Evolution of Mammalian KELL Blood Group Glycoproteins and Genes (KEL): Evidence for a Marsupial Origin from an Ancestral M13 Type II Endopeptidase Gene

Roger S Holmes*
Eskitis Institute for Drug Discovery and School of Biomolecular and Physical Sciences, Griffith University, Nathan, QLD, Australia

Abstract

KELL is a member of the M13 family of type II neutral endopeptidases, which functions as a blood group antigen in human and animal populations. KELL amino acid sequences and structures and KEL gene locations were examined using bioinformatic data from several mammalian genome projects. Mammalian KELL sequences shared 55-99% identity, as compared with 21-31% sequence identities with other M13-like family members. Four predicted N-glycosylation sites were conserved among the mammalian KELL proteins examined. Sequence alignments, key amino acid residues and conserved predicted secondary and tertiary structures were also studied, including active site residues, predicted disulfide forming Cys residues, cytoplasmic, transmembrane and extracellular sequences and KELL C-terminus amino acid sequences. Mammalian KEL genes usually contained 18 or 19 coding exons on the direct strand. Transcription factor binding sites within the human KEL promoter may regulate transcription within erythroid cells. Phylogenetic analyses examined the relationships and potential evolutionary origins of the mammalian KEL gene with six other vertebrate neutral endopeptidase M13 family genes. These suggested that KEL originated in an ancestral marsupial genome from a gene duplication event of a neutral endopeptidase M13-like gene.

Keywords: KELL glycoproteins; Blood group antigens; KEL; Sequence conservation; Marsupial origin; Gene evolution

Abbreviations: KELL: Blood Group Glycoprotein; KEL: KELL Blood Group Glycoprotein Gene; MME: Membrane Metallo-Endopeptidase; ECE: Endothelin-Converting Enzyme; PHEX: Phosphate Regulating Neutral Endopeptidase; ECELL1: ECE-Like Protein 1; RBC: Red Blood Cell; XY protein: A Multipass Membrane Protein; QTL: Quantitative Trait Locus; BLAST: Basic Local Alignment Search Tool; BLAT: Blast-Like Alignment Tool; NCBI: National Center for Biotechnology Information; KO: Knock Out; AceView: NCBI Based Representation of Public mRNAs; SWISS-MODEL: Automated Protein Structure Homology-Modeling Server; UTR: Untranslated Gene Region

Introduction

KELL blood group glycoprotein (KELL, also called CD antigen 238; EC 3.4.24.-) is one of at least seven members of the M13 family of neutral endopeptidases, which are zinc-containing type II transmembrane enzymes [1-4]. KELL blood group glycoprotein contains several antigens that are highly immunogenic and serve as the third most effective system in triggering an immune reaction, after the ABO and Rh blood groups [5-7]. KELL is also a single-pass transmembrane protein which is linked to the XY protein, found in red blood cell (RBC) membranes [8,9], and serves as an endothelin-converting enzyme, an endopeptidase which cleaves 'big' endothelin-3 to form an active vasoconstrictor peptide [10,11].

Other M13 neutral endopeptidases have been described: Membrane endopeptidase (MME) or nephrilysin (NEP) inactivates signaling peptides involved in regulating blood pressure, the immune system and neuronal activity [12-14]; membrane metallo-endopeptidase-like 1 (MMEL1) or nephrilysin-like protein 1 (NEPL1), reported as a susceptibility locus for multiple sclerosis, primary biliary cirrhosis and rheumatoid arthritis with a proposed sperm function role [15-18]; endothelin-converting enzyme 1 (ECE1, EC=3.4.24.71) [19,20], and endothelin-converting enzyme 2 (ECE2, EC=3.4.24.71) participate in regulatory peptide processing [21,22]; endothelin-converting enzyme-like 1 (ECELL1) serves an essential role in the nervous control of respiration [23,24]; and phosphate-regulating neutral endopeptidase (PHEX), which is involved in bone mineralization, and has a proposed role in renal phosphate reabsorption [25,26].

The gene encoding KELL (KEL in humans; KEL in mice) is highly expressed in erythroid tissues, but also in other tissues, including testis, heart, spleen and skeletal muscle [3,9]. The structures of the human KEL and the mouse KEL genes have been reported, containing 18 or 19 exons of DNA encoding KELL sequences [9,10,27,28]. The molecular basis of the major human KEL antigens has been determined, which result from KEL point mutations and single amino acid substitutions [29]. At least 30 KELL antigens are reported [27,30,31], including the highest frequency KEL polymorphism in human populations, designated as K/E, due to a C→T substitution in exon 6 causing a 193Thr to 193Met substitution and disruption of an N-glycosylation site [32]. Other high incidence KELL antigens have been reported, including KALT, which is sensitive to treatment of RBCs by trypsin [27,33]. A ‘null’ human KEL phenotype (designated as K0), which abolishes KELL expression in erythroid tissues, resulting from a mutation at the splicing exon 3 donor site [33], while other K alleles have been reported in low frequency in a Chinese population [34]. Knock-out (KEL-Kel-) (KO) mice lacking RBC KELL glycoprotein exhibited changes in red cell ion transport and some mild motor dysfunction, have provided a useful model to study the biochemical and physiological roles for this protein [31]. RBC Gardos channel activity, which normally functions as a potassium chloride cotransport and calcium-activated potassium channel, and assists with maintaining RBC hydration status [34], was increased in KO erythrocytes, although lacking endothelin-converting endopeptidase activity. Antibodies generated in the circulation system in response to KELL antigens are usually immunoglobulin G, which are credited.

*Corresponding author: Roger S Holmes, Eskitis Institute for Drug Discovery, School of Biomolecular and Physical Sciences, Griffith University, Nathan, QLD 4111, Australia, Tel: +61 7 37356008; E-mail: r.holmes@griffith.edu.au

Received April 14, 2013; Accepted July 17, 2013; Published July 22, 2013

Citation: Holmes RS (2013) Evolution of Mammalian KELL Blood Group Glycoproteins and Genes (KEL): Evidence for a Marsupial Origin from an Ancestral M13 Type II Endopeptidase Gene. J Phylogen Evolution Biol 1: 112. doi:10.4172/2329-9002.1000112

Copyright: © 2013 Holmes RS. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
may result in severe hemolytic transfusion reactions and hemolytic disease of the fetus and newborn [5,7]. Recent studies have suggested that these reactions may occur not only as a result of immune RBC destruction, but also by the suppression of erythropoiesis by anti-KELL-K antibodies, which can lead to severe anemia in the fetus or newborn [35,36].

This paper reports the predicted gene structures and amino acid sequences for several mammalian KELL genes and proteins, the predicted structures for mammalian KELL proteins and the structural, phylogenetic and evolutionary relationships for these genes and enzymes with those for six other vertebrate M13 neutral Type II endopeptidase gene families. The results suggest that the mammalian KELL gene arose from the duplication event of an ancestral mammalian M13 Type II-like endopeptidase gene, with the appearance corresponding to the emergence of marsupial mammals during evolution.

Methods

Mammalian KEL gene and KELL protein identification

BLAST studies were undertaken using web tools from the National Center for Biotechnology Information (NCBI) [37]. Protein BLAST analyses used mammalian KELL and six other vertebrate M13 neutral endopeptidase amino acid sequences previously described [1,2,9,23,38-43] (Table 1 and 2).

BLAT analyses were subsequently undertaken for each of the predicted KELL amino acid sequences and other M13 neutral endopeptidase-like genes, using the UC Santa Cruz Genome Browser with the default settings to obtain the predicted locations for each of these vertebrate genes, including predicted exon boundary locations and gene sizes (Table 1 and 2) [44]. The AceView website was used to obtain structures for the major human KEL and mouse KEL transcripts [28].

Predicted structures and properties of mammalian KELL proteins

Predicted secondary and tertiary structures for human and other mammalian KELL proteins were obtained using the SWISS-MODEL web-server [43], and the reported tertiary structure for human ECE1 complexed with phosphoramidon [20] (PDB:3dwbA), with a modeling residue range of 60-713 for human KELL. Molecular weights, N-glycosylation sites and predicted transmembrane [46], cytosolic and extracellular sequences for mammalian KELL and related vertebrate M13 Type II endopeptidase proteins were obtained using Expasy web tools (http://au.expasy.org/tools/pi_tool.html). Identification of conserved domains for mammalian KELL proteins was made using NCBI web tools [47] (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi).

Comparative human (KEL) gene expression

The genome browser (http://genome.ucsc.edu) was used to examine GNF Expression Atlas 2 data, using U133A and GNF1H expression chips for the human KEL gene (http://biogps.gnf.org) [48]. Gene array expression 'heat maps' were examined for comparative gene expression levels among human and mouse tissues, showing high (red); intermediate (black); and low (green) expression levels.

Phylogeny studies and sequence divergence

Phylogenetic analyses were undertaken using the http://phylogeny.fr platform [49]. Alignments of mammalian KELL sequences, six other vertebrate M13 neutral endopeptidase sequences and a nematode (Caenorhabditis elegans) M13 endopeptidase-like sequence were assembled using PMUSCLE [50] (Table 1 and 2). Alignment ambiguous regions were excluded prior to phylogenetic analysis, yielding alignments for mammalian KELL sequences with other vertebrate neutral endopeptidase and nematode (Caenorhabditis elegans) M13-like sequences. The phylogenetic tree was constructed using the maximum likelihood tree estimation program PHYML [51].

Table 1: Mammalian KEL and human ECE1 and MME genes and proteins.

| Gene KELL | Species | RefSeq ID | Ensembl/BIN | GenBank ID | UNIPROT ID | Amino Acids | Chromosome location | Coding Exons (strand) | Gene Size bps | Subunit NN | Gene Expression | % Identity human KELL | % Identity human ECE1 | % Identity human MME |
|-----------|---------|-----------|-------------|------------|-------------|-------------|---------------------|--------------------|---------------|-------------|-----------------|----------------------|----------------------|----------------------|
| Human (KELa) Homo sapiens NM_000420 BC003135 P23276 732 7:142,638,342-142,658,569 19 (-ve) 20,618 82,824 0.5 8.1 100 30 21 |
| Human (KELb) Homo sapiens na CH471199 na 713 7:142,638,342-142,659,588 18 (-ve) 20,247 80,793 0.5 8.1 100 30 21 |
| Chim Pan troglodytes XP_519445 na na 732 7:144,486,459-144,507,070 19 (-ve) 20,612 82,864 0.8 5.9 74 31 22 |
| Gorila Gorilla gorilla XP_004064634 na na 713 7:141,432,380-141,452,992 19 (-ve) 20,613 82,725 0.9 11 30 21 |
| Gibbon Nomascus leucogenys XP_003270898 na na 732 7:143,328,325-330,342 18 (+ve) 20,618 82,867 1 21 |
| Rhesus Macaca mulatta XP_003920065 EA417199 na 713 3:182,467,326-182,487,814 18 (-ve) 20,487 80,867 0.7 93 31 23 |
| Equin Horse Equus caballus XP_004487572 na na 713 4:69,015,434-69,034,243 18 (+ve) 19,810 90,732 4 30 |
| Cow Bos taurus XP_001780561 na na 721 4:109,587,066-109,625,396 18 (-ve) 18,610 81,503 0.5 7.1 70 31 22 |
| Rabbit Oryctolagus cuniculus XP_002712030 na na 784 7:195,751-619,196,396 19 (+ve) 20,978 88,047 0.8 7.3 79 29 22 |
| Pandalus Alkogoldea melanoleuca XP_003294225 na na 738 *191114,349,706-368,327 19 (+ve) 16,622 83,365 0.7 79 31 22 |
| Opus eos Meriones domesticus XP_001346828 na na 768 8:205,592,264-205,612,188 18 (ve) 19,925 86,352 0.5 6.3 55 27 21 |

MME

Human Homo sapiens NM_000902 BC101632 P08473 750 3:184,801,957-154,869,245 22 (ve) 56,289 85,514 0.5 5.5 21 36 100 |

ECE1

Human Homo sapiens NM_048826 BC050453 O95672 775 2:23,344,866-233,351,363 17 (ve) 6,498 87,791 0.4 6.6 30 100 36 |

RefSeq: the reference amino acid sequence; *predicted Ensembl amino acid sequence; na-not available; GenBank IDs are derived NCBI http://www.ncbi.nlm.nih.gov/GenBank/; Ensembl ID was derived from Ensembl genome database http://www.ensembl.org; UNIPROT refers to UniprotKB/Swiss-Prot IDs for individual endopeptidase-like proteins (see http://kr.expasy.org); "refers to an unknown scaffold; bps refers to base pairs of nucleotide sequences; pI refers to theoretical isoelectric points; the number of coding exons are listed; high % identities are shown in bold; high gene expression values are in bold; (a) and (b) refer to isoform sequences for human KELL.
Results and Discussion

Alignments of mammalian KELL amino acid sequences

The deduced amino acid sequences for gibbon (Nomascus leucogenys), horse (Equus caballus) and rat (Rattus norvegicus) KELL proteins are shown in Figure 1, together with previously reported sequences for human [2] and mouse KELL proteins [9] (Table 1). Alignments of human with other mammalian sequences examined were between 55-98% identical, suggesting that these are members of the same gene family, whereas comparisons of sequence identities of mammalian KELL proteins with human ECE1 and MME proteins exhibited lower levels of sequence identities: 27-31% and 19-22%, respectively, indicating that these are members of distinct M13 Type II endopeptidase-like gene families (Table 1).

The amino acid sequences for mammalian KELL proteins contained between 713 for human KELL isoform 'b' and 784 for rabbit KELL residues, whereas most mammalian KELL sequences contained either 732 amino acids (for human KELL isoform 'a') or 713 residues (for human KELL isoform 'b') (Figure 1; Table 1). Previous studies have reported several key regions and residues for human and mouse KELL proteins (human isoform 'b' KELL amino acid residues were identified in each case). These included an N-terminus cytoplasmic tail (residues 1-28), followed by a hydrophobic transmembrane twenty residue segment, which may anchor the enzyme to the plasma membrane [2]. These N-terminal cytoplasmic tail and transmembrane regions revealed a high degree of amino acid sequence conservation (Figure 1). This region was further characterized by conserved 'book-end' residues for mammalian KELL, Arg27-Arg/Trp28 at the N-terminal end, and Tyr48 at the C-terminal end of the membrane anchoring segment, which may contribute to the membrane spanning properties. Figure 2 provides an alignment of the cytoplasmic N-terminal sequences for several mammalian KELL sequences, including human KELL 'a' and 'b' isoform sequences. These contained identical transmembrane and endopeptidase domains, but with N-terminal cytoplasmic tails of different lengths (residues 1-47 and 1-24 for the human 'a' and 'b' isoforms, respectively). This explains the differences in amino acid composition for these human KELL proteins (732 and 713 amino acids, respectively). Figure 2 also shows N-terminal cytoplasmic tail and transmembrane sequences for other mammalian cytoplasmic domain KELL sequences, including rabbit KELL, which contained an extended N-terminal cytoplasmic domain sequence of 110 amino acid residues.

Table 2: Vertebrate ECE1, ECE1, ECE2, MME, MMEL1 and PHEX genes and proteins.

| Gene KELL | Species                   | RefSeq ID    | GenBank ID | UNIPROT ID | Amino Acids | Chromosome location | Coding Exons (strand) | Gene Size bps | Subunit MW | Gene Expression | pl |
|-----------|--------------------------|--------------|------------|------------|-------------|---------------------|----------------------|---------------|------------|----------------|----|
| Human     | Homo sapiens             | NM_00113347  | BC172256   | P42892     | 770         | 1:21,546,451-21,616,907 | 19 (ve)              | 70,457       | 87,163     | 4.4           | 5.6|
| Mouse     | Mus musculus             | NM_199007    | AB036648   | G4FZ2X     | 769         | 4:173,469,640-173,518,818 | 19 (ve)              | 49,178       | 87,085     | 3.7           | 5.6|
| Chicken   | Gallus gallus            | NM_204717    | AF69287    | Q6DGN6     | 752         | 21:6,600,258-6,607,612  | 17 (ve)              | 7,395        | 84,988     | na           | 5.1|
| Zebrafish | Danio rerio              | NP_001071260 | BC125652   | F1RAS8     | 752         | 11:28,797,446-11,280,975 | 18 (ve)              | 32,530       | 85,206     | na           | 5.8|

Table 2: Vertebrate ECE1, ECE1, ECE2, MME, MMEL1 and PHEX genes and proteins.

RefSeq: the reference amino acid sequence; *predicted Ensembl amino acid sequence; na-not available; GenBank IDs are derived from NCBI (http://www.ncbi.nlm.nih.gov/genbank/); Ensembl ID was derived from Ensembl genome database (http://www.ensembl.org); UNIPROT refers to UniprotKB/Swiss-Prot IDs for individual endopeptidase-like proteins (see http://kr.expasy.org); bps refers to base pairs of nucleotide sequences; pl refers to theoretical isoelectric points; the number of coding exons are listed; high gene expression levels are in **bold**.

Citation: Holmes RS (2013) Evolution of Mammalian KELL Blood Group Glycoproteins and Genes (KEL): Evidence for a Marsupial Origin from an Ancestral M13 Type II Endopeptidase Gene. J Phylogen Evolution Biol 1: 112. doi:10.4172/2329-9002.1000112
This C-terminal region is predicted to be localized extracellularly, and to contain an active M13-like endopeptidase sequence capable of metabolizing physiologically active peptides. Four (horse and pig KELL sequences) to eight (mouse KELL sequence) N-glycosylation sites were predicted for the mammalian KELL sequences examined, which shared 16 distinct N-glycosylation sites (Figure 1; Table 3). Most mammalian KELL sequences exhibited at least four common N-glycosylation sites, to 16 distinct N-glycosylation sites (Figure 1; Table 3). It is relevant to note that a high frequency human KEL polymorphism (site 10, Table 3) is responsible for removing an N-glycosylation site, resulting in the synthesis of antigens K1 and K2 with distinct antigenic properties [27,32]. This lends support to a significant role being played by at least this N-glycosylation site in contributing to the structure and antigenic properties of this protein.

Twelve active site residues were conserved for the mammalian KELL sequences examined (Figure 1; Supplementary Table 1s; Figure 1), which are deduced from the M13 peptidase domain active site residues reported for the NCBI domain studies [47]. These included an active site catalytic residue (Glu563); an active site proton donor (Asp619); 3 residues involved in binding the active site Zinc (His562, His566 and Glu615), deduced from 3D studies of a related enzyme, MME [14]; and 7 other active site residues, also deduced from the MME tertiary structure, namely Asn521, Ala522, Ile599, Tyr658, Ala659, His675 and Arg681 (Figure 1; Supplementary Table 1s; Figure 1). Of particular significance, however, was the amino acid substitution observed for the KELL ‘active site’ Glu563 residue, which was replaced by a basic amino acid (lysine) for the rat and rabbit KELL sequences. This is likely to have significance, however, was the amino acid substitution observed for the KELL ‘active site’ Glu563 residue, which was replaced by a basic amino acid (lysine) for the rat and rabbit KELL sequences. This is likely to have
a major impact on the endopeptidase activity for these proteins and may reflect a ‘null’ phenotype for rat and rabbit KELL proteins, similar to that observed for the human K_2 KELL variant phenotype [33]. Most of the other residues in the active site and C-terminal KELL regions were predominantly conserved for many of the mammalian KELL proteins examined, which may reflect strong functional roles for these sequences as well (Figure 1).

There were several conserved Cys residues among the mammalian KELL sequences examined, including Cys53 (Figure 1), which forms a heterodimeric disulfide bond with the XK red cell membrane protein, a multipass transport protein [8]. Ten other Cys residues were conserved among the mammalian KELL sequences examined, which aligned with 10 of 12 Cys residues reported as forming disulfide bonds for the human MME sequence [53]. These included the following potential disulfide bond positions for human KELL: Cys58-Cys63; Cys81-Cys698; Cys89-Cys663; Cys136-Cys391; and Cys591-Cys729, which suggests that mammalian KELL proteins contain five disulfide bonds in similar positions to those reported for five of the six human MME disulfide bonds [33], in addition to the disulfide bond linking KELL with the XK red cell membrane protein [8].

Alignments of the C-terminal sequences for human, opossum and tammar wallaby (Sarcophilus harrisii), KELL proteins are shown in Supplementary Figure 1s, together with the corresponding KEL_3' nucleotide sequences. In contrast to all other mammalian C-terminal KELL sequences examined, the opossum KELL sequence contained three additional amino acids (Leu-Ser-Ala) to the C-terminal Trp residue observed for all other mammalian KELL sequences. It is likely that mutations in this region of the ancestral opossum KEL sequence may have caused a shift of the stop codon downstream, with a corresponding extension of the C-terminal opossum KELL sequence.

Predicted secondary and tertiary structures for mammalian KELL proteins

Predicted secondary structures for human, opossum, and tammar wallaby (Sarcophilus harrisii), KELL proteins are shown in Supplementary Figure 1s, together with the corresponding KEL_3' nucleotide sequences. In contrast to all other mammalian C-terminal KELL sequences examined, the opossum KELL sequence contained three additional amino acids (Leu-Ser-Ala) to the C-terminal Trp residue observed for all other mammalian KELL sequences. It is likely that mutations in this region of the ancestral opossum KELL sequence may have caused a shift of the stop codon downstream, with a corresponding extension of the C-terminal opossum KELL sequence.

**Table 3: Predicted N-Glycosylation sites for mammalian KELL sequences.**

Amino acids are represented as N (asparagine), P (proline); S (serine); G (glutamine), T (threonine); G (glycine); L (leucine), I (isoleucine), V (valine), M (methionine), W (tryptophan); H (histidine); E (glutamate); D (aspartate); K (lysine) and R (arginine); sites are numbered in sequence from the N-terminus.

| Endopeptidase | Site 1 | Site 2 | Site 3 | Site 4 | Site 5 | Site 6 | Site 7 | Site 8 | Site 9 | Site 10 | Site 11 | Site 12 | Site 13 | Site 14 | Site 15 | Site 16 | Total No of sites |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------------------|
| KELL          |       |       |       |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Human         | 75NTSV| 96NSFS|       |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Chimpanzee    | 75NTSV| 96NSFS|       |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Gorilla       | 75NTSV| 96NSFS|       |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Gibbon        | 75NTSV| 96NSFS|       |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Rhesus        | 75NTSV| 165NTWS|      |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Squirrel monkey| 75NTSV| 90NTSF|       |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Mouse         | 75NRSA| 93NGTS|       |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Rat           | 75NRSA| 93NGTS|       |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Cow           | 84NTSV| 102NRGTG|    |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Horse         | 75NTSV| 161NTGS|       |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Pig           | 110NRATG|      |       |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Rabbit        | 165NTGS| 177NKSR|    |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Panda         | 100NTSV| 190NTWS|       |       |       |       |       |       |       |       |       |       |       |       |       |      |                   |
| Opossum       | 124NASV| 140NTLS| 147NESR|       |       |       |       |       |       |       |       |       |       |       |       |      |                   |

Figure 3: Comparisons of the known tertiary structure for human ECE1 with the predicted structure for human KELL. The structure for the human ECE1 subunit and the predicted structure for the human KELL subunit are based on the reported structure for human ECE1 [20], and obtained using the SWISS MODEL web site based on PDB 3dw6A http://swissmodel.expasy.org/workspace/ The rainbow color code describes the 3-D structures from the N- (blue) to C-termini (red color); predicted α-helices, β-sheets and proposed active site cleft are shown.

Page 5 of 9
26 and 27, which contained several predicted active site residues for human KELL. A predicted tertiary structure for human KELL is shown in Figure 3, together with human EC1 [20]. The tertiary structure of the extracellular domain (residues 60-713) for human KELL was similar to that described for human EC1 [20]. Twenty-eight a-helices and nine b-sheet structures were observed, which is similar to that described for the predicted secondary structure for this enzyme. In addition, two major domains for these enzymes were observed, that enclose a large cavity previously shown to contain the enzyme's active site. The more C-terminal of these two domains has been shown to have a fold similar to that of thermolysin which contains the active site residue, whereas the other domain may serve to control access of substrates to the active site [20].

Overall, the predicted human KELL structure closely resembled that reported for human EC1 [20], and MME [14]. Previous modeling studies have reported that the human KELL protein contains two globular domains, consisting mostly of a-helices, with the domain situated closest to the transmembrane sequence containing both the N- and C-terminal sequences and the active site [54]. In addition, the outer domain of the protein contained all of the amino acids involved in forming at least 30 human KELL antigenic sites.

**Gene locations, exonic structures and regulatory sequences for mammalian KEL genes**

Table 1 summarizes the predicted locations for mammalian KEL genes based upon BLAT interrogations of several mammalian genomes, using the reported sequences for human [2] and mouse [9], KELL and the predicted sequences for other mammalian KELL proteins and the UC Santa Cruz genome browser [44]. The mammalian KEL genes examined were transcribed on the minus strand. Figure 1 summarizes the predicted exonic start sites for human, mouse, gibbon, horse and rat KEL genes, based on the KELL 'b' isoform, with each having 18 coding exons, in identical or similar positions to those predicted for the human KEL gene.

Figure 4 shows the predicted structures for the major human and mouse KEL transcripts. In each case, two major transcripts were observed, including the reference sequences (NM_000420 and NM_032540), which were 2811 and 2531 bps in length, with extended 5'- and 3'-untranslated regions (UTR), for the human and mouse KEL transcripts. The two major human KEL transcript isoforms, designated as 'a' and 'b', encode proteins with distinct N-terminal amino acid sequences, and contain 732 and 713 amino acids with 19 and 18 coding exons, respectively. The predicted structures for the major human and mouse KEL promoters are shown in Table 1 and Figure 2 [28]. The human KEL promoter region does not contain a TATA box, but has potential transcription factor binding sites for GATA-1 and Sp1 [29], as well as several other KEL promoter transcription factor binding sites (Table 4). Of particular significance to KEL gene regulation for erythroid cell development are the following sites: NF-E2: a transcriptional factor essential for erythroid maturation and differentiation, which also participates in the transcriptional activation of the mammalian beta-globin gene locus [55]; GATA1: a transcriptional activator localized in the promoters of the globin gene family and other erythroid-specific genes [56]; and EVI1: a transcription activator which is essential for the proliferation and maintenance of hematopoietic stem cells [57]. It would appear that the KEL gene promoter is well endowed with gene regulatory sequences, which may contribute to the high levels of KEL expression in mammalian erythroid cells, and to the maintenance of this expression during development.

---

**Table 1:** Predicted exonic structures and regulatory sequences for mammalian KEL genes

| TFBS Name       | Name              | Chr 7 Position | Function/Role                                                                 | Sequence     | UNIPROT ID   |
|-----------------|-------------------|----------------|-------------------------------------------------------------------------------|--------------|--------------|
| NFE2            | Transcription factor NF-E2 | 142,661,722-142,662,732 | Essential for regulating erythroid maturation and differentiation | TGATGACCTCAC | Q16621       |
| BACH2           | Transcription regulator protein BACH2 | 142,661,721-142,661,731 | Transcriptional regulator                                                   | GGTGAATCAC   | Q9BYV9       |
| GATA1           | Erythroid transcription factor | 142,661,694-142,661,706 | Transcriptional activator acting as a general switch for erythroid development | GGGTGATAAGAG| P15976       |
| FOX2            | Forkhead box protein J2 | 142,660,952-142,660,965 | Transcriptional activator                                                   | ACAATATAC    | Q9POK8       |
| EVI1            | MDS1 and EVI1 complex locus protein EVI1 | 142,660,943-142,660,958 | Transcriptional regulator involved in hematopoiesis                        | TGGCAAGATACAA | P14404      |
| ETVI            | MDS1 and EVI1 complex locus protein EVI1 | 142,659,507-142,659,522 | Transcriptional regulator involved in hematopoiesis                        | TGAGAAGCTGAGAT | P14404      |
| TLX2            | T-cell leukemia homeobox protein 2 | 142,658,874-142,658,883 | Transcriptional activator that binds DNA via its homeobox | AGGTAAGTGG  | Q9Q48        |

---

**Table 4:** Identification of Transcription Factor Binding Sites (TFBS) within the Human KEL gene promoter.

The identification of KEL TFBS within the KEL promoter region was undertaken using the human genome browser (http://genome.ucsc.edu). UNIPROT refers to UniprotKB/Swiss-Prot IDs for individual TFBS sequences (http://kr.expasy.org).
**KEL tissue expression**

Supplementary Figure 2 presents ‘heat maps’ showing comparative KEL expression in various human tissues obtained from GNF Expression Atlas 2 Data using U133A and GNF1H (human) chips (http://genome.ucsc.edu; http://biogps.gnf.org) [48]. These data supported high level tissue expression for human KEL, particularly in bone marrow (BM) erythroid cells, which is consistent with previous reports for this gene [5-7]. Another study reported that the KEL protein is expressed in nonerythroid tissues, including testis and skeletal muscle, although KEL transcripts were also reported in several other tissues [3]. Supplementary Figure 2 supports this wider KEL tissue expression, particularly for testis and fetal liver. Overall, human KEL and mouse Kell tissue expression levels were 0.5-0.8 times the average level of gene expression which supports a key role for this protein as an endopeptidase, but with a highly specific erythroid tissue expression profile [5-7,28].

**Phylogeny and divergence of mammalian KEL and other vertebrate M13 type II endopeptidase genes and proteins**

A phylogenetic tree (Figure 5) was calculated by the progressive alignment of 15 mammalian KELL amino acid sequences with several other vertebrate M13 type II endopeptidase-like sequences, which was ‘rooted’ with the worm (*Caenorhabditis elegans*) membrane metalloendopeptidase (NEP22) sequence Table 1 and 2. The phylogram showed clustering of the KELL sequences into groups consistent with their evolutionary relatedness, as well as groups for each of the vertebrate M13 type II endopeptidase families, which were distinct from the worm NEP22 sequence. These groups were significantly different from each other (with bootstrap values >90). This data suggests that the KEL gene and KELL protein have appeared early in marsupial and eutherian mammalian evolution, for which a proposed common ancestor for these genes may have predated or coincided with the appearance of marsupial mammals during vertebrate evolution. The KEL gene (and KELL protein) was apparently absent from monotreme (platypus: *Ornithorhynchus anatinus*), bird (chicken: *Gallus gallus*), reptile (lizard: *Anolis carolensis*), amphibian (frog: *Xenopus tropicalis*) and fish (zebrafish: *Danio rerio*) genomes, whereas other vertebrate M13 type II endopeptidase genes (MME, MMEL1, ECE1, ECE2, ECEL1 and PHEX) were present throughout vertebrate evolution (Figure 5) [57]. In addition, these results suggested that the mammalian KEL gene evolved, following a gene duplication event of an ECE-like gene, given that ECE1, ECE2 and KEL appear to be more closely related to each other, than to other M13 type II endopeptidase genes. This is consistent with a recent study of the evolution of vertebrate *ECE1* genes and proteins [57].

**Conclusions**

This study suggests that mammalian KEL genes and encoded KELL proteins represent a distinct gene and protein family of M13 Type II neutral endopeptidase-like proteins which share conserved sequences, and active site residues with those reported for other related M13 Type II endopeptidases, MME, MMEL1, ECE1, ECE2, ECEL1 and PHEX, previously studied [19-22,25,58]. KELL has a distinctive property among these M13 Type II endopeptidase gene families. A genetic distance scale is shown. The number of times a clade (sequences common to a node or branch) occurred in the bootstrap replicates are shown. Only replicate values of 90 or more, which are highly significant, are shown with 100 bootstrap replicates performed in each case. A proposed sequence of gene duplication events is shown arising from an ancestral endopeptidase-like gene.
KELL is encoded by a single gene (KEL), among the marsupial and eutherian mammalian genomes studied, which is highly expressed in human and mouse erythroid cells and usually contained 18 or 19 coding exons on the negative strand. Several transcription factor binding sites were localized within the human KEL gene promotor region, including GATA1, EVII and NFE2, which are essential in regulating erythroid cell differentiation and maintenance of expression, and may contribute significantly to the high level of gene expression in bone marrow erythroid cells.

Predicted secondary and tertiary structures for vertebrate KELL proteins showed a strong similarity with other M13 type II endopeptidase-like proteins. Several major structural domains were apparent for mammalian KELL proteins, including a N-terminal cytoplasmic tail of varying lengths between human KELL isoforms and other mammalian KELL sequences; a transmembrane domain which anchors the enzyme to the cell membrane; and an extracellular domain, containing the active site (including a zinc binding site), which is responsible for endopeptidase activity; and four conserved N-glycosylation sites, which may contribute significantly to the antigenic properties for this blood group protein. Several KELL endopeptidase domain cysteine residues were conserved among mammalian sequences, including Cys33 (see Figure 1), which forms a heterodimeric disulfide bond with the XK red cell membrane protein, a multipass transport protein [8]. Ten other Cys residues were also conserved which aligned with 10 of 12 Cys residues forming disulfide bonds for the human MME sequence [53]. This suggests that mammalian KELL proteins may contain five disulfide bonds in similar positions to those reported for five of the six human MME disulfide bonds [53]. Alternatively, these Cys residues may serve other KELL structural or functional roles.

A phylogenetic study used 15 mammalian KELL sequences with a range of other vertebrate M13 type II endopeptidase sequences, which indicated that the KEL gene had appeared early in marsupial mammalian evolution, prior to or coincident with the appearance of marsupial mammals, and existed as a distinct gene family within this group, together with the MME and MMEL1; ECE1 and ECE2; ECEL1; and PHEX gene groups. Moreover, the study indicated that the KEL gene may have originated from a gene duplication event of an ancestral mammalian ECE-like gene, in common with the ECE1 and ECE2 genes, which originated much earlier in vertebrate evolution, as both of these genes have been reported in bird, reptile, amphibian and fish genomes [57].

References
1. Rawlings ND, Barrett AJ (1993) Evolutionary families of peptidases. Biochem J 290: 205-218.
2. Lee ME, Temizer DH, Clifford JA, Quertemous T (1991) Cloning of the GATA- binding protein that regulates endothelin-1 gene expression in endothelial cells. J Biol Chem 266: 16188-16192.
3. Russo D, Wu X, Redman CM, Lee S (2000) Expression of Kell blood group protein in nonerythroid tissues. Blood 96: 340-346.
4. Bland ND, Pinney JW, Thomas JE, Turner AJ, Isaac RE (2008) Bioinformatic analysis of the neprilysin (M13) family of peptidases reveals complex evolutionary and functional relationships. BMC Evol Biol 8: 16.
5. Reid ME, Mohandas N (2004) Red blood cell blood group antigens: structure and function. Semin Hematol 41: 93-117.
6. Mohandas N, Narla A (2005) Blood group antigens in health and disease. Curr Opin Hematol 12: 135-140.
7. Dean L (2005) The Kell blood group. In: Blood Groups and Red Cell Antigens, Chapter 8, National Center for Biotechnology Information, Bethesda MD, USA.
8. Russo D, Redman C, Lee S (1998) Association of XK and Kell blood group proteins. J Biol Chem 273: 13950-13956.
9. Lee S, Russo DC, Pu J, Ho M, Redman CM (2000) The mouse Kell blood group gene (Kell): cDNA sequence, genomic organization, expression, and enzymatic function. Immunogenetics 52: 53-62.
10. Lee S, Lin M, Mele A, Cao Y, Farmar J, et al. (1999) Proteolytic processing of big endothelin-3 by the kell blood group protein. Blood 94: 1440-1450.
11. Clapéron A, Rose C, Gane P, Collec E, Bertrand O, et al. (2005) The Kell protein of the common K2 phenotype is a catalytically active metallopeptidase, whereas the rare Kell K1 antigen is inactive. Identification of novel substrates for the Kell protein. J Biol Chem 280: 21272-21283.
12. Letarte M, Vera S, Tran R, Addis JB, Onizuka RJ, et al. (1988) Common acute lymphocytic leukaemia antigen is identical to neutral endopeptidase. J Exp Med 168: 1247-1253.
13. Ship MA, Richardson NE, Sayre PH, Brown NR, Mastellet EL, et al. (1988) Molecular cloning of the common acute lymphoblastic leukemia antigen (CALLA) identifies a type II integral membrane protein. Proc Natl Acad Sci U S A 85: 4819-4823.
14. Oefner C, D’Arcy A, Hennig M, Winkler FK, Dale GE (2000) Structure of human neutral endopeptidase (Nephrilysin) complexed with phosphoramidon. J Mol Biol 296: 341-349.
15. Bonvoulou R, Lemieux N, Crine P, Boileau G, DesGroseillers L (2001) Molecular cloning, tissue distribution, and chromosomal localization of MMEL2, a gene encoding for a novel human member of the neutral endopeptidase-24.11 family. DNA Cell Biol 20: 493-498.
16. Hirschfield GM, Liu X, Han Y, Gorkov IP, Lu Y, et al. (2010) Variants at IRF5-TNPO3, 17q12-21 and MMEL1 are associated with primary biliary cirrhosis. Nat Genet 42: 655-657.
17. Ban DM, McCauley JL, Zuvich R, Baker A, Bergamaschi L, et al. (2010) A non- synonymous SNP within membrane metalloendopeptidase-like 1 (MMEL1) is associated with multiple sclerosis. Genes Immun 11: 660-664.
18. Danoy P, Wei M, Johanna H, Jiang L, He D, et al. (2011) Association of variants in MMEL1 and CTLA4 with rheumatoid arthritis in the Han Chinese population. Ann Rheum Dis 70: 1793-1797.
19. Schmidt M, Kröger B, Jacob E, Seubelberger H, Subkowski T, et al. (1994) Molecular characterization of human and bovine endothelin converting enzyme (ECE-1). FEBS Lett 356: 238-243.
20. Schulz H, Dale GE, Karimi-Nejad Y, Oefner C (2009) Structure of human endothelin-converting enzyme 1 complexed with phosphoramidon. J Mol Biol 385: 178-187.
21. Lorenzo MN, Khan RY, Wang Y, Tai SC, Chan GC, et al. (2001) Human endothelin converting enzyme-2 (ECE2): characterization of mRNA species and chromosomal localization. Biochim Biophys Acta 1522: 49-52.
22. Mzhavia N, Pan H, Che FY, Fricker LD, Devi LA (2003) Characterization of endothelin converting enzyme-2. Implication for a role in the nonsclassical processing of regulatory peptides. J Biol Chem 278: 14704-14711.
23. Valdenaire O, Richards JG, Faurl RL, Schweizer A (1999) XCE, a new member of the endothelin-converting enzyme and neutral endopeptidase family, is preferentially expressed in the CNS. Brain Res Mol Brain Res 64: 211-221.
24. Schweizer A, Valdenaire O, Köster A, Lang Y, Schmidt G, et al. (1999) Neonatal lethality in mice deficient in XCE, a novel member of the endothelin-converting enzyme and neutral endopeptidase family. J Biol Chem 274: 20450-20456.
25. Du L, Desbarats M, Vie J, Glorieux FH, Cawthorn C, et al. (1996) cDNA cloning of the murine Pex gene implicated in X-linked hypophosphataemia and evidence for expression in bone. Genomics 36: 22-28.
26. Miao D, Bai X, Panda D, McKee M, Karaplis A, et al. (2001) Osteomalacia in hyp mice is associated with abnormal pex gene expression and with altered bone matrix protein expression and deposition. Endocrinology 142: 926-939.
27. Lee S, Zambas E, Green ED, Redman C (1995) Organization of the gene encoding the human Kell blood group protein. Blood 85: 1364-1370.
28. Thierry-Mieg D, Thierry-Mieg J (2006) AceView: A comprehensive cDNA-supported gene and transcripts annotation. Genome Biol 7: S12.
29. Redman CM, Lee S (1995) The Kell blood group system. Transfus Clin Biol 2: 243-249.
30. Lee S, Wu X, Son S, Naime D, Reid M, et al. (1996) Point mutations characterize KEL10, the KEL3, KEL4, and KEL21 alleles, and the KEL17 and KEL11 alleles. Transfusion 36: 490-494.
Citation: Holmes RS (2013) Evolution of Mammalian KELL Blood Group Glycoproteins and Genes (KEL): Evidence for a Marsupial Origin from an Ancestral M13 Type II Endopeptidase Gene. J Phylogen Evolution Biol 1: 112. doi:10.4172/2329-9002.1000112

31. Zhu X, Rivera A, Golub MS, Peng J, Sha Q, et al. (2009) Changes in red cell ion transport, reduced intratumoral neovascularization, and some mild motor function abnormalities accompany targeted disruption of the Mouse Kell gene (Ktl). Am J Hematol 84: 492-498.

32. Lee S, Wu X, Reid M, Redman C (1995) Molecular basis of the K.6,-7 [Js(a+b-)] phenotype in the Kell blood group system. Transfusion 35: 822-825.

33. Lee S, Debnath AK, Wu X, Scolfied T, George T, et al. (2006) Molecular basis of two novel high-prevalence antigens in the Kell blood group system, KALT and KTIM. Transfusion 46: 1323-1327.

34. Yu LC, Tuw YC, Chang CY, Lin M (2001) Molecular basis of the Kell-null phenotype: a mutation at the splice site of human KEL gene abolishes the expression of Kell blood group antigens. J Biol Chem 276: 10247-10252.

35. Yang Y, Wang L, Wang C, Chen H, Guo Z, et al. (2009) Two novel null alleles of the KEL gene detected in two Chinese women with the Kn(null) phenotype. Transfus Med 19: 235-244.

36. Hoffman JF, Joiner W, Nehrk K, Potapova O, Foye K, et al. (2003) The hSK4 (KCNN4) isoform is the Ca2+-activated K+ channel (Gardos channel) in human red blood cells. Proc Natl Acad Sci U S A 100: 7366-7371.

37. Tuson, Hue-Roye K, Koval K, Imlay S, Desai R, et al. (2011) Possible suppression of fetal erythropoiesis by the Kell blood group antibody anti-Kp(a). Immunohematology 27: 58-60.

38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410.

39. Guo R, Quarters LD (1997) Cloning and sequencing of human PEX from a bone cDNA library: Evidence for its developmental stage-specific regulation in osteoblasts. J Bone Miner Res 12: 1009-1017.

40. Grief M, Mumm S, Waeltz P, Mazzarella R, Whyte MP, et al. (1997) Expression and cloning of the human X-linked hypophosphatemia gene cDNA. Biochem Biophys Res Commun 231: 635-639.

41. Dixon PH, Christie PT, Wooding C, Trump D, Grief M, et al. (1998) Mutational analysis of PHEX gene in X-linked hypophosphatemia. J Clin Endocrinol Metab 83: 3615-3623.

42. Kiyu-Seo S, Sasaki M, Yokohama H, Nakagomi S, Hirayama T, et al. (2000) Damage-induced neuronal endopeptidase (DINE) is a unique metallopeptidase expressed in response to neuronal damage and activates superoxide scavengers. Proc Natl Acad Sci U S A 97: 4345-4350.

43. Bernt A, Vargas MA, Desgresellers L, Boileau G (2004) Endothelin-converting enzyme-like 1 (ECEL1) is present both in the plasma membrane and in the endoplasmic reticulum. Biochem J 380: 881-888.

44. Kent WJ, Sugrey CW, Furey TS, Roskin KM, Pringle TH, et al. (2002) The human genome browser at UCSC. Genome Res 12: 996-1006.

45. Schwede T, Kopp J, Guex N, Peitsch MC (2003) SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res 31: 3381-3385.

46. Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. J Mol Biol 305: 567-580.

47. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, et al. (2011) CDD: a Conserved Domain Database for the functional annotation of proteins. Nucleic Acids Res 39: D225-D229.

48. Su AI, Willshire T, Batalov S, Lapp H, Ching KA, et al. (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A 101: 6062-6067.

49. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, et al. (2008) Phylolgy. fr: Robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36: W465-W469.

50. Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.

51. Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52: 696-704.

52. Ofner C, Roques BP, Fournie-Zaluski MC, Dale GE (2004) Structural analysis of nephrilysin with various specific and potent inhibitors. Acta Crystallogr D Biol Crystallogr 60: 392-396.

53. Lee S, Debnath AK, Redman CM (2003) Active amino acids of the Kell blood group protein and model of the ectodomain based on the structure of neutral endopeptidase 24.11. Blood 102: 3029-3034.

54. Shyu YC, Lee TL, Ting CY, Wen SC, Hsieh LJ, et al. (2005) Sumoylation of the mouse Kell blood group protein and model of the ectodomain based on the structure of neutral endopeptidase 24.11. Blood 102: 3029-3034.

55. Trinh CD, Evans T, Felsenfeld G, Boguski MS (1990) Structure and evolution of a human erythroid transcription factor. Nature 343: 92-96.

56. Shimabe M, Goya S, Watanabe-Okochi N, Yoshimi A, Ichikawa M, et al. (2009) Pbx1 is a downstream target of Evi-1 in hematopoietic stem/progenitors and leukemic cells. Oncogene 28: 4364-4374.

57. Holmes RS, Cox LA (2013) Comparative studies of vertebrate endothelin-converting enzyme-like 1 genes and proteins. Res Rep Biochemistry 3: 1-16.