation in calcium chloride improves the competence of Escherichia coli cells. Gene 6:23-38.

Ellman, G. L. 1959. Tissue sulfhydryl groups. Archives of Biochemistry and Biophysics 82:70-77.

Farrel, E. K., and Merkler, D.J. 2008. Biosynthesis, degradation and pharmacological importance of the fatty acid amides. Drug discovery today 13:558-568.

Goldenberg, D., Azar, I., Oppenheim, A. B., Brandi, A., Pon, C. L. and Gualerzi, C. O. 1997. Role of Escherichia coli csp promoter sequences and adaptation of translational apparatus in the cold shock response. Molecular and General Genetics MGG 256:282-290.

Gonzalez, F. J. 2005. Role of cytochromes P450 in chemical toxicity and oxidative stress: studies with CYP2E1. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 569:101-110.

Grant, D. M., Josephy, P. D., Lord, H.L. and Morrison, L.D. 1992. Salmonella typhimurium strains expressing human arylamine N-Acetyltransferases: metabolism and mutagenic activation of aromatic amines. Cancer Research 52:3961-3964.

Grøn, I., Gregersen, N., Kølvraa, S. and Rasmussen, K. 1978. Formation of N-dicarboxyl-mono-glycines catalyzed by glycine-N-acylase. Biochemical Medicine 19:133-140.

Group, T. I. S. M. W. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928-933.

Hampf, M., Swart, A.C. and Swart, P. 1996. Sequence of the 11 beta-hydroxylase gene from the Cape baboon (Papio ursinus). Endocrine Research 22:495-499.

Hanahan, D. 1983. Studies on transformation of Escherichia coli with plasmids. Journal of Molecular Biology 166:557-580.

Heresco-Levy, U., Javitt, D. C., Ermilov, M., Mordel, C., Silipo, G., and Lichtenstein, M. 1999. Efficacy of high-dose glycine in the treatment of enduring negative symptoms of schizophrenia. Arch Gen Psychiatry 56:29-36.

Honchel, R., Aksoy, I. A., Szumlanski, C., Wood, T. C., Otterness, D. M., Wieben, E. D. and Weinshilboum, R. M. 1993. Human thiopurine methyltransferase: molecular cloning and expression of T84 colon carcinoma cell cDNA. Molecular Pharmacology 43:878-887.

Honchel, R., Aksoy, I. A., Szumlanski, C., Wood, T. C., Otterness, D. M., Wieben, E. D. and Weinshilboum, R. M. 1993. Human thiopurine methyltransferase: molecular cloning and expression of T84 colon carcinoma cell cDNA. Molecular Pharmacology 43:878-887.

Hubbard, T. J. P., Murzin, A. G., Brenner, S.E. and Chothia, C. 1997. SCOP: a structural classification of proteins database. Nucleic Acids Research 25:236-239.

Iida, Y., and Sasaki, F. 1983. Recognition patterns for exon-intron junctions in higher organisms as revealed by a computer search. The journal of biochemistry 94:1731-1738.

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

Janknecht, R., de Martynoff, G., Lou, J., Hipskind, R. A., Nordheim, A. and Stunnenberg, H. G. 1991. Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. Proceedings of the National Academy of Sciences 88:8972-8976.

Jakoby, W. B., and D. M. Ziegler. 1990. The enzymes of detoxication. Journal of Biological Chemistry 265:20715-8.

Jasinska, A. S., Susan, L., Matthew, S., Erin., L. Oliver, S., Eric, F., Lynn, B., Julia, J., Matthew, B., Herry, D., Ken, H., Thomas, P., Roberta, F. and Nelson, O. R. 2007. A genetic linkage map of the vervet monkey (Chlorocebus aethiops sabaeus). Mammalian Genome 18:347-360.

Jones, D. T. 1999. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. Journal of Molecular Biology 287:797-815.

Jones, P. G., R. A. VanBogelen, and F. C. Neidhardt. 1987. Induction of proteins in response to low temperature in Escherichia coli. Journal of Bacteriology 169:2092-2095.

Kang, T. Y., El-sohemy, A., Cornelis, M.C., Eny, K.M. and Bae, B.C. 2005. Glutathione S-transferase genotype and risk of systemic lupus erythematosus in Koreans. Lupas 14:381-384

Kazuo, A. 1991. Enzymes that metabolize acyl-coenzyme a in the monkey--their distribution, properties and roles in an alternative pathway for the excretion of nitrogen. International Journal of Biochemistry 23:429-434.

Kelley, M., and Vessey, D.A. 1992. Structural comparison between the mitochondrial arylkyl-CoA and arylacetyl-CoA N-acyltransferases. Biochemical journal 288:315-317.

Kohle, C., and Bock, K. W. 2007. Coordinate regulation of Phase I and II xenobiotic metabolisms by the Ah receptor and Nrf2. Biochemical Pharmacology 73:1853–1862.

Kolvraa, S., and Gregersen, N. 1986. Acyl-CoA: glycine N-acyltransferase: organelle localization and affinity toward straight-and branched-chained acyl-CoA esters in rat liver Biochemical Medicine and Metabolic Biology 36:98-105.

Kotewicz, M. L., Sampson, C. M., D'Alessio, J. M., and Gerard, G. F. 1988. Isolation of cloned moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. Nucleic Acids Research 16:265-277.

Krieger, I., and Tanaka, K. A. Y. 1976. Therapeutic Effects of Glycine in Isovaleric Acidemia. Pediatric Research 10:25-29.

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

Lau, E. P., Haley, B. E.B., and Roland E. 1977. The 8-azidoadenine analog of S-benzoyl (3'-dephospho) coenzyme A - A photoaffinity label for acyl CoA:glycine N-acyltransferase. Biochemical and Biophysical Research Communications 76:843-849.

Leavy, H. L. E., Arline M.Lott, Ira T.Kurtz, and Donald J. 1973. Isovaleric acidemia: results of family study and dietary treatment. Pediatrics 52:83.

Lim, K., and Chae, C.B. 1989. A simple assay for DNA transfection by incubation of the cells in culture dishes with substrates for beta-galactosidase. Biotechniques 7:576-579.

Liska, D. J. 1998. The detoxification enzyme systems. Alternative Medicine Review 3:187-198.

Loots, D. T., Mienie, L. J. and Erasmus, E. 2007. Amino-acid depletion induced by abnormal amino-acid conjugation and protein restriction in isovaleric acidemia. European Journal of Clinical Nutrition 61:1323-1327.

Mandel, M. a. H., A. 1970. Calcium-dependent bacteriophage DNA infection. Journal of Molecular Biology 53:159-162.

Marko, M. A., Chipperfield,R. and Birnboim, H.C. 1982. A procedure for the large-scale isolation of highly purified plasmid DNA using alkaline extraction and binding to glass powder. Analytical Biochemistry 121:382-387.

Mawal, Y. R., and Qureshi, A.I. 1994. Purification to homogeinity of mitochondial acyl-CoA: Glycine N-acyltarnsferase from human liver. Biochemical and Biophysical Research Communications 205:1373-1379.

Merkler, D. J., Merkler, K.A., Stern, W. and Flemming, F.F. 1996. Fatty acid amide biosynthesis: a possible new role for peptidylglycine a-amidating enzyme and acyl-Coenzyme A:glycine N-acyltransferase. Archives of biochemistry and biophysics 330:430-434.

Mitchell, G. A., Gauthier, N., Lesimple, A., Wang, S.P., Mamer, O., and Qureshi, I. 2008. Hereditary and acquired diseases of acyl-coenzyme A metabolism. Molecular Genetics and Metabolism.

Moldave, K. and. Meister., A. 1957. Synthesis of phenylacetylglutamine by human tissue. Journal of Biological Chemistry 229:463-476.

Murzin, A. G., Brenner, S.E., Hubbard, T. and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. Journal of Molecular Biology 247:536-540.

Nandi, D. L., Lucas, S.V., and Webster, L.T. 1979. Benzoyl-Coenzyme A:glycine N-acyltransferase and phenylacetyl-Coenzyme A:glycine N-acyltransferase from bovine liver mitochondria: purification and characterization. The Journal of Biological Chemistry 254:7230-7237.

Nishihara, K., Kanemori M., Kitagawa, M., Yanagi, H. and Yura, T. 1998. Chaperone coexpression plasmids: differential and synergistic roles of DnaK-DnaJ-GrpE and GroEL-GroES in assisting folding of an allergen of japanese cedar pollen, Cryj2, in Escherichia coli. Applied Environmental Microbiology 64:1694-1699.

Piper, W. N., Condie, L. W. and Tephly, T. R. 1973. The role of substrates for glycine acyltransferase in the reversal of chemically induced porphyria in the rat. Archives of Biochemistry and Biophysics 159:671-677.

Rinaldo, P., Schmidt-Sommerfeld, E.P., Anthony, P. H., Simon J. R., Woolf, D. A. and Leonard, J. V. 1993. Effect of treatment with glycine and l-carnitine in medium-chain acyl-coenzyme A dehydrogenase deficiency. The Journal of Pediatrics 122:580-584.

Roe, C. R., Millington, D.S., Maltby, D.A., Kahler, S.G. and Bohan, T.P. 1984. L-Carnitine therapy in isovaleric acidemia. The american society for clinical investigation 74:2290-2295.

Russell, D. W., and Sambrook, J. 2001. Molecular cloning:a laboratory manual, 3 ed, vol. 1. Cold spring harbor laboratory press, New York.

Sanger, F., Nicklen, S. and Coulson, A. R. 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74:5463-5467.

Schachter, D., and Taggard, J. V. 1953. Glycine N-acylase: purification and properties. The Journal of Biological Chemistry:263-275.

Schachter, D., and Taggart, J.V. 1954 (a). Glycine N-acylase: purification and properties. Journal of Biological Chemistry 208:263-275.

Schachter, D., and Taggart, J. V. 1954 (b). Product inhibition in the glycine N-acylase reaction. Journal of Biological Chemistry 211:271-278.

Schuster, S. C., Miller, W., Ratan, A., Tomsho, L.P., Giardine, A., Kasson, L.R., Harris, R.S, Petersen, D.C., Zhao, F., Alkan., C., Kidd., J.M., Sun, Y., Drautz, D.I.,Bouffard, I.,Muzny, D.M., Reid, J.G., Nazareth, L.V., Wang, Q., Burhans, R.,Riemer, C., Wittekindt, N.E., Moorjani, P., Tindall, E.A., Danko, C.G.,Teo, W.S., Buboltz, A.M., Zhang, Z., Ma, Q., Oosthuysen, A., Steenkamp, A.W., Oostuisen, H., Venter, P., Gajewski, J., Zhang, Y., Pugh, B.F., Makova, K.D., Nekrutenko, A., Mardis, E.R., Patterson, N., Pringle, T.H., Chiaromonte, F., Mullikin, J.C., Eichler, E.E., Hardison,

R.C., Gibbs, R.A., Harkins, T.T., and Hayes, V.M. 2010. Complete Khoisan and bantu genomes from Southern Africa. Nature 463:943-947.

Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. and Klenk, D. C. 1985. Measurement of protein using bicinchoninic acid. Analytical Biochemistry 150:76-85.

Snyders, M. (MSc in progress at the time of print) Cloning and Expression of Recombinant Bovine Glycine N-acyltransferase from Bovine Liver. North West University, Potchefstroom.

Sorensen, H., and Mortensen, K. 2005. Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. Microbial Cell Factories 4:1.

Stachulski, A. V., and Lennard, M. S. 2000. Drug Metabolism: The body's defense against chemical attack. Journal of Chemical Education 77:349-353.

Tanaka, K., M. A. Budd, M. L. Efron, and K. J. Isselbacher. 1966. Isovaleric Acidemia: A New Genetic Defect of Leucine Metabolism. Proceedings of the National Academy of Sciences of the United States of America 56:236-242.

Tanaka, K., Watakabe, A. and Shimura, Y. 1994. Polypurine sequences within a downstream exon function as a splicing enhancer. Molecular and cellular biology 14:1347-1354.

Tassaneeyakul, W. B., Donald, J. M., Michael, E. T., Wongwiwat, V., Maurice, E., Andersson, T.T., Robert, H. and Miners, J.O. 1994. Caffeine metabolism by human hepatic cytochromes p450: Contributions of 1A2, 2E1 and 3A isoforms. Biochemical Pharmacology 47:1767-1776.

Ullmann, A., Jacob, F. and Monod, J. 1967. Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of Escherichia coli. Journal of Molecular Biology 24:339-343.

van der Waide, J., and Steijns, L.S.W. 1999. Cytochrome P450 enzyme system: genetic polymorphisms andimpact on clinical pharmacology. Annual clinical biochemistry 36:722-729.

van der Westhuizen, F. H. 1998. An investigation into the origin of acyl-amino acid conjugation in the human and bovine metabolism. Thesis: submitted for the degree Philosophiae Doctor in the Department of Biochemistry and Microbiology of the Potchefstroom University for Christian Higher Education.

van der Westhuizen, F. H., Pretorius, P.J., and Erasmus, E. 2000. The utilization of alanine, glutamic acid, and serine as amino acid substrates for glycine N-acyltransferase. Journal of Biochemical Molecular Toxicology 14:102-109.

Venter, J. C. 2001. The sequence of the human genome. Science 291:1304-1351.

Vessey, D. A. 1978. The biochemical basis for the conjugation of bile acids with either glycine or taurine. Biochemical Journal 174:621-626.

Vessey, D. A., and Lau, E. 1998. Determination of the sequence of the arylacetyl acyl-CoA:amino acid N-acyltransferase from bovine liver mitochondria and Its homology to the aralkyl acyl-CoA:amino acid N-acyltransferase. Journal of biochemical molecular toxicology 12:275-279.

Vetting, M. W. S., de Carvalho, L. P. Y., Michael, H., Subray, S., Magnet, S. R., Steven, L. and Blanchard, J.S. 2005. Structure and functions of the GNAT superfamily of acetyltransferases. Archives of Biochemistry and Biophysics 433:212-226.

Waluk, D. P., Schultz, N., and Hunt, M. C. 2010. Identification of glycine N-acyltransferase-like 2 (GLYATL2) as a transferase that produces N-acyl glycines in humans. The Journal of the Federation of American Society of Experimental Biology 24:2795-2803.

Warburton, D. 2000. The molecular basis of lung morphogenesis. Mechanism of development 92:55-81.

Watson, A. F., and Langstaff, E. 1927. A note on the ammonium sulphate precipitation of the active principle of the the culture of the culture filtrates of C. Iphtheriae. Biochemical Journal 21:426-433.

Webster, L. T., Siddiqui, U. A., Lucas, S. V., Strong, J.M., and Mieyal, J.J. 1976. Identification of separate acyl-CoA: glycine and acyl-CoA:L-glutamine N-acyltransferase activities in mitochondrial fractions from Liver of rhesus monkey and man. The Journal of Biological Chemistry 251:3352-3358.

Wheeler, M. D., Ikejema, K., Enomoto, N., Stacklewitz, R. F., Seabra, V., Zhong, Z., Yin, M., Schemmer, P., Rose, M. L.,Rusyn, I., Bradford, B., and Thurman, R. G. 1999. Glycine: a new anti-inflammatory immunonutrient. Cellular and Molecular Life Sciences 56:843-856.

Yamamoto, A., Nonen, S., Fukuda, T., Yamazaki, H., and Zuma, J. 2009. Genetic polymorphisms of glycine N-acyltransferase in Japanese individual. Drug metabolic pharmacokinetics 24:114-117.

Yamashina, S., Ikejima, K., Enomoto, N., Takei, Y. and Sato, N. 2005. Glycine as a therapeutic immuno-nutrient for alcoholic liver disease. Alcoholism: Clinical and Experimental Research 29:162S-165S.

Zhang, H., Lang, Q, Li, J., Zhong, Z., Xie, F., Ye, G., Wan, B. and Yu, l. . 2007. Molecular cloning and characterization of a novel human glycine-N-acyltransferase gene GLYATL1, which activeates transcriptional activity of HSE pathway. International Journal of Molecular Science 8:433-444.

Appendix I: List of Tables

Appendix II: List of Figures

Figure 2.6: 1% Agarose gel analysis of the gel extracted PCR amplicons generated with GLYAT specific primers from total RNA of a chacma baboon.

Figure 2.7: A flow diagram illustrating how transformation of the pBabTAV001, pColdIII and pColdIII-EHDV into JM109 cells was conducted

Figure 2.8: A diagramatic representation produced using DNAMAN™ computer software illustrating the pBabTAV001 plasmid map with the expected restriction pattern by the enzymes NdeI, XhoI, BamHI, EcoRI and PstI

Figure 2.9: Restriction enzyme digestion of pBABTAV001 and pColdIII-EDHV

Figure 2.10: Amplification of the 3'-fragment (A) and 5'-fragment (B) which includes an open reading frame and an untranslated region (UTR) of the cDNA encoding for GLYAT of chacma baboon.

Figure 2.11: Gel extraction of PCR products containing (A) exons 1, 3, 4, 5 and 6 (B) exon 2 that encoding GLYAT of vervet monkey

Figure 2.12: Double stranded DNA sequence and the deduced amino acid sequence of the chacma baboon GLYAT transcript.

Figure 3.5: SDS-PAGE analysis of proteins expressed by Origami™ cells transformed with chaperone and pColdIII plasmids.

Figure 3.6: SDS-PAGE analysis of protein crude after expression of the chacma baboon recombinant GLYAT in Origami™ cells with the chaperones induced with 0.1 µg/L tetracycline.

Figure 3.7: Comparison of enzymatic activity in cytoplasmic extracts from livers of a chacma baboon and a bovine

Appendix III: List of Materials

Item	Catalogue number	Manufacture
Acrylamide	1.00209.100	Merck
Agar	BX.10.500	Merck
Agarose	H111206	Hispanagar
Ammonium persulfate	A3678	Sigma
Ammonium sulfate	A2939	Sigma
Ampicillin	A9393	Sigma
Benzoic acid	242381	Sigma
Benzonase	70746-3	Merck
Benzoyl coenzyme A	B1638	Sigma
Bicinchoninic acid solution	B9643	Sigma
Bovine serum albumin	775 827	Roche
Bugbuster	70584-4	Novagen
Chaperone plasmid set	3340v0902	TAKARA
Chloramphenicol	442513	Supelco
Coomassie brilliant blue	B8647	Sigma
Disposable cuvettes	165-2089	Ratiolab
DTNB	D8130	Sigma
DTT	43815	Sigma
EDTA	3658	Fluka

Ethanol 100%	100,983.25	Merck
Ethanol 97% (Drum)	C32102189	Rochelle Chemicals
Ethidium bromide	160539	Sigma
Ethyl acetate	1.09623.2500	Merck
ExTaq	RR001	Takara
Fermentas Generuler	SM1173	Fermentas
Fermentas instaclone PCR cloning kit	K1214	Fermentas
Fermentas pageruler	SM0811	Fermentas
Flexigene DNA purification kit	51206	Qiagen
MMLV High Performance Reverse Transcriptase	RT80125K	Epicentre biotechnologies
PCR clean up/ gel extraction kit (Nucleospin extract II)	740.609.250	Machery-Nagel
PureYield™ Plasmid Midiprep System	A2496	Promega
pCold DNA	3364 v.0708	Takara
T4 Ligase	9PIM180	Promega
RNeasy mini kit for RNA extraction	74106	Qiagen

Appendix IV: List of Abbreviations

Abbreviation	Meaning
bp	Base pair (s)
kDa	Kilo daltons
mg/ ml	Milligram per milli liter
ml	Milli liter
mm	Milli miter
mM	Milli mole per liter
min	Minute (s)
ng	Nanogram
pmol	Pico moles
sec	Second (s)
µg	microgram
U	Enzyme units
° C	Degrees Celsius
%	Percentage sign

Appendix V: A draft of a communication to be submitted to: The Journal of Molecular Biology

Novel open reading frames encoding glycine-N-acyltransferase of chacma baboon and vervet monkey reveal a possible novel variation at the protein level which groups their proteins as distinct from that of humans

MC Mahlanza and AA van Dijk*

*Biochemistry Division, North-West University, Private Bag X6001, 2520 Potchefstroom, South Africa. Tel: (+27) 18 299 2320

*Author to whom correspondence should be addressed: Albie.vandijk@nwu.ac.za

Abstract

Glycine-N-acyltransferase (GLYAT, EC 2.3.1.13) is one of the enzymes responsible for the biotransformation of endogenous and xenobiotic toxic compounds in mammals. It conjugates glycine to acyl-coA in mitochondria mainly in the liver and kidneys. We report novel open reading frames encoding GLYAT of the chacma baboon (*Papio ursinus*) and vervet monkey (*Chlorocebus pygerythrus)*. These new sequences reveal 8 amino acid differences, located in the GNAT fold, from the corresponding human enzyme. There have been no reports to date on characterisation of genes encoding for GLYAT of either the chacma baboon or vervet monkey. We cloned an open reading frame encoding the GLYAT of the chacma baboon in a TA vector and sequenced it using Sanger's DNA sequencing method. In addition, we amplified exons making up an

open reading frame encoding GLYAT from genomic DNA obtained from vervet monkey leukocytes. These amino acid sequences were aligned with those corresponding to the human, chimpanzee and rhesus monkey, whose GLYAT sequences were obtained from the *Ensembl* sequence database. The GLYAT amino acid sequences of the chacma baboon, vervet monkey and rhesus monkey formed a related group, distinct from other primates. The chacma baboon and vervet monkey sequences appeared to be 99% identical to the rhesus monkey sequence and 92.6% identical to the human sequence. The human and chimpanzee sequences were 99.7% identical to each other. We conclude that the vervet monkey, chacma baboon and rhesus monkey are closely related in terms of their GLYAT molecular phylogeny, whereas humans and chimpanzees form a separate group, in this respect.

Keywords: conjugation, GNAT, acyl-coA, acyl coenzyme A, ACGNAT, glycine acylase

A liver from a chacma baboon and blood sample from a vervet monkey were obtained to investigate the molecular characteristics of their glycine-N-acyltransferase (EC 2.3.1.13; also known as GLYAT, ACGNAT, CAT, GAT or glycine acylase). The GLYAT gene has not been characterised in chacma baboons and vervet monkeys. GLYAT detoxifies potentially harmful acyl-coAs in mammals by conjugating them to glycine, which renders them water soluble and excreted in urine and bile as acylglycines[1]. Patients with organic acidemias such as isovaleric acidemia, 3-methylcrotonylglycinuria and propionic acidemia usually respond well to glycine therapy and excrete acylglycines

in their urine[2,3]. The work presented here, is intended to be one of the first efforts that could serve as a basis for the rationale of designing recombinant GLYAT enzymes that can biotransform a much larger variety of substrates and possibly be targeted to treat organic acidemias.

GLYAT conjugates glycine to acyl-coA in mitochondria mainly in the liver and kidneys. It is a transferase enzyme with two substrates: an acyl donor and an acceptor. These substrates bind to GLYAT in a sequential reaction mechanism[4]. The GLYAT family belongs to the GCN5-related N-acetyltransferase superfamily because members of the GLYAT family display the characteristic core GNAT fold conserved by the more than 10 000 superfamily members[5]. This core GNAT fold is encoded by the C-terminal of the GLYAT proteins. Although there is an abundance of publications on structures of proteins belonging to the GNAT superfamily, the GLYAT protein structure remains unsolved. Hence, the information on GLYAT family members (GLYAT, GLYATL1, GLYATL2 and GLYATL3) has been gathered using any one of the three approaches: activity assays[1,6,4,7,8,9,10,11]; protein degradation studies[12,13] and amino acid sequences translated from open reading frames[8,11,14,15]. We use the last approach to contribute possible useful information about GLYAT protein in primates.

We report novel open reading frames encoding for GLYAT of the chacma baboon (*Papio ursinus*) and vervet monkey (*Chlorocebus pygerythrus)*. Their translated amino acid sequences suggested 8 variations relative to the corresponding enzyme of humans

in the region encoding the core GNAT fold. In addition we identified a possible novel amino acid variation at positions 32 and 201 of the vervet monkey and chacma baboon enzyme respectively.

Ethical approvals for the use of primate tissue

The chacma baboon liver tissue was obtained from a project that was approved by the Ethics Committee of North-West University (ethical approval number: NWU-00005-09-A1). A blood sample of a vervet monkey was obtained from a vervet monkey from a project also approved by the same Ethics Committee (ethical approval number of NWU-0022-09-S5).

Novel open reading frames encoding glycine-N-acyltransferase of the chacma baboon and vervet monkey suggest changes in amino acid sequence in the region encoding for the GNAT fold

The open reading frames encoding GLYAT of chacma baboon and vervet monkey were sequenced then translated into an amino acid sequence. The translated amino acids representing GLYAT of chacma baboon and vervet monkey were aligned to those coresponding to the human, chimpanzee and rhesus monkey, which were obtained from the *Ensembl* sequence database. The alignment is shown in figure 1 (a) and a phylogenic tree constructed (figure 1 (b)). It is known in literature that the conserved core GNAT fold is made up of amino acids from the C-terminal which runs through

positions 207 to 296 of the protein in contrast to positions 1 to 206 corresponds to the N-terminal of the protein. The conserved core GNAT fold is described as having an active function in binding and catalysis of the substrate. The theory of the GNAT fold does not prohibit any other amino acid, which is between positions 1 to 206, from interacting with substrate or interfering with the reaction mechanism at any stage. The C-terminal of the GLYAT protein from chacma baboon was compared to templates on Swiss-Model online server and it was observed that position 201 of GLYAT is located on the beta sheet region that begins the GNAT fold. We observed that the vervet monkey sequence introduced a novel variation at position 32 of the alignment by contributing an isoleucine (I32). On the other hand, the chacma baboon sequence introduced a novel variation at position 201 of the alignment by contributing a threonine (T201). These two amino acid positions were identified on the basis that they were the only positions where either the vervet monkey or chacma baboon showed a difference that had not been shown before by other primates. The implications of this T201 are unknown at the moment. The phylogenic tree showed that the vervet monkey, chacma baboon and rhesus monkey formed a distinct group separate from humans and chimpanzee. The chacma baboon and vervet monkey GLYAT amino acid sequences appeared to be 99% identical to the rhesus monkey sequence and 92.6% identical to the human sequence. The human and chimpanzee sequences were 99.7% identical to each other.

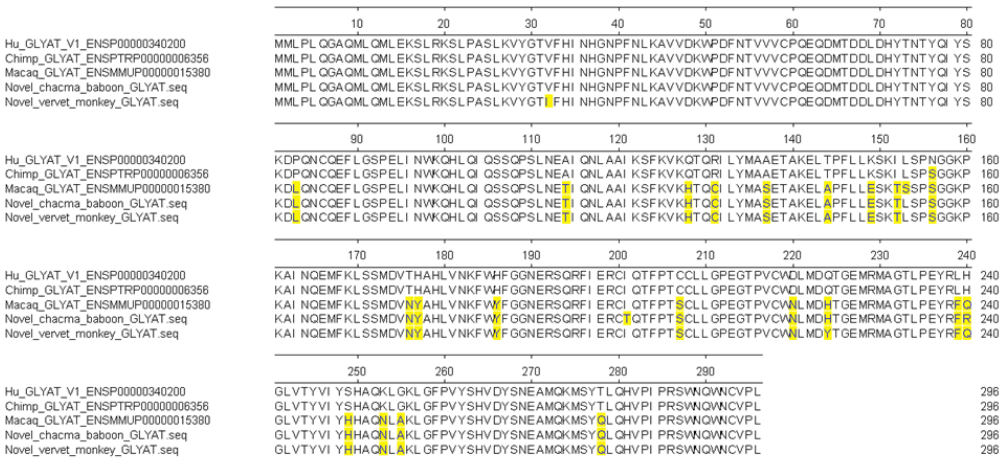


Figure 1(a): Multiple alignment of the novel GLYAT proteins of the chacma baboon and vervet monkey compared to that of human, chimpanzee and macaque obtained from Ensembl sequence database. The alignment was done using the Clustal V method. The highlighted amino acids are those that are exactly the same as the human GLYAT. Specific attention is drawn to amino acid positions 32 and 201, where the vervet monkey and chacma baboon introduces novel variations respectively.

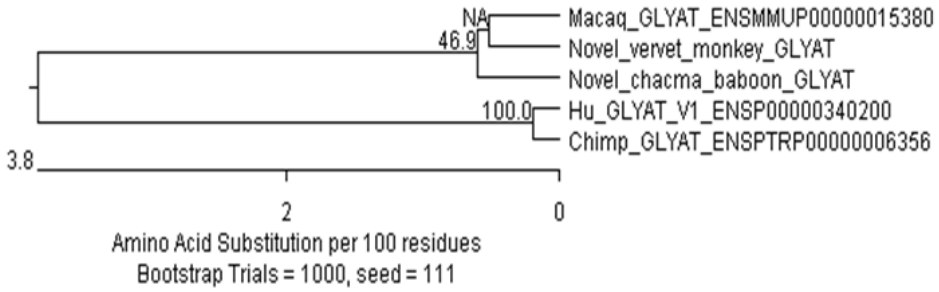
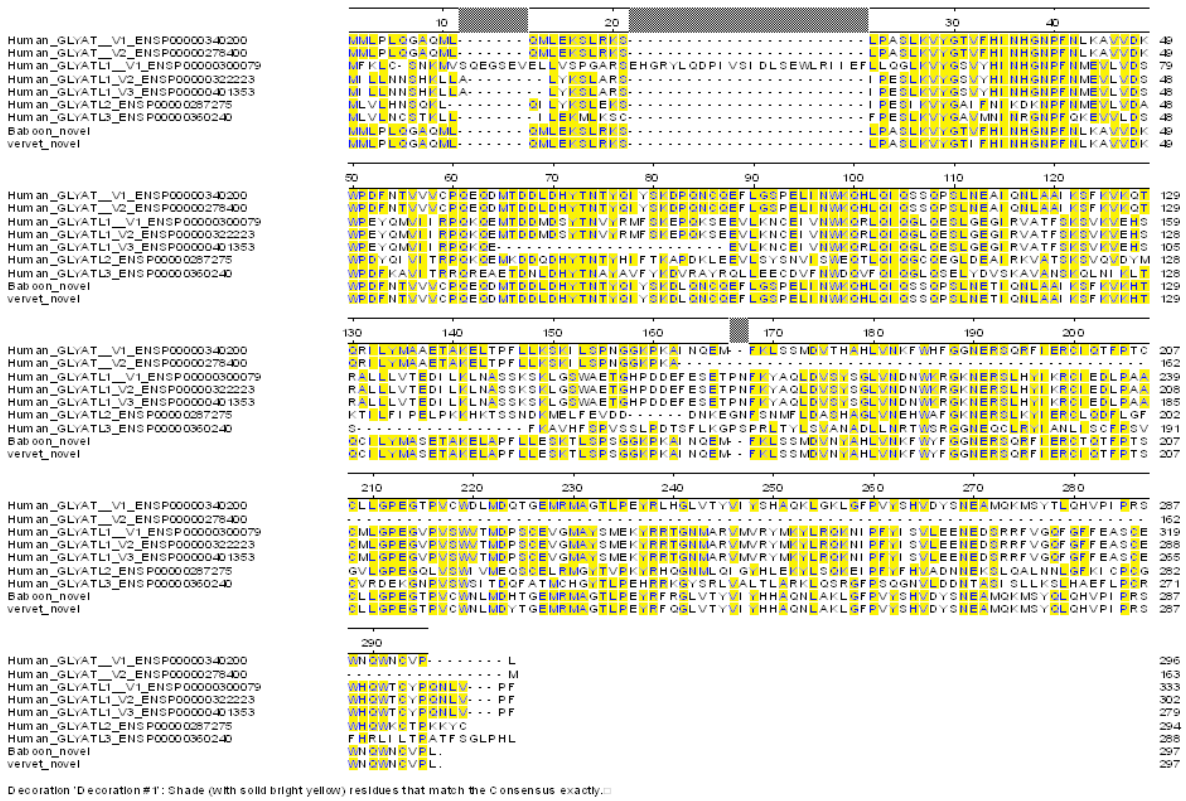


Figure 1. (b): A phylogenic tree analysis of GLYAT proteins from human, chimpanzee, macaque, chacma baboon and vervet monkey. The GLYAT proteins from macaque, chacma baboon and vervet monkey grouped distinctively from those coresponding to human and chimpanzee. The chacma baboon and vervet monkey sequences appeared to be 99 % identical to macaque sequence and 92.6 % identical to the human sequence. The human and chimpanzee sequences were 99.7 % identical to each other.

GLYATL3 of human has glutamate substituted with threonine in the protein region of the conserved GNAT fold

Glutamate at position 226 (E226) of bovine GLYAT corresponding to position 227 (E227) in primates has been pointed out as having a possible role in the catalytic mechanism of GLYAT[16]. This position was selected by those authors because it was the

most conserved position in GNAT superfamily members along the amino acids sequence making up the conserved GNAT fold of the GNAT superfamily. They used site-directed mutagenesis to illustrate the fact that E227 is of catalytic importance. In the work we present, the translated amino acid sequences representing GLYAT of chacma baboon and vervet monkey were aligned to all human GLYAT family members presently available in the *Ensembl* sequence database (figure 2). However, the observation made after aligning human GLYAT family members contradicted this hypothesis. The glutamate (E227) of GLYATL3 appeared to be substituted with threonine (T227). The authors that presented the E227 residue concluded that any GLYAT member without the glutamate residue at that position may not have any catalytic activity. However, it is possible that the threonine residue is characteristic of either the GLYATL3 enzyme or the specific individual animal that contributed the sequence. However, not much can be said about this contradiction at this stage since more experimental work appears to be required to either prove or disprove the hypothesis.



Decoration 'Decoration # 1': Shade (with solid bright yellow) residues that match the Consensus exactly.

Figure 2: Multiple sequence alignment of the translated GLYAT amino acid sequences of chacma baboon and vervet monkey compared to human GLYAT family members available on Ensembl sequence database. The alignment was conducted using the Clustal V method. The highlighted amino acids are those that are exactly the same as those for human GLYAT. Dashes were automatically inserted by the computer in order to obtain a better fit of the alignment. Human GLYATL1 isoform a (V1) of humans seems to have a long insertion of 23 amino acids starting at position 22. Human GLYATL1 isoform c (V2) seems to have a long deletion of 23 amino acids, starting from position 65. Amino acid at position 226 (E226) is suspected as having a role in GLYAT catalytic mechanism[9].

We conclude that the vervet monkey, chacma baboon and rhesus monkey are closely related with regard to their GLYAT molecular phylogeny, whereas humans and chimpanzee form a separate group. Characterisation of GLYAT from additional non-human primates will help achieve better understanding of GLYAT diversity and gain more insight into mammalian detoxification systems.

Experimental approach

Cloning and DNA sequencing of open reading frame encoding GLYAT of chacma baboon

Total cellular RNA was extracted from the liver of a chacma baboon using a RNeasy kit (QIAGEN, cat. # 74104). cDNA was synthesised using the MMLV High Performance Reverse Transcriptase kit (EPICENTRE® cat #RT80125K). An open reading frame encoding baboon GLYAT was amplified with human primers (see primers in table 1 for primer details) using the *Ex Taq* DNA polymerase (TAKARA® cat #PR001A). PCR product containing a cloned copy of an open reading frame encoding for GLYAT of chacma baboon were purified using a gel extraction kit (Macherey-Nagel Nucleospin Extract II® kit cat #740609.50) then cloned into a TA vector (Fermentas® instaclone PCR cloning kit cat #K1214). The M13 primer pair was used to sequence across the multiple cloning site of the TA vector (see Table 1 for primer details).

Amplification and DNA sequencing of an open reading frame encoding GLYAT from genomic DNA of vervet monkey

The FlexiGene® mammalian genomic DNA extraction kit was used to extract genomic DNA from buffy coat obtained from the blood of a vervet monkey. Separate exons making up an open reading frame encoding GLYAT of vervet monkey were amplified using human primers (see table 1 for primer details). PCR was performed using the Ex Taq DNA polymerase (TAKARA® cat #PR001A). PCR products containing exons making up GLYAT were purified using a PCR clean up kit (Macherey-Nagel Nucleospin Extract II® kit cat #740609.50) and then sequenced using the same primer used for PCR.

Table 1: A list of primers used for PCR. GAPDH primer pair was used as a positive control. Exons 1-6 primer pairs were used to amplify and sequence vervet monkey GLYAT exons. ORF primer pair was used to amplify chacma baboon GLYAT open reading frame. M13 primer pair was used to sequence across the muliple cloning site of the TA vector.

Region amplified	Primer direction	Nucleotide sequence	Annealing Temperature °C
GAPDH	Forward Reverse	5'- gaaggtgaaggtcggagtc-3' 5'-gaagatggtgatgggatttc-3'	62
Exon 1	Forward Reverse	5'-cagattcttttgccagcctagtac-3' 5'-cactcatgtagcatggatcccatataca-3'	56
Exon 2	Forward Reverse	5'-cagctcgttctcagaggagtcag-3' 5'-gcagtgtttagactaagg-3'	60
Exon 3	Forward Reverse	5'-agtggttgtctgcctctctgtg-3' 5'-gccctggctctaccatattgc-3'	65
Exon 4	Forward Reverse	5'-caggatatgacagatgaccttgat-3' 5'-tctggagcttggaggaag-3'	60
Exon 5	Forward Reverse	5'-ggaaagccagagtgaatgcag-3' 5'-tagcaccaagcccagaacc-3'	65
Exon 6	Forward Reverse	5'-gattctcacagacaccaaatctgctg-3' 5'-cttcactctgttcctctttcatca-3'	56
ORF	Forward Reverse	5'-aattcatatgatgttaccattacaaggtgc-3' 5'-aattctcgagcagaggcacacagttccac-3'	62
M13 (sequencing of the TA vector)	Forward Reverse	5'-gttttcccagtcacgac-3' 5'-caggaaacagctatgac-3'	

References

1. Schachter D, & Taggard, J. V. 1953. Glycine N-acylase: purification and properties. The Journal of Biological Chemistry:263-275.

2. Krieger I, Tanaka KAY. 1976. Therapeutic Effects of Glycine in Isovaleric Acidemia. Pediatric Research 10(1):25-29.

3. Roe CR, Millington, D.S., Maltby, D.A., Kahler, S.G. and Bohan, T.P. 1984. L-Carnitine therapy in isovaleric acidemia. The american society for clinical investigation 74:2290-2295.

4. Nandi DL, Lucas, S.V., & Webster, L.T. 1979. Benzoyl-Coenzyme A:Glycine N-Acyltransferase and Phenylacetyl-Coenzyme A:Glycine N-Acyltransferase from Bovine Liver Mitochondria: purification and characterization. The Journal of Biological Chemistry 254(15):7230-7237.

5. Vetting MW, S. de Carvalho LP, Yu M, Hegde SS, Magnet S, Roderick SL, Blanchard JS. 2005. Structure and functions of the GNAT superfamily of acetyltransferases. Archives of Biochemistry and Biophysics 433(1):212-226.

6. Webster LT, Siddiqui, U. A., Lucas, S. V., Strong, J.M., & Mieyal, J.J. . 1976. Identification of separate acyl-CoA: glycine and acyl-CoA:L-glutamine N-acyltransferase activities in mitochondrial fractions from Liver of rhesus monkey and man. The Journal of Biological Chemistry 251(11):3352-3358.

7. Mawal YR, & Qureshi, A.I. 1994. Purification to homogeinity of mitochondial acyl-CoA: Glycine N-acyltarnsferase from human liver. Biochemical and Biophysical Research Communications 205(2):1373-1379.

8. Zhang H, Lang, Q, Li, J., Zhong, Z., Xie, F., Ye, G., Wan, B. & Yu, l. . 2007. Molecular cloning and characterization of a novel human glycine-N-acyltransferase gene GLYATL1, which activeates transcriptional activity of HSE pathway. International Journal of Molecular Science 8:433-444.

9. van der Westhuizen FH, Pretorius, P.J., & Erasmus, E. 2000. The utilization of alanine, glutamic acid, and serine as amino acid substrates for Glycine N-acyltransferase. Journal of Biochemical Molecular Toxicology 14(2):102-109.

10. Kasuya F, Yamaoka, Y., Osawa, E., Igarashi, K. and Fukui, M. 2000. Difference of the liver and kidney in glycine conjugation of ortho-substituted benzoic acids. Chemico-Biological Interactions 125:39-50.

11. Waluk DP, Schultz, N. and Hunt, M.C. 2010. Identification of glycine N-acyltransferase-like 2 (GLYATL2) as a transferase that produces N-acyl glycines in humans. FASEB journal.

12. Vessey DA, and Lau, E. 1998. Determination of the sequence of the arylacetyl acyl-CoA:amino acid N-acyltransferase from bovine liver mitochondria and Its homology to the aralkyl acyl-CoA:amino acid N-acyltransferase. Journal of biochemical molecular toxicology 12(5):275-279.

13. Kelley M, and Vessey, D.A. 1992. Structural comparison between the mitochondrial arylkyl-CoA and arylacetyl-CoA N-acyltransferases. Biochemical journal 288:315-317.

14. Yamamoto A, Nonen, S., Fukuda, T., Yamazaki, H., and Zuma, J. 2009. Genetic polymorphisms of glycine N-acyltransferase in Japanese individual. Drug metabolic pharmacokinetics 24(1):114-117.

15. Cardenas CLL, Bourgine, J., Cauffiez, C., Allorge, D., Lo-Guidice, J.M., Broly, F., and Chevalier, D. 2010. Genetic polymorphisms of Glycine N-acyltransferase (GLYAT) in a French Caucasian population. Xenobiotica 0(0):1-9.

16. Badenhorst, C.P.S., Snyderss, M & van Dijk, A.A. (2010). Bacterial expression and elucidation of the catalytic mechanism of the glycine N-acyltransferase. J Inherit Metab Dis. (Abstract) 33, pS53.