Novel tryptase sequences from non-human primates

Supervisor: Dr John E Hunt

Word Count: 5 095

Novel nucleotide sequences from this report have been submitted into Genbank as EF_436230, EF_436231, and EF_436232.

ABSTRACT

Tryptases are tetrameric serine proteases secreted by human mast cells and play a prominent role in diseases such as asthma, inflammatory bowel disease, and rheumatoid arthritis. Human tryptases are part of a multi-gene family and include the α, β, δ, γ, and ε tryptases. However, there is debate over whether the α- and δ-tryptases are functional. Although there are multiple tryptases in most mammals, homologs of human δ-tryptase, or indeed of α/β-tryptases, are not present in rodent genomes. Surprisingly little attention has been paid to tryptases in non-human primates, with only two published sequences (one each from the chimpanzee and the cynomolgus monkey). Therefore, we screened genomic DNA from non-human primates for novel δ-tryptase and non-δ-tryptase homologs by PCR. No δ-tryptase-like sequences were detected in the genomes of six non-human primates selected from different branches of the evolutionary tree (including the chimpanzee, orangutan, gorilla, two macaques, and a lemur). We successfully amplified and cloned α/β-tryptase-like sequences from the chimpanzee, cynomolgus monkey, and rhesus monkey. The amplified sequences covered the region from residue 60B (chymotrypsin numbering) to the carboxy terminus, and were at least 90% identical to human α/β-tryptases at both the nucleotide and protein level. Phylogenic tree modelling revealed that the known and novel primate sequences cluster into two distinct groups – α-tryptase-like and β-tryptase-like. The presence (and conservation) of α-like tryptases in humans and in primates that diverged from humans approximately twenty five million years ago, suggests that these proteases might have important physiological roles and are most probably enzymatically active. In addition, the conserved nature of the α/β tryptase dichotomy in humans and non human primates indicates that these primates might serve as better animal models for the study of tryptases in inflammatory diseases than the rodent models presently used.
INTRODUCTION

Mast cells are granulated effector cells of the immune system and are believed to play a role in inflammation, tumour angiogenesis, immunoregulation, and tissue repair (1). Tryptases are active at neutral pH, and their production helps distinguish mast cells from other leucocytes (2, 3).

Tryptases are tetrameric serine proteases secreted by human mast cells and have a molecular weight of ≈134kDa (monomer ≈26-35kDa) (1-4). Tryptases account for ≈25% of the total protein content of mast cell granules, and are stored in their active form prior to release during mast cell degranulation (5). However, their actions appear to be restricted to the extracellular milieu (6). The arrangement of tryptase’s catalytically active subunits produces a small oval central pore (size ≈50×30Å), resulting in restricted accessibility for substrates and inhibitors (6, 7). The tryptase monomer is arranged with six externally exposed domains, which interact with its external environment, including its neighbouring monomers, when in the tetramer formation (8). These domains are tryptase’s surface loops, and are named the “37-loop”, the “60-loop”, the “70- to 80-loop”, the “97-loop”, the “147-loop”, and the “173-loop” (9). As these loops surround the active site, any changes in these loops can potentially alter the substrate specificity of tryptase (8, 9). In addition, tryptase contains a catalytic triad (His-57, Asp-102, Ser-195 (chymotrypsin numbering¹)), which is essential for its proteolytic activity (6).

Although a number of substrates have been identified for tryptases in vitro, their true biological roles and targets are still unclear (7). However, they are reported to induce microvascular leakage and inflammatory cell accumulation, and regulate mast cell activation (4). They are therefore important mediators of inflammation (4, 7), and have a prominent role in diseases such as asthma, inflammatory bowel disease, and rheumatoid arthritis (10-13). In addition, tryptase activity can be inhibited by synthetic protease inhibitors, and several such therapeutic approaches have shown clinical efficacy in the treatment of asthma and ulcerative colitis (10, 11, 14).

Multiple human tryptases have been identified, including α, β, δ, γ, and ε (7), however, uncertainty exists as to whether all forms are functional. The genes that encode human tryptases are located in a cluster within a 2.5Mb region, on the short arm of chromosome 16, at position 16p13.3 (1, 6, 7). All known human tryptase genes have a six exon/five intron organisation, which is approximately 1.8kb long (7, 15). This gene architecture differs from that of other mast cell or leucocyte serine proteases (6). Tryptases contain a 30-amino acid prepropeptide followed by a 245-amino acid catalytic

¹ All subsequent residues in this report will be referred to according to the chymotrypsin numbering system.
domain (15). Although the 5′ regulatory region is similar to other serine proteases, it is unique due to its separation from the initiator Met codon by the first intron (6, 15).

In most mammalian genomes, tryptases belong to a well-conserved multi-gene family, with highly conserved intron phase and size, and intron/exon boundary positions (7). Tryptases have been found in several mammals (e.g. human, dog, mouse, rat, gerbil, sheep and cow) (7), and there has been two reports of primate tryptases to date. Guida and coworkers (16) sequenced a tryptase gene in Macaca fascicularis (cynomolgus monkey), and Sakaguchi and coworkers (17) purified tryptase from the same species in the same year. There is also a predicted sequence of chimpanzee tryptase published on GenBank (accession no. XM_511185). The macaque tryptase protein was found to be 91% and 88% identical to what Guida et al. called human tryptase 1 and 2, respectively (16), with characteristics similar to those of human tryptases (17). Although Guida et al. did not report the presence of any other monkey tryptases (16), we believe that it is possible that multiple tryptases exists in individual non-human primates, as multiple tryptases have been found in all mammalian models studied to date (7).

The first human tryptase cDNA was cloned in 1989 by Miller and coworkers (18). It was named α-tryptase following the cloning of β-tryptase in the subsequent year (19). β-tryptase was then divided into βI, βII, and βIII, with the discovery of skin tryptases similar to the original β-tryptase by Vanderslice and coworkers (15). α-tryptase was later divided into two allelic variants - αI and αII (1). Not only is the amino acid sequence of α- and β-tryptase ≈90% identical (1, 19), but they are 97% identical at the nucleotide level (16). Recent evidence suggests that the α-tryptases are allelic variants of βI-tryptase as they occupy the same locus. The other β gene locus contains the allelic partners βII and βIII (3). As α- and βI-tryptase share the same locus, Soto et al found that ≈29% of individuals in their study had α-tryptase deficiency, with a genotype of ββββ and no α-tryptase genes (20). Unlike β-tryptase, α-tryptase is not stored as mature proteins in mast cell granules, but rather, is secreted as an inactive precursor (1, 7). Pallaoro et al (21) reported a sequence for a third form of human tryptase – mouse mast cell protease 7 (mMCP-7)-like tryptase, later renamed δ-tryptase (2). It is thought to be a chimera resulting from gene conversion events between an α/β-type tryptase gene and one related to the mMCP-7 gene (3).

There has been much discussion regarding the activity of α-tryptase, and there is debate over whether α-tryptase is a functional protein. Certain amino acid differences between α- and β-tryptase results in vastly different protein structures. It was believed that the glutamine at position -3 (arginine in β-tryptase) affected the three-dimensional folding, and therefore the function, of α-
tryptase (22). However, Huang et al. subsequently showed that this was not the case (23). Huang et al. also believes that residue 216 lies on one of tryptase’s surface loops. The amino acid difference at this position – glycine (β-tryptase) versus aspartic acid (α-tryptase) – results in a change in the substrate-binding region of α-tryptase, thus leading to a restricted substrate specificity relative to β-tryptase (22, 23). Huang et al. (23) have shown that despite this, α-tryptase is still enzymatically active. However, it is also less stable than β-tryptase (24).

Although multiple tryptases have been found in many mammals (7), a homolog of human δ-tryptase is not present in the murine genome, and has not yet been reported in any non-human primate (3). In our study, we attempted to amplify δ-tryptase sequences from primates of different branches of the evolutionary tree – humans, chimpanzees, orangutans, gorillas, macaques, and lemurs (figure 1). We were unable to find any δ-tryptases in the primates we researched. Thus, we proceeded to find non-δ tryptases and we were successful in obtaining α/β-like sequences from the chimpanzee, and two macaques. Chimpanzees are the closest primate relatives to humans, with a common ancestor ≈4-6 million years ago (mya). Comparatively, humans are believed to have diverged from macaques (Old World Monkeys) ≈25mya (shown in figure 1) (25, 26). We believed that we could design PCR primers confidently as the genome and its chromosomal arrangement are highly conserved in humans and non-human primates (26-28). DNA sequence comparisons at many loci showed that there is a strong similarity between humans and African great apes, with ≈98.5% sequence identities (25, 26, 29).

Figure 1: Primate phylogenetic relationships. Figure represents the evolutionary history of primates. Primates researched in this study are boxed in red. Primates were chosen to represent different branches of the evolutionary tree.
As tryptases appear to play important roles in diseases such as asthma, ulcerative colitis and rheumatoid arthritis, an appropriate animal model with tryptase pathophysiology similar to humans will be very useful in future clinical trials. Due to our close evolutionary relationship with primates, especially hominids, they may prove to be the most useful animal model for the study of tryptase activity in humans.

**MATERIALS & METHODS**

*DNA Samples*

Genomic DNA from the gorilla (*Gorilla gorilla*), cynomolgus monkey (*Macaca fascicularis*), orangutan (*Pongo pygmaeus*), chimpanzee (*Pan troglodytes*) & lemur (*Lemur variegatus variegatus*) were obtained from the Australian Museum (catalogue numbers EBU_17384, EBU_7967, EBU_17404, EBU_40812 and EBU_10675 respectively). Macaque genomic DNA (*Macaca fascicularis* – cynomolgus monkey and *Macaca mulatta* – rhesus monkey) were purchased commercially (BioChain, Hayward, CA, USA).

*Polymerase chain reaction amplifications*

Generic tryptase primers (forward 5′-TGCCCAGGGACGTCAAGGAT-3′, reverse 5′-AAGGGTCCTCAGGACAGGGAA-3′) were designed in regions conserved in all mammals using Invitrogen’s Vector NTI Advanced 9.0 software program. The reverse primer was located in the STS (sequence tagged site) for tryptase (UniSTS: 61561), a unique DNA segment in the human genome, to minimise amplification of non-tryptase gene fragments. The polymerase chain reaction (PCR) amplifications were optimized using a PCR Optimisation Kit (Invitrogen, Mount Waverley, VIC) according to manufacturer’s instructions. Subsequently, buffer F (2.0mM Mg\(^{2+}\), pH 9.0) was utilized in 50µl reactions with 50ng genomic DNA, 7.5ng of each primer and Platinum Taq High Fidelity DNA polymerase. Thermal cycling was performed on a Perkin Elmer GeneAmp PCR System 2400 thermal cycler with an initial denaturing step at 94°C for 5mins, followed by 2 cycles of a denaturing step at 94°C for 30s, a primer annealing step at 50°C for 30s, and an elongation step at 72°C for 30s, followed by 30 cycles with annealing step at 55°C, and a final extension at 72°C for 8mins.

*TA Cloning*

TA cloning was performed using the Invitrogen TA Cloning Kit according to manufacturer’s instructions. Following ligation of the PCR product with the pCR2.1-TOPO vector (4µl PCR product, 1µl vector, 1µl salt solution (1.2M NaCl, 0.06M MgCl\(_2\))), 2µl of the ligation mix was incubated with 50µl of competent cell mixture (Promega JM109 cells). The cells were plated on a
LB Agar plate with 100µg/ml ampicillin, and incubated overnight. Colonies were subsequently picked at random and screened for the presence of correctly sized inserts through PCR (25 cycles at annealing temperature of 55°C) using the M13 primers (forward (F) 5'-GTAAAACGACGGCCAG-3', and reverse (R) 5'-CAGGAAACAGCTATGAC-3').

**DNA Sequencing**

Single colonies with inserts of the correct size were grown in LB media with 100µg/ml ampicillin for 4-6 hours at 37°C with shaking (220rpm). Plasmid DNA was then isolated using the Promega Wizard Plus SV Miniprep DNA Purification System, and its nucleotide sequence determined using the ABI PRISM BigDye Terminator v3.1 Cycle Sequencing Kit, according to the respective manufacturer’s instructions. In the first instance, sequencing was performed using the vector primers M13F and M13R (as above). The sequencing reaction conditions were 96°C for 1min, then 25 cycles of 96°C for 10s, 50°C for 5s, and 60°C for 4mins, followed by 7mins at 4°C, performed on a Perkin Elmer GeneAmp PCR System 2400 thermal cycler. Samples were then precipitated in ethanol and dried by evaporation at room temperature in the dark. The sequencing reactions were sent to the Ramaciotti Centre for Gene Function Analysis (UNSW) for sequencing.

Sequencing of the entire insert in both directions was completed using internal primers designed according to the results of the previous sequencing run (primers shown in table 1). Sequences were aligned with known human and non-human primate tryptase sequences using Invitrogen’s Vector NTI Advance 9.0 AlignX program, which uses a modified Clustal W algorithm to find identities and any base or amino acid substitutions (30). Novel sequences from the non-human primates were checked against the human SNP databases (16, 31) to ensure that the sequences did not originate from contamination with human genomic DNA.

<table>
<thead>
<tr>
<th>Primate</th>
<th>F/R</th>
<th>Sequence (5’→3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chimpanzee (Pan troglodytes)</td>
<td>F</td>
<td>GTGAAGGTCCCCCATAAT</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>CAAAAAGCCCTGAGTCA</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ATTCACCTTGACACCA</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AGACCCACCATCATTGT</td>
</tr>
<tr>
<td>Cynomolgus monkey (Macaca fascicularis)</td>
<td>F</td>
<td>GTGAAGGTCCCCCATAAT</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>TGGTGACAGGTAAT</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ATTCACCTTGACACCA</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>AGACCCACCATCATTGT</td>
</tr>
<tr>
<td>Rhesus monkey (Macaca mulatta)</td>
<td>F</td>
<td>AAATACCACCTCGCCT</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>ATGTCATCGCGATGAT</td>
</tr>
</tbody>
</table>

Table 1: Internal primers used in sequencing. The “F/R” column indicates whether the primer is a forward (F) or a reverse (R) primer.
Phylogenetic Tree Modelling

Novel tryptase sequences from this study and relevant published sequences (human $\beta$-tryptase, human $\alpha$-tryptase, chimpanzee tryptase, and cynomolgus monkey tryptase) were analysed using Protpars (protein sequence parsimony) and DNApars (DNA parsimony) in the PHYLIP package. Both programs infer an unrooted phylogeny from protein and DNA sequences respectively. Protpars assesses phylogeny by determining the number of steps needed for each amino acid change. Only changes of amino acids that are consistent with the genetic code are included, with allowances for changes through a third amino acid (32). DNApars (32) utilises an alternate method outlined by Fitch (33). The results obtained from these programs were then graphically represented using the web server Phylodendron which graphs the data in both unrooted tree diagrams and single root phenograms (34).

RESULTS

Primate $\delta$-tryptase-homologs

Despite using primers based on the human $\delta$-tryptase sequence and extensive PCR optimization, $\delta$-tryptase homologs were not detected in any of the six primate genomes screened (data not shown). In control reactions we successfully amplified $\delta$-tryptase from human genomic DNA using the same primers, and other non-$\delta$ tryptase genes from primate genomic DNA using generic tryptase primers.

Primate non-$\delta$-tryptase-homologs

Using the pan-specific tryptase PCR primers, amplimers of the expected size were generated from the genomic DNA of the chimpanzee, the cynomolgus monkey, and the rhesus monkey (see figure 2). These PCR products were cloned using the pCR2.1-TOPO vector (Invitrogen, Mount Waverley, VIC), and their nucleotide sequence determined. All three sequences were similar to each other as well as to the human $\alpha$- and $\beta$-tryptases. The chimpanzee and cynomolgus monkey sequences were more similar to $\beta$-tryptase (96.4% and 99% respectively), whereas the rhesus monkey sequence was more similar to $\alpha$-tryptase (90.4%) (see table 2, figure 5). Consequently amino acid comparisons were made with the corresponding human tryptase ($\alpha$ or $\beta$). Although the same primers and conditions were used for all primates, no amplimers of the correct size were generated from the genomic DNA of the other primates.
Figure 2: Agarose gel electrophoresis of PCR reactions performed on non-human primate genomic DNA and using generic tryptase primers. The primers used generate a 993bp amplimer from human genomic DNA, and an amplimer of similar size in the chimpanzee, and the cynomolgus and rhesus monkeys. All gels show a 1kb ladder in the first lane, and human and no template controls in the last two. A) Using cynomolgus monkey (from the Australian Museum), orangutan, and chimpanzee genomic DNA and a primer concentration of 10ng/µl. B) Using lemur genomic DNA with varying primer concentrations (shown on figure). C) Using cynomolgus monkey, rhesus monkey, and gorilla genomic DNA, with varying primer concentrations (shown on figure).

As we will refer both to our novel sequences and to published sequences in the following discussion, we will refer to the published chimpanzee tryptase as PT1, Guida et al.’s (16) cynomolgus monkey tryptase as MF1, and our novel tryptases as PT2 (chimpanzee), MF2 (cynomolgus monkey), and MM1 (rhesus monkey).
A) PT2

The fragment of the tryptase gene, PT2, amplified from the chimpanzee (deposited in GenBank as accession no. EF_436230) was 994 base pairs (bp) long (including 775bp of the cDNA). This translates into a 197 amino acid segment of the mature protein and spans from residue 60B in exon 4 to the stop codon in exon 6 (figure 3A). The chimpanzee nucleotide sequence is highly homologous to human β-tryptase and human α-tryptase (97.5% and 96.3% identical respectively), and to PT1 (GenBank accession no. XM_511185) (99.4% identical) (see table 2A). Similarly, the translated protein is 96.4% identical to human β-tryptase, 94.4% identical to human α-tryptase, and 98.5% identical to PT1 (see table 2B). Of the thirteen amino acid differences that distinguish human α-tryptase from human β-tryptase, ten reside within the segment of sequence we obtained. The

Figure 3: Nucleotide and deduced amino acid sequences of PT2, MF2, and MM1. Shown are the nucleotide sequences of A) PT2, B) MF2, and C) MM1. Exons 4-6 are boxed in red. The putative amino acid sequence (shown in three letter code) is shown above the nucleotide sequence. Slight differences in amplifier length (PT2 (994bp), MF2 (993bp), and MM1 (988bp)) are due to deletions in intron 5 (MF2 and MM1 versus PT2), and a 4bp deletion in exon 6 that occurs after the stop codon in MM1.

Chimpanzee (PT2)

The fragment of the tryptase gene, PT2, amplified from the chimpanzee (deposited in GenBank as accession no. EF_436230) was 994 base pairs (bp) long (including 775bp of the cDNA). This translates into a 197 amino acid segment of the mature protein and spans from residue 60B in exon 4 to the stop codon in exon 6 (figure 3A). The chimpanzee nucleotide sequence is highly homologous to human β-tryptase and human α-tryptase (97.5% and 96.3% identical respectively), and to PT1 (GenBank accession no. XM_511185) (99.4% identical) (see table 2A). Similarly, the translated protein is 96.4% identical to human β-tryptase, 94.4% identical to human α-tryptase, and 98.5% identical to PT1 (see table 2B). Of the thirteen amino acid differences that distinguish human α-tryptase from human β-tryptase, ten reside within the segment of sequence we obtained. The
The chimpanzee sequence had 7 β-like amino acids and 2 α-like amino acids (shown in table 3), and one residue not shared with either.

Of the three members of the catalytic triad, two lie within the sequenced segment (Asp-102 and Ser-195), and were both present (see figure 4). This sequence has six amino acids which differ from human β-tryptase, and three amino acids which differ from PT1. Interestingly, at residue 60C, the phenylalanine of PT1 is not present, rather, as in the human, there is a valine. Also, the histidine at position 95 and the tyrosine at 239 differ from both the human and PT1. Five of the six changes between PT2 and human β-tryptase were not conserved, and of these, three lie on surface loops. When compared to PT1, two of the three changes were not conserved, and only one lies on a surface loop.

A) cDNA

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>α</th>
<th>PT1</th>
<th>PT2</th>
<th>MF2</th>
<th>MM1</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>96.1</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT1</td>
<td>97.3</td>
<td>96.3</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT2</td>
<td>97.5</td>
<td>96.3</td>
<td>99.4</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF2</td>
<td>98.7</td>
<td>97</td>
<td>96.9</td>
<td>96.6</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>MM1</td>
<td>90.7</td>
<td>90.6</td>
<td>89.9</td>
<td>89.5</td>
<td>90.8</td>
<td>100</td>
</tr>
</tbody>
</table>

B) Protein

<table>
<thead>
<tr>
<th></th>
<th>β</th>
<th>α</th>
<th>PT1</th>
<th>PT2</th>
<th>MF1</th>
<th>MF2</th>
<th>MM1</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>93.1</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT1</td>
<td>96</td>
<td>94.2</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PT2</td>
<td>96.4</td>
<td>94.4</td>
<td>98.5</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF1</td>
<td>89.8</td>
<td>88.7</td>
<td>88.3</td>
<td>88.7</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MF2</td>
<td>99</td>
<td>94.4</td>
<td>97</td>
<td>96.4</td>
<td>89.8</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>MM1</td>
<td>90.4</td>
<td>90.4</td>
<td>88.3</td>
<td>87.8</td>
<td>97.8</td>
<td>89.3</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2: Percentage identities of primate sequences. As indicated in the key above, yellow indicates identities of 100%, blue indicates identities <100% and ≥98%, green indicates identities <98% and ≥95%, and fuchsia <95% and ≥90%. A) Identities between the cDNA sequences starting from exon 4. B) Identities between the protein sequences starting from residue 60B of the mature protein. C) Featuring MF2, comparing the identities with human α- and β-tryptase of the cDNA and protein sequence.

Cynomolgus monkey (MF2)

The 993bp sequence (775bp cDNA) obtained from the cynomolgus monkey using the same primers (Deposited as GenBank accession no. EF_436231), also translated into a 197 amino acid segment of the mature protein (figure 3B), situated in the same area of the gene (residue 60B onwards). This sequence was also found to be more similar to human β-tryptase than human α-tryptase, with a surprisingly higher identity to human β-tryptase than the chimpanzee sequence discussed.
previously. The cDNA sequence was 98.7% identical to human β-tryptase and 97% to α-tryptase (see table 2A), whereas the translated protein sequence was 99% and 94.4% identical respectively (see table 2B). There was 89.8% identity between this protein sequence and MF1 (16). Unfortunately, there was no published nucleotide sequence to compare MF2 to. The close similarity to human β-tryptase is revealed by the presence of nine of the ten “fingerprint” amino acids of β-tryptase in MF2, and none of the α-like residues.

| Residue | 21 (36) | 23 (37A) | 46 (59) | 55 (63) | 85 (96) | 86 (97) | 88 (99) | 103 (114) | 106 (117) | 111 (122) | 138 (150) | 175 (176) | 215 (216) | β-like | α-like |
|---------|---------|---------|---------|---------|---------|---------|---------|-----------|-----------|-----------|-----------|-----------|----------|--------|
| β       | H       | P       | V       | A       | T       | A       | I       | V         | H         | T         | R         | V         | G        |        |
| α       | R       | R       | L       | T       | I       | I       | T       | I         | R         | M         | P         | I         | D        |        |
| PT1     | R       | P       | V       | A       | T       | A       | I       | R         | T         | S         | V         | G        |         |        |
| MF1     | H       | ?       | V       | D       | A       | V       | T       | V         | H         | T         | P         | I         | D       | 9/13    | 3/13    |
| PT2     | A       | I       | A       | I       | R       | T       | S       | V         | G        |           |           |           |          | 7/10    | 2/10    |
| MF2     | A       | T       | A       | I       | A       | H       | T       | R         | V         | G        |           |           |          | 9/10    | 0/10    |
| MM1     | D       | T       | V       | T       | V       | H       | T       | P         | I         | D        |           |           |          | 4/10    | 4/10    |

Table 3: Fingerprint amino acids of human α- and β-tryptase. The table shows the 13 amino acids that differ between α- and β-tryptase, and the respective amino acids in the non-human primate sequences. The table indicates the residue number in the mature tryptase protein, with the respective chymotrypsin residue number in brackets. β-like amino acids are shown in red, whilst α-like amino acids are shown in blue. The two columns “α-like” and “β-like” indicates the number of amino acids that the sequence has which is the same as human α- and β-tryptase respectively. The “?” in the row MF1 indicates an unknown amino acid as it was not provided in Guida et al.’s paper (16).

Like PT2, the two members of the catalytic triad that can be found within the sequenced segment are present. Only two of the amino acids in this sequence differs from human β-tryptase, however, there are nineteen which differs from MF1 (16). Of the two that differ from the human, neither lies on a surface loop. The only change that is not conserved is the tyrosine to cysteine substitution at residue 234. Conversely, eleven of the nineteen differences with MF1 are not conserved. Five of these lie on surface loops, and one other has been recorded to be important in the structural arrangement of tryptase (8, 9). Since, this sequence is so similar to human β-tryptase despite the length of evolutionary separation between the two species, it was checked against human single nucleotide polymorphism (SNP) databases (16, 31) in case of contamination with human genomic DNA. However, none of the nucleotide differences between this sequence and the human sequence are recorded as SNPs.
Figure 4: Comparison of the amino acid sequences of α- and β-like tryptases from human and non-human primates. Amino acid sequences starting from residue 60B (the beginning of the novel tryptases) were aligned using Invitrogen’s Vector NTI Advance 9.0 AlignX program. Amino acids that are identical in all sequences are shown in yellow, amino acids that appear in over 50% of sequences are shown in blue. Amino acid changes deemed to be similar to the majority are shown in green, and those deemed to be weakly similar are written in purple. Amino acids are numbered from the beginning of the segment, with the respective chymotrypsin residue number in Italics. The purple boxes indicate the regions coding for surface loops, featuring (in order from the beginning) the end of the “60-loop”, the “70- to 80-loop”, the “97-loop”, the “147-loop”, and the “173-loop”. The two members of the catalytic triad (Asp-102 and Ser-195) in this segment are indicated by the arrows, and the “#” indicates residue 216 which has been reported to be crucial in producing the substrate specificity differences seen in α- and β-tryptase (23). The “-” indicates an unknown amino acid, as the MF1 sequence did not continue to the end of the protein (16). Novel tryptases (PT2, MF2, MM1) discussed in this report are underlined.

Rhesus monkey (MM1)

Of the three novel primate tryptase sequences obtained in this study, the rhesus monkey sequence (MM1) (deposited as GenBank accession no. EF_436232) differs the most from human β-tryptase. Like the other sequences, this 988bp (771bp cDNA) sequence translated into a 197 amino acid segment of the mature protein which starts from residue 60B (figure 3C). However, it’s cDNA is only 90.6% and 90.7% identical to human α- and β-tryptase respectively, and 90.4% identical to both α- and β-tryptase at the protein level. MM1 was also found to be 97.8% identical to MF1 and 89.3% identical to MF2 (Tables 2A and B). Unlike the previous two sequences, there is an even number of α-like and β-like fingerprint amino acids, with four of each in the 10 that lie within the sequenced segment (Table 3). Although this sequence appears to be equally similar to α- and β-
tryptase, there is one residue that suggests that this is likely to be an α-like tryptase. Residue 216 has been reported to be important in determining tryptase’s substrate specificity, and is the key difference between α- and β-tryptase (23). Like α-tryptase, MM1 has an aspartic acid instead of a glycine (found in β-tryptase) in this position.

Despite the many differences between human α-tryptase and MM1, the two members of the catalytic triad are present in this sequence. Out of the 197 amino acid sequence, nineteen were found to be different from human α-tryptase, of which thirteen are not conserved changes. Of these, four lie on surface loops. This sequence differs from MF1 by only five of the 186 amino acids in which the two segments overlap. None of these are conserved changes, and only two lie on surface loops. When compared with MF2, there was only 89.3% identity (compared with the 97.8% with MF1), and twenty amino acid differences (see figure 4).

**DISCUSSION**

The majority of known tryptase sequences are from humans or rodents, and surprisingly little attention has been paid to non-human primates. At the time of writing this report, only two published tryptase sequences were from non-human primates – one from the chimpanzee (GenBank accession no. XM_511185), and the other from the cynomolgus monkey (16). Of the six primate genomes screened for this study, we found three novel tryptase sequences, one each from the chimpanzee, cynomolgus monkey and rhesus monkey. The sequences were all partial sequences, covering the region that codes for residue 60B of the mature protein and continues until the c terminus. The number of amino acid differences between these sequences and the respective published sequence suggests that they represent novel genes rather than allelic variants of previously published non-human primate sequences.

Tryptases constitute a highly conserved gene family in mammals (7). This was supported by our data in which all three novel sequences were at least 90% similar to human tryptases despite the fact that macaques diverged from humans ≈25mya (see table 2). Similarly, the exon/intron boundaries were found to be very conserved, with only a couple of base pairs difference in the length of the introns. Moreover, the exons were of the same length except for a 4bp deletion in MM1 that occurred after the stop codon (see figure 3). MF2 was extremely similar to human β-tryptase, only differing in two amino acids. As this sequence may have been the result of contamination with human DNA, we checked it against human SNP databases, and none of the nucleotide differences between MF2 and human β-tryptase were listed as human SNPs. We found it surprising that this
sequence was more similar to human β-tryptase than both chimpanzee sequences (PT1 and PT2) despite the vast difference in evolutionary history, with a 99% identity in the protein sequence compared to 96% and 96.4% respectively. Humans and cynomolgus monkeys diverged ≈25mya, and human and chimpanzees only 4-6mya (see figure 1). Although we do not know the reason behind this, possible reasons for this unexpected finding may be uncovered through further study. As the two macaques are closer to each other in evolution compared to humans, it was not surprising to find that MM1 had a 97.8% identity with MF1, and only five amino acid differences. However, when compared with MF2, there was only 89.3% identity and twenty amino acid differences. Furthermore, although both sequences come from the cynomolgus monkey, there is only 89.8% identity between MF1 and MF2. This supports the hypothesis that MF1 and MF2 are two different types of tryptase, rather than being allelic variants of the same tryptase. MM1 appears to be the same type of tryptase as MF1, both of which seem to be α-like sequences, whereas PT1, PT2, and MF2 appear to be β-like sequences.

To our knowledge, these sequences that we have discovered are all functional. The catalytic triad that is essential for the activity of tryptases appears to be present, with both Asp-102 and Ser-195 unchanged in any sequence (see figure 4). Unfortunately, His-57, the third member of the catalytic triad, does not lie within our partial sequence. Therefore, although it seems likely, we cannot be certain that it remains unchanged in our sequences. Although the sequences were very similar to published primate sequences (including human), there were a number of non-conserved amino acid changes in surface loop regions of the sequences. Of these, a couple are reported to be important in the structural arrangement of tryptase. One of these changes is Tyr95His in PT2, which lies on the 97-loop. This tyrosine has been reported to have a couple of structural roles. Firstly, its aromatic side chain nestles into the bend of the opposing 173 flap. Furthermore, the phenolic group of Tyr-95 of monomer D (along with the endogenous side chain of Gln-98) blocks the S3/S4 subsite on top of the indole moiety of Trp-215 (8, 9). Therefore, the substitution of Tyr-95 to a histidine may affect the structural arrangement, and consequently function, of this chimpanzee tryptase. Another significant change occurs in MM1. Residue 145 in the human (aspartic acid) has been reported to have a role in the structural arrangement of tryptase. Its acidic residue is part of a trio of amino acids which, when arranged around Ile-16, may form a negatively charged anchoring site. This arrangement is believed to destabilise the structured active site of the tryptase monomer (8, 9). The substitution of this amino acid for an asparagine in MM1 may therefore affect the activity of this tryptase. On further study, these differences may prove to change the respective tryptase’s interaction with its environment or its substrate binding cleft in such a way that may render the tryptase non-functional or alter its substrate specificity. In addition, there were also some non-
conserved changes that occurred outside the surface loops, such as the histidine to tyrosine substitution at position 239 in the chimpanzee sequence (PT2). Protein modelling or expression studies would be required to determine the effects of these and other amino acid changes on the functionality of these novel tryptases. A further barrier to our conclusions is that there may be amino acid changes in the unsequenced portion of the protein that prevents or alters enzymatic activity of the tryptase.

To further examine the relationship between these novel tryptases and the relevant published sequences, phylogeny trees were constructed using the program PHYLIP, which constructed trees based on the similarity between the sequences and the calculated amount of steps needed for the amino acid changes to occur (32). Instead of appearing to resemble the evolutionary relationship between the primates as might be expected, the tryptases were arranged into 2 to 3 distinct groups (see figure 5). As expected, the chimpanzee sequences were grouped together, however, the cynomolgus monkey sequences were placed into two different groups – one with β-tryptase, and one with α-tryptase. The cDNA tree (figure 5A) shows three distinct groups of sequences, with human β-tryptase in its own group, the chimpanzee sequences grouped together in another, and the macaques grouped together with human α-tryptase in the third. As there was no nucleotide sequence provided for Guida et al.’s (16) cynomolgus monkey sequence (MF1), it does not appear on the tree. Similarly, the rooted protein tree (figure 5B) show three groups, however, they differ slightly from the cDNA groups. Human β-tryptase is still in its own group and the rhesus monkey sequence is still grouped with human α-tryptase along with MF1. However, MF2 (the novel cynomolgus monkey sequence) was grouped with the chimpanzee sequences. Furthermore, a similar grouping occurred in the unrooted protein tree (figure 5C), which arranged the sequences into two distinct groups, separating the α-like tryptases (human α-tryptase, MF1, and MM1) and the β-like tryptases (human β-tryptase, MF2, PT1, and PT2). In the rooted trees, we found that setting different sequences as the outgroup (the first branch) produced trees with slightly different branch points and groupings, however, regardless of the outgroup set, all unrooted trees show a separation between the α- and β-like sequences. This proves that although small groupings may change, there are definitely two types of tryptases in this set, α- and β-like tryptases. It was very interesting that although MF2 was grouped with the α-like sequences in the cDNA tree, it was grouped with the β-like sequences in both trees featuring protein sequences. This could be due to the differences seen in the identities between this sequence and human α- and β-tryptase (table 2C). In the cDNA, there is a 97% and 98.7% identity with human α- and β-tryptase respectively. However, in the protein sequence, even though its identity with β-tryptase remains at 99%, it only
Novel tryptase sequences from non-human primates

Independent Learning Project: March 2007

has a 94.4% identity with $\alpha$-tryptase. This is due to the fact that the majority of the changes in the cDNA between MF2 and human $\beta$-tryptase occur outside the coding region (after the stop codon), whereas all cDNA differences with human $\alpha$-tryptase occur within the coding region.

Figure 5: Phylogeny tree featuring the novel tryptases and the relevant human and non-primate $\alpha$ and $\beta$-like tryptase. Tree branch points constructed using the programs Protpars and DNApars from the PHYLIP Phylogeny Inference Package for protein and cDNA trees respectively, which use the parsimony methods (32). Data obtained from programs graphically represented using the web server version of Phylodendron. Novel tryptases from this paper are boxed in red. A) Phenogram (a rooted tree with a single root) based on the cDNA sequences. B) Phenogram (a rooted tree with a single root) based on the protein sequences. C) Unrooted tree based on the protein sequences.
Of the five tryptase-like genes in humans (7), the \(\alpha/\beta\) types have been most widely researched. As such, there has been many comparisons made between these two types of tryptase and although the figures do not vary widely, a variety of identities between the two have been reported. Like Guida et al. (16) who reported an identity of 97%, we found that \(\alpha\)- and \(\beta\)-tryptase were 96.1% identical at the nucleotide (cDNA) level (table 2A). However, our data varied greater with the published literature with regards to the protein sequence. Though it has been said that \(\alpha\)- and \(\beta\)-tryptase are \(\approx 90\)% identical at the protein level (1, 19), we found a higher identity of 93.1% (table 2B). This difference is most likely due purely to the alleles used for the calculation.

Apart from humans, the most work that has been done on tryptase has been done in mice, however, the mice tryptase do not group conveniently into the human \(\alpha/\beta\) dichotomy (6). As we have discovered \(\alpha\)- and \(\beta\)-like tryptases in these non-human primates, it indicates that they will be better animal models for investigating human tryptase pathophysiology. In addition, although there has been much discussion in the literature regarding the functionality of \(\alpha\)-tryptase, the discovery of \(\alpha\)-like tryptases in primates relatively distant from humans in evolutionary terms (common ancestor \(\approx 25\)mya) suggests that there is an evolutionary significance in possessing such a tryptase, and therefore, \(\alpha\)-tryptase is most likely enzymatically active.

Although these tryptases have been concluded to be either \(\alpha\)- or \(\beta\)-like on the trees, it is interesting to note that these relationships are not as clearly shown through the identities and fingerprint amino acids. What was determined to be \(\alpha\)-like tryptases (MF1 and MM1) had roughly even numbers of \(\alpha\)- and \(\beta\)-like fingerprint amino acids, with identities that were very similar. In fact, although grouped with the \(\alpha\)-like tryptases, MF1 appears to be closer to \(\beta\)-tryptase in both identities and in the fingerprint amino acids. Conversely, although there were some \(\alpha\)-like amino acids, the \(\beta\)-like tryptases had a clear majority of \(\beta\)-like fingerprint amino acids, as well as identities that clearly suggests a closer relationship with human \(\beta\)-tryptase. This seems to indicate that \(\alpha\)- and \(\beta\)-tryptase are more similar in non-human primates and that they had drifted apart during evolution. In addition, as \(\alpha\)-tryptase has been classed as being an allelic variant of \(\beta\)I-tryptase (20), this mixture of fingerprint amino acids could be due to the evolutionary changes that occur between alleles. Despite these similarities with \(\beta\)-tryptase, what we determined to be \(\alpha\)-like tryptases all have the aspartic acid at residue 216, which has been reported to be singularly responsible for the difference in substrate specificity between \(\alpha\)- and \(\beta\)-tryptase (23). Huang et al. believes this amino acid to lie on one of the surface loops (23). However, the surface loops shown in that paper differ from the surface loops proposed by other papers, and as can be seen in figure 4, residue 216 does not lie on
one of these surface loops. Therefore, although Huang et al. (23) used a site-directed mutagenesis approach to prove their hypothesis, if indeed residue 216 does not lie on a surface loop, the significance of this amino acid on the substrate-binding cleft of tryptase, and subsequently our conclusions regarding the $\alpha$-like sequences, may be questioned.

To fully establish the activity of these novel tryptases, sequencing of the full gene, protein modelling, and protein expression/substrate specificity studies would need to be performed. As a result it would be possible to determine the exact classification (eg $\alpha$- or $\beta$-like tryptase) as well as discover whether these primates would be good models for the investigation of the role of tryptase in the development of inflammatory diseases. We believe that these primate tryptases would prove useful in understanding the role of tryptases in humans as, not only do primates have a close evolutionary relationship with humans, but they also appear to have the same $\alpha/\beta$-tryptase dichotomy.
REFERENCES


