Molecular Characterization of the Murinoglobulins*

(Received for publication, March 18, 1991)

Lut Overbergh, Sophie Torrekens, Fred Van Leuven, and Herman Van den Bergh
From the Center for Human Genetics, University of Leuven, Campus Gasthuisberg, B-3000 Leuven, Belgium

Proteinase inhibitors of the α2-macroglobulin (α2M) type, although well characterized in vitro, still evade a precise description of their actual role in vivo. The main reason for this is the absence of any clinical evidence for the malfunctioning of α2M in humans. Moreover, despite their ubiquitous presence in animals of very different taxa, animal models are notoriously absent in this field. With the advent of transgenic animals an important tool became available in this respect. As a first step in this direction we are analyzing at the molecular level all the members of this proteinase inhibitor family in the mouse. To retrieve related sequences we screened a mouse liver cDNA library with human α2M cDNA. The sequences from two isolated clones partially coded for a protein with a high degree of sequence identity with human α2M, rat α2M, and rat α113. Protein sequence data from the large and small subunits of mouse α2M and of the protein isolated from mouse plasma allowed us to designate the clones as coding for murinoglobulin (MUG), an α2M-related single-chain proteinase inhibitor. Rescreening resulted in the isolation of 24 clones, of which 21 were related or identical to the original MUG clones. Restriction analysis led to three groups of clones of which representative members were sequenced. Two highly homologous cDNA sequences were derived, coding for proteins that displayed the typical features of α2M-type proteinase inhibitors: the overall size, the positions of a putative bait region and of the internal thiol ester, and the positional conservation of cysteine residues and putative asparagine-glycosylation sites. A third related member for which only one incomplete cDNA clone was obtained and sequenced, proved to be aberrant: the bait region contained what appeared to be an intron which escaped proper splicing. A second apparent intron was present at the 3' end of the cDNA while a frameshift mutation near the 5' end (insertion of a G) caused premature termination of the reading frame when compared to the other MUG sequences. These features were confirmed from an isolated genomic clone and extended at the genomic level: the corresponding gene, a transcriptionally weakly active pseudogene, contained the small intron but was spliced into a larger intron. The presence of suitable intron/exon splice sites show that a relatively small part of the intron is being introduced as an exon in the mRNA. Although the resulting reading frame is not open, which makes this mRNA useless, the mechanism as such is highly intriguing since it occurs in the bait region, the least conserved sequence in all members of the α2M family.

The mouse plasma protein, murinoglobulin, is characterized as a single-chain (190 kDa) proteinase inhibitor, belonging to the family of α2-macroglobulin (1, 2). This family of plasma proteins with comparable structural and functional characteristics is present in the circulation of vertebrates and invertebrates but also in the egg white of birds and reptiles (3, 4). As a general mechanism, a proteinase activates the inhibitor by specific proteolysis in the bait region, which, by an unknown mechanism leads to reaction at the cysteiny1-glutamyl internal thiol ester site and to a conformational change, whereby the proteinase is trapped and/or covalently bound to the inhibitor. This binding is mainly steric in the tetrameric forms of the family, with human α2M as the best studied member, but is covalent in the single-chain inhibitors, thereby accentuating their structural homology with complement components C3 and C4 (3). While in the tetrameric proteinase inhibitors steric inhibition is sufficiently strong, monomeric forms need a covalent linkage between the activated glutamyl residue of the original thiol ester and a terminal amino group of a lysine or another nucleophilic group on the proteinase, for inhibition to be effective (5-8).

The thiol ester and the receptor-recognition domain are highly conserved regions within different αM, whereas the bait region is a stretch of amino acids which is unique for every member of the αM-family, containing specific sites for various proteinases. The bait region determines the preference of these inhibitors for different sets of proteinases. Finally, expression of the receptor-binding domain is responsible for rapid elimination of the αM-proteinase complexes (4).

In rat plasma, three different αMs are observed: the constitutive α1M, the acute-phase reactant α2M, and the single-chain α1-inhibitor3 (α113) (8-12). Four different groups of cDNA-clones (13-15) and five groups of genomic clones have been isolated for α113 (16), while in rat plasma only two different forms of α113 were demonstrated (5, 6, 9), corresponding to cDNA clones α113/2J and α113/27J (14). In normal rats α113 is present in the circulation as a monomeric protein of 180 kDa in massive concentrations, up to 14 mg per ml (5, 12, 14), which is the highest value reported for any member of the α2M family. It is a negative acute-phase protein, since its concentration is reduced considerably in

* This investigation was supported by Grant 3.0069.89 from the Fonds voor Geneeskhundig Wetenschappelijk Onderzoek, by a grant of the Interuniversity-network for fundamental research (1987-1991) from the Belgian government, and by a grant from the Leuvenlijn-aktie Nationaal Fonds voor Wetenschappelijk Onderzoek (Belgium). The costs of publication of this article were defrayed in part by the to the GenBank®/EMBL Data Bank with accession number(s) M65736, M65298, and M65237.

1 The abbreviations used are: α2M, α2-macroglobulin; α113, α1-inhibitor 3; MUG, murinoglobulin; αM, α-macroglobulin; α1M, α1-macroglobulin; SDS, sodium dodecyl sulfate; bp, base pairs; kb, kilobase(s); PCR, polymerase chain reaction.
acute and chronic inflammation as well as in response to tissue injury (9, 14).

In the mouse the murinoglobulin originally described (1) is a single-chain proteinase inhibitor, thought of as the murine equivalent of α1Ⅲ in the rat. In the absence of molecular data this needs to be demonstrated. In a study aimed at the molecular cloning of all members of the αM family in mice we have isolated several different cDNA clones. In this report we describe the isolation and complete sequencing of two murinoglobulins at the cDNA level. Characterization of a third, partial cDNA and of part of the corresponding gene proved it to be derived from a pseudogene. Only one murinoglobulin was isolated from mouse plasma, the NH₂-terminal sequence of which corresponded to the one predicted by translation of the most prominent cDNA.

MATERIALS AND METHODS

Materials

Libraries used were: adult male (C57/Black6) liver cDNA in λGT11 (Clontech), adult female (B6CBA/F1J) liver cDNA in XZap (Stratagene), genomic DNA in λGEM11 from the liver of an adult mouse (ICR Swiss; Promega). Human α2M cDNA was kindly provided by G. Fey (La Jolla, CA).

Methods

Screening—λ phage were plated on the appropriate Escherichia coli hosts (Y1090 for λGT11, B54 for XZap, and KW251 for λGEM11) and grown on NZYD-T agar substrate at 37 °C until plaques were of appropriate size. Replica’s were made on nylon filters (Du Pont Colony Screen) and hybridized with the probes indicated. The cDNA probes were isolated inserts in low melting agarose, labeled by hexanucleotide mediated incorporation of [³²P]dCTP (17-18). Oligonucleotides were endlabeled with [³²P]dATP or with digoxigenin-dUTP (Boehringer Mannheim).

Subcloning of λGT11 and λGEM11 Inserts—λ inserts containing the putative inserts were excised with EcoRI (for λGT11 inserts) or ScaI (for λGEM11 inserts) and ligated into pUC18.

Excision and Rescue of XZap Inserts—XZap DNA containing the putative murinoglobulin cDNA inserts were excised by the automatic excision process as described by the manufacturer. The colonies, containing the rescued pBluescript plasmid with the cloned cDNA insert, were processed by standard procedures (19, 20).

DNA Sequencing—General methods of DNA manipulation were according to Maniatis (19) and Davis (20). Purified DNA was sequenced on both strands by the dideoxy chain-termination method of Sanger (21) (Sequenase, U. S. Biochemicals). Deletion clones were made with different restriction enzymes. Remaining gaps were sequenced with internal oligonucleotides as primers, synthesized on either Millipore or ABI equipment according to standard procedures made with different restriction enzymes. Remaining gaps were sequenced with internal synthetic oligonucleotides. The two clones overlapped 2081 nucleotides of identical sequence while combined they resulted in a sequence of 3532 nucleotides, including a terminal amino acid sequence of the 35-kDa subunit of mouse α2M (bases 1128–2132) (C), or with a 5’-terminal 889-bp restriction fragment of clone LZ37 (D). Restriction digests were done with the following enzymes: E, EcoRI; N, Ncol; S, SacI; EV, EcoRV; and H, HindIII.

Not all the restriction sites of the tested enzymes are shown, but only these sites missing in one of the classes of murinoglobulins. Open circles denote the absence of the restriction site.

RESULTS

Isolation of Murinoglobulin cDNA Clones—A mouse liver cDNA library in λGT11 was screened with a partial cDNA probe from human α2M (24). The inserts of two positive

plagues, designated 15 (2.6 kb) and 20 (2.9 kb), were subcloned in pUC18. Both clones were sequenced by the dideoxy chain-termination method, after preparation of deletion clones with suitable restriction enzymes. Remaining gaps were sequenced with internal synthetic oligonucleotides. The two clones overlapped 2081 nucleotides of identical sequence while combined they resulted in a sequence of 3532 nucleotides, including a 29-nucleotide poly(A) tail (Fig. 1A). The derived amino acid sequence showed considerable differences with the NH₂-terminal amino acid sequence of the 35-kDa subunit of mouse α2M.² Comparison with the published cDNA sequences of human α2M, rat α2M, and especially rat α1Ⅲ suggested that the clones represented murinoglobulin.

Screening of a second mouse liver cDNA library (in λZap) using clone 15 as a probe, produced 24 different positive clones. Characterization by restriction mapping with the enzymes EcoRI, SacI, Ncol, HindIII, and EcoRV revealed four different groups of clones. One group of three clones were shown to correspond to mouse α2M. Of the other, a large group of 17 clones showed the same restriction pattern as the original clone 15, which we designated murinoglobulin 1 (MUG1). Three clones displayed a different restriction pattern (missing one EcoRI, one SacI, two Ncol, and one EcoRV restriction site, Fig. 1B) and were classified as MUG2. Only

² F. Van Leuven and S. Torrekens, unpublished results.
one clone, LZ52, displaying a third restriction pattern, (missing two SacI sites, one HindIII, and one EcoRV site compared to MUG1, Fig. 1B) was isolated and designated MUG3.

Screening of the same cDNA library with a restriction fragment of human α2M revealed 36 positives, of which two corresponded to MUG1 (Fig. 1C). Another screening for more full-length MUG clones was done with a 5'-terminal 889-bp restriction fragment of LZ37 (MUG2). 10 positive clones were obtained, seven corresponding to MUG1 and three to MUG2 (Fig. 1D). The NcoI restriction site is remarkable. It was present in LZ1 at position 488–493, but missing in the seven MUG1 clones of the third screening. The presence of this NcoI restriction recognition site in LZ1 was confirmed by sequencing. In LZ3/5, LZ3/6, and LZ3/10 nucleotide sequencing showed a cytosine to thymine point mutation at position 489, causing the disappearance of the recognition sequence for NcoI. This polymorphism for MUG1 indicates at least a two-allele system.

Sequencing of Murinoglobulin cDNA Clones—The complete sequence of MUG1 cDNA was obtained by combining three partially overlapping clones, 15, 20, and LZ1, with a total length of 4626 base pairs (Figs. 2 and 3). For MUG2 sequence data from two clones, LZ23 and LZ37 were combined to obtain a full-length sequence of 4566 base pairs (Figs. 2 and 4). Clone LZ23 was sequenced completely. Combined with sequencing of the 5' and 3'-terminal fragments of LZ37, which is a full-length clone of 4.5 kb, this yielded the complete MUG2 sequence. To ascertain that both clones belonged to the same variant murinoglobulin, the bait region of clone LZ37 and the thiol ester domain, which is typical as a consequence of a deletion of 75 nucleotides relative to MUG1, were also sequenced, to confirm their identity with the corresponding sequences from LZ23.

The one clone, LZ52, representing MUG3 was sequenced completely (Figs. 2 and 5).

In all cases we employed the same sequencing strategy described above, which combines the preparation of deletion clones after subcloning of the A inserts into pUC18 (for λGT11 inserts) or after rescue into pBluescript plasmids (for λZap inserts) and the synthesis of specific internal oligonucleotides as sequence primers. Since the different murinoglobulins are highly homologous, many primers could be used for the different variants.

Analysis of MUG1 and MUG2 cDNA Sequences—MUG1 cDNA contained an open reading frame of 4428 nucleotides, whereas in MUG2 only 4353 nucleotides were in-frame. An ATG start codon was sequenced in both variants and part of the 5'-nontranslated region was obtained (39 nucleotides for MUG1 and 64 for MUG2). In the 3'-untranslated region a canonical AATAAA polyadenylation signal is present 15 nucleotides upstream from the poly(A) tail in MUG1, whereas only 13 nucleotides separates them in MUG2 (Figs. 3 and 4).

At the nucleotide level a sequence identity of 93% is observed between the two full-length murinoglobulin variants. Compared to rat α13 an overall sequence identity of 77% is seen with the prototype α13/2J, whereas only an identity of 50% is observed with mouse α2M.

Translation of the nucleotide sequences in open reading frames of 1476 (MUG1) and 1451 (MUG2) codons (Fig. 3 and 4) yields an overall amino acid sequence identity of 89%. The lowest level of amino acid sequence identity, only 56%, is observed in the putative bait region (position 682–734 in MUG1). In other regions the homology is almost complete, which is most notable at both the NH2- and COOH-terminal ends. A rather large deletion of 25 amino acids is typical for MUG2 (corresponding to residues 1094–1120 in MUG1).

The predicted sequence of MUG1 contains 26 cysteine residues, one more than MUG2. In both murinoglobulin variants 24 of these residues are positionally conserved. In variant MUG2 the cysteine residues 402 and 437 are missing, while an extra cysteine residue is present at position 1273 (Figs. 3, 4, and 6). From the derived amino acid sequence the calculated mass of the mature proteins is 165061 for MUG1 and 162343 for MUG2. The predicted potential asparagine-glycosylation sites amount to 11 in MUG1 and only nine in MUG2; of those, eight are identically positioned in both MUG variants (Figs. 3, 4, and 7). Compared to human α2M (24), for which disulfide bonds and glycosylation sites were experimentally determined (25), the homology is striking (Figs. 6 and 7).

Expression of mRNA—Despite direct detection by Northern blotting of mRNA species of about 5 kb extracted from mouse liver and other tissues, we could not conclusively show the differential presence of specific mRNA of the different MUG variants. Cross-reaction with the abundant mouse α2M mRNA and with the other variants prevented unequivocal results. Therefore, we attempted to obtain this information by PCR amplification. Total RNA, isolated from mouse liver and also from uterus, was reverse-translated, and the bait regions of different MUG variants were amplified by the PCR method. Different sets of oligonucleotide primers were devised based on the nucleotide sequence information of the bait regions which are the most typical and diverse regions of these proteins. The results supported the evidence that the three different MUG variants were expressed in the liver, which substantiates their isolation from the cDNA library prepared from this tissue. Signals corresponding to the three different types of cDNA isolated were also found by hybridization of the PCR-amplified fragments with three different oligonucleotide primers, typical and specific for each of the bait regions of the three MUG variants. In all cases the variant MUG1 was present as the predominant species. Additionally, MUG2 and MUG3 mRNA was present in both tissues but in much lower concentration than MUG1 mRNA.

Plasma Murinoglobulin: Protein Isolation and Sequencing—Murinoglobulin was isolated from fresh, heparinized mouse plasma obtained by cardiac puncture from anesthetised animals. Isolation was done by a simple procedure which includes polyethylene glycol precipitation, gel filtration, and hydrophobic interaction chromatography. A single protein species

\[ F. \text{Van Leuven and L. Overbergh, unpublished observations.} \]
was obtained as judged from analytical gel filtration, rate electrophoresis in 8% native gels, SDS-polyacrylamide gel electrophoresis on gradient gels, and by NH2-terminal sequencing. By these criteria we isolated a protein of apparent molecular mass of 180 kDa, initially thought of as a second MUG variant. Thus the two highly different regions were actually intron sequences, was confirmed by the isolation of a native protein was estimated at about 100 pmol of approximately 100 pmol of protein.

Further analysis is necessary to establish the equivalence of glutamine or histidine at position 1-255. No sequence homology was retrieved from the EMBL or GenBank sequence databases consulted. Region 624-798, located in the putative bait region, indicated the possibility that this also might be an intron. This hypothesis, that the two highly different regions were actually intron sequences, was confirmed by the isolation of a genomic clone coding for this cDNA. This genomic clone of

**Analysis of the Aberrant Murinoglobulin**—The cDNA sequence obtained from the partial clone designated MUG3 contained 3316 base pairs including, at the 3' end, both a poly(A) tail and a typical AATAAA canonical polyadenylation signal (Fig. 5). The sequence identity with both MUG variants is high but not complete. A frameshift mutation is obviously present at position 1461 (compare to MUG1, Fig. 3 position 2763). This induces a stop of translation in the next codon, disturbing an open reading frame, which prohibits efficient translation.
presumed intron at the 5' end of the MUG3 cDNA (Fig. 10). Moreover, the intron in the bait region is observed, but the function of a protein. Of further considerable importance is the possibility that a particular intron is buried in a larger intron. This was demonstrated in a recent study of mouse a2M, but also murinoglobulin, the single-chain inhibitor of the a2M family. Since in humans no clinical studies have been carried out, animal models have been developed to study these inhibitors. In the tech-

In conclusion, the MUG3 cDNA was sequenced using synthetic oligonucleotides and the cDNA size was determined by PCR amplification using the same primers. The sequence analysis confirmed the presence of the putative intron region and provided clues to the functioning of these inhibitors we are interested in. From this perspective we have initiated studies on the murinoglobulins, which are the physiological behavior of murinoglobulins, and the a2M, which is not known. The role of proteinase inhibitors in the a2M family, which is also known as murinoglobulin, was studied. In this context, the a2M family is of particular interest because in humans no clinical studies have been carried out. Animal models have been developed to study these inhibitors, which has provided useful insights into the physiological behavior of murinoglobulins. In the technical aspects of genetic manipulation have culminated into a tool for the characterization at the molecular level of all the members of this family of proteinase inhibitors, not only mouse a2M but also murinoglobulin, the single-chain inhibitor previously described (1, 2).
smaller and not identically positioned deletion observed in ulins. Unequivocal identification as murinoglobulins was mouse a2M and by comparison with the published cDNA sequence LG1/6 are.

The differences between variants MUG1, MUG2, and the same cysteine residues are also conserved in the prototype rat All3/2J (14). Including the cysteine residue present in the thiol ester site, disulfide bonds in MUG1, and MUG2 are.

FIG. 5. Partial cDNA sequence and derived amino acid sequence of MUG3 (clone L252). The two introns are represented in lower case. The boxed guanine at position 1461 denotes an insertion.

By screening a cDNA mouse liver library with a human a2M probe, we isolated not only mouse a2M clones (which is reported elsewhere) but also different variants of murinoglobulins. Unequivocal identification as murinoglobulins was based upon NH2-terminal sequencing of a murinoglobulin isolated from plasma and of the large and small subunits of mouse a2M and by comparison with the published cDNA sequence of a13, the single chain inhibitor in the rat (14, 15).

The differences between variants MUG1, MUG2, and the aberrant MUG3 are located predominantly in the bait region. A deletion of 75 nucleotides is present in the thiol ester domain of MUG2, which is reminiscent of a similar but smaller and not identically positioned deletion observed in one of the variants of rat a13 (14). Point mutations (base substitutions, deletions, and insertions) are present throughout the whole sequence, suggesting that the different cDNAs did not originate by alternative splicing of mRNA derived from one murinoglobulin gene. The more acceptable hypothesis is the existence of different genes.

FIG. 6. Location of cysteine residues. Top, disulfide bonds in human a2M (3); dotted lines show two cysteine residues involved in interchain disulfide bonds. Positions of cysteine residues in rat a13/2J are published (14). Positions of cysteine residues in MUG1 and MUG2 were obtained from translation of the cDNA sequences. TES denotes the thiol ester site.

FIG. 7. Location of putative NH2-linked glycosylation sites. The potential glycosylation sites are defined as Asn-X-Ser or Asn-X-Thr. Human a2M (25), rat a13/2J (14), MUG1, and MUG2 are compared.

substitutions, deletions, and insertions) are present throughout the whole sequence, suggesting that the different cDNAs did not originate by alternative splicing of mRNA derived from one murinoglobulin gene. The more acceptable hypothesis is the existence of different genes.

Southern blot analysis of mouse genomic DNA with a 350-bp-amplified fragment of the putative bait region, known to react with all three MUG cDNA, indicated the existence of different MUG genes. Restriction of the genomic DNA with BamHI, EcoRI, and EcoRV revealed in each case four different bands, totaling between 25 and 30 kb. The existence of at least three different genes for murinoglobulins appears thereby very likely.

That MUG1 and MUG2 are members of the aM-family is further evidenced by the presence of a typical thiol ester site, which is involved in covalent binding of the proteinase. In monomeric aM this site is an absolute necessity for the further evidenced by the presence of a typical thiol ester site, thereby very likely.

That MUG1 and MUG2 are members of the aM-family is further evidenced by the presence of a typical thiol ester site, which is involved in covalent binding of the proteinase. In monomeric aM this site is an absolute necessity for the further evidenced by the presence of a typical thiol ester site, thereby very likely.
Murinoglobulins

FIG. 8. SDS-polyacrylamide gel electrophoresis of mouse plasma and of murinoglobulin and mouse a2M. Lane 1, markers with M, as indicated; lane 2, purified murinoglobulin, 5 µg; lane 3, mouse plasma, 1.5 µl; lane 4, purified mouse a2M, 7.5 µg. Proteins were treated in Tris buffer (50 mM, pH 8.0) with 1% SDS, 1% 2-mercaptoethanol, for 5 min at 98 °C. Analysis was on a polyacrylamide gradient gel (6–20%).

FIG. 9. Comparison of the signal peptide of MUG1 and MUG2 with the NH2-terminal sequences of the purified MUG protein. The signal peptide and NH2-terminal amino acid sequence of MUG1 and MUG2 from translated cDNA are compared with the NH2-terminal protein sequence of the purified MUG. Identical residues are indicated by colons. Residues which were not confidently assigned are given in lower case.

The two variant murinoglobulins differ most extensively in their putative bait regions, where an amino acid sequence identity of only 56% is observed, as opposed to an overall protein sequence identity of 89%. It has been proposed that the divergence in this region, which is a general feature of the aM family as observed in human a2M and pregnancy zone protein; rat a2M, a1M, and two variants of a113; and in mouse a2M and the MUG variants, could be due to positive Darwinian selection, thereby creating new regions for the attacking proteinases as postulated in other systems (30–32).

Another explanation for divergence in this region is the neutralist theory, suggesting that the bait region is located in a loop which is not of crucial importance for the protein structure, so that mutations in this region are not selected against, as opposed to other parts of the protein (14, 30). When comparing the inhibitor characteristics of mouse a2M and MUG (1, 2), as well as the two variants of a113 (6, 30), essentially the same inhibitory patterns are found. These findings make the second theory more acceptable. A combination of both theories is of course possible and must await verification by experiments.

For the third aberrant MUG no full-length, normal translatable transcript was obtained from two liver cDNA libraries. Isolation of a genomic clone corresponding to the gene responsible for the aberrant cDNA sequence of MUG3 did not
conclusively settle the question as to the existence of a normal counterpart. The frameshift mutation which was found in the cDNA clone at position 1461, is also present in this genomic clone, located at an intron-exon boundary. The “intron” sequence (175 bp) which was present in the bait region of the aberrant cDNA clone and which we defined by comparison to the other MUG variants, was demonstrated to be present in the genomic clone as part of a much larger intron of 2.2 kb. The intron sequence is thus to be regarded as an exon, though it shows no open reading frame. These results suggest that the third aberrant MUG is derived from a transcriptionally active pseudogene which was partially isolated as the genomic clone. It is impossible to decide at this point in time if this pseudogene is complete and derived from the gene of murinoglobulins 1 or 2 or from a third active murinoglobulin gene for which we have no cDNA evidence yet. Analysis of the putative bait region of MUG3, leaving out the intron, and comparison to the bait regions of the “active” murinoglobulins 1 and 2, reveals the following (Fig. 11): of 46 comparable residues, 19 residues (41%) are positionally conserved in all three MUGs (including 5 prolines which are also seen in the same positions in the rat α113 variants), while 13 residues (28%) are uniquely present in MUG3. Of the remaining 14 residues, 5 amino acids (11%) are shared between MUG1 and MUG3, while 9 (20%) are common between MUG3 and MUG2. Although such considerations do not take into account the chemical similarities of different amino acids nor relative importance in terms of protease substrate specificity, they do demonstrate conclusively that MUG3 is not significantly related to any of the other murinoglobulins. This would mean that a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third active gene, still to be discovered, is encoding a third activ...