Identification and analysis of mammalian KLK6 orthologue genes for prediction of physiological substrates

Georgios Pampalakis\textsuperscript{a}, Maria Arampatzidou\textsuperscript{a}, Grigoris Amoutzias\textsuperscript{b}, Sofia Kossida\textsuperscript{c}, Georgia Sotiropoulou\textsuperscript{a,*}

\textsuperscript{a} Department of Pharmacy, School of Health Sciences, University of Patras, Rion-Patras 26500, Greece
\textsuperscript{b} Department of Ecology and Evolution, Biophore, University of Lausanne, 1015 Lausanne, Switzerland
\textsuperscript{c} Bioinformatics Group, Foundation for Biomedical Research of the Academy of Athens, Athens 11527, Greece

Received 1 June 2007; received in revised form 25 October 2007; accepted 17 November 2007

Abstract

Human kallikrein-related peptidase 6 (KLK6) is a novel serine protease that is aberrantly expressed in human cancers and represents a serum biomarker for the molecular diagnosis and monitoring of ovarian cancer. Here, we report the cloning and analysis of human kallikrein-related peptidase 6 gene \((\text{KLK6})\) orthologues in model organisms and farm animals. The corresponding full-length cDNAs were assembled from partial sequences retrieved from EST and genomic databases. Alignment of inferred protein sequences indicated a high degree of conservation of the encoded enzyme. We found that, similarly to (HUMAN)KLK6, monkey, cattle, mouse and rat orthologue genes encode for multiple transcript variants. This strengthens our previously published data showing that (HUMAN)KLK6 transcription is coordinately regulated by alternative promoters. Analysis of the KLK6 upstream genomic region led to the identification of multiple conserved regulatory regions with motifs for nuclear receptor transcription factors. Interestingly, we found that specific CpG dinucleotides in the proximal promoter, that were shown to regulate (HUMAN)KLK6 gene expression via DNA methylation, are conserved in orthologue genes, indicating epigenetic regulation of the KLK6 gene. Construction of a protein–protein interaction network indicated that KLK6 likely acts on the TGF-b1 signal transduction pathway to regulate certain cytoskeletal proteins, such as vimentin and keratin 8, thus, KLK6 may control cell shape that, in turn, regulates cell migration and motility.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Human kallikrein-related peptidase 6 (tissue kallikrein 6/protease M/zyme/neurosin); Orthologues; Splice variants; Conserved regulatory sequences; KLK6 interactome

1. Introduction

The gene encoding human kallikrein-related peptidase 6 (tissue kallikrein 6 or protease M/zyme/neurosin) was originally identified and cloned by differential display based on its inactivated expression in metastatic breast cancer cells when compared to corresponding primary tumor cells (Anisowicz et al., 1996). KLK6 is a member of the human tissue kallikrein family, a cluster of 15 genes tandemly localized on human chromosome 19q13.4 that are predicted to encode for trypsin- or chymotrypsin-like serine proteases. Tissue kallikreins constitute the largest serine protease family within the entire human genome (Borgoño and Diamandis, 2004). The enzymatic activity of KLK6 has been implicated in neurodegenerative and inflammatory diseases of the central nervous system, including multiple sclerosis, Alzheimer’s and Parkinson’s disease (Borgoño and Diamandis, 2004; Iwata et al., 2003; Bernett et al., 2002). It was shown that the enzymatic activity of KLK6 is regulated by an autoactivation/autolytic inactivation mechanism (Bayés et al., 2004; Blaber et al., 2007).
In this study, we applied biocomputational approaches to analyze partial DNA sequences retrieved from EST and genomic databases, in order to identify orthologue genes of (HUMAN)KLK6 and assemble their full-length cDNA sequences (virtual transcripts). Recently, we suggested that transcription of KLK6 is controlled by three alternative promoters (P1, P2 and P3) that all drive the synthesis of wild-type protein and display tissue-specific activity (Pampalakis et al., 2004; Pampalakis and Sotiropoulou, 2006). Interestingly, differential regulation of P1 and P2 promoters has been associated with CNS demyelinating diseases (Christophi et al., 2004). Production of multiple transcript variants is corroborated by the observation reported here that orthologue genes are also regulated by alternative promoters that exploit intrinsic sequences. In addition, we have analyzed the KLK6 promoter sequences in mammals and identified multiple conserved consensus sequences for binding of transcription factors. In addition, conserved positions of CpG dinucleotides were identified within the KLK6 promoter sequences from different species. Previously, we showed that cytosine methylation at these sites underlies the aberrant expression of KLK6 in human tumor cells (Pampalakis and Sotiropoulou, 2006). Consistently with the previous identification of human KLK6 isoforms produced by splicing out coding exons (splice variants), mining of EST orthologue sequences revealed novel splice variants in cattle. It is likely that accumulation of genome sequencing data will lead to the identification of additional KLK6 isoforms in other species. In order to identify potential substrates and biological functions of KLK6, we have constructed a protein–protein interaction network for KLK6 (the KLK6 interactome) based on human and rat data. Expression of KLK6 up-regulates keratin 8 (KRT8) and down-regulates vimentin (VIM). These changes are known to control the cytoskeletal structure and reduce cellular motility and migration.

In conclusion, we showed that KLK6 activity is regulated both at the level of transcription and post-translationally. The mechanisms that control the expression of KLK6 are likely conserved among species and involve the usage of multiple promoters and production of splice variants. The encoded protein is highly conserved indicating a conserved biological function. The KLK6 interactome predicted that this secreted serine protease may also act intracellularly to regulate cell migration and motility.

2. Materials and methods

2.1. Nomenclature

The nomenclature that was recently proposed for the tissue kallikrein family by Lundwall et al. (2006) was followed here. Briefly, gene names are given in italics and protein names with standard font. Animal symbols used in the present study, were obtained from SWISS-PROT (http://www.expasy.ch/cgi-bin/speclist). Under this system, the prefix (HUMAN) denotes Homo sapiens, (BOVIN) Bos taurus, (CANFA) Canis familiaris, (PANTR) Pan troglodytes, (PIG) Sus scrofa, (MACMU) Macaca mulatta, (MOUSE) Mus musculus and (RAT) Rattus norvegicus.

2.2. Nucleotide and protein databases

Expressed sequence tag sequences were retrieved from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov) and ENSEMBL (http://www.ensembl.org). Search of EST databases was performed by megaBLAST, BLASTN and TBLASTN algorithms (Altschul et al., 1997), while the KLK6 gene structure was retrieved from MapViewer (http://www.ncbi.nlm.nih.gov/mapview).

2.3. Nucleotide and protein tools

Translation of nucleotide sequences and motif prediction were performed with tools available in Expasy (http://www.expasy.org). Sequence alignments were performed by BL2seq (BLAST two sequences) or DNAMAN (LYNNon Corporation, Quebec, Canada) for multiple alignments.

2.4. Identification of regulatory elements

Identification of regulatory elements upstream of KLK6 gene was conducted with the program ECR (evolutionary conserved regions, http://www.dcode.org).

2.5. Protein–protein interaction network

Protein–protein interactions of human KLK6 were identified and displayed by the ingenuity pathway analysis (IPA) (Ingenuity Systems, Inc., Redwood City, California, http://www.ingenuity.com/). The pathway was constructed with a direction from KLK6 to the nodes, which are upstream of the differentially expressed genes and regulate their expression.

3. Results

3.1. Identification of full-length cDNA sequences of KLK6 orthologue genes

Fig. 1 shows a schematic representation of full-length orthologue cDNA sequences of KLK6. Accession numbers of representative EST clones used in this study are given in Table 1. EST databases containing sequences from animal genomes were searched with the KLK6 protein sequence using the TBLASTN algorithm (Altschul et al., 1997). Full-length cDNAs (virtual transcripts) were derived as contigs of partially overlapping EST sequences. The derived virtual transcripts encoding KLK6 orthologues were searched against genomic animal databases, in order to identify intron–exon junctions and the corresponding gene structures. All splice junctions satisfied the rule GT (donor)–AG (acceptor) that filters out transcripts possibly resulting from spliceosomal errors (Modrek and Lee, 2002). For all identified orthologues, translational start sites lied within a conserved Kozak sequence (Kozak, 1987) and a polyadenylation signal was identified in the 3′ untranslated region. The mouse, rat and dog orthologues were identified by Homologene (http://www.ncbi.nlm.nih.gov/). Full-length transcripts were
obtained by EST database mining. KLK6 orthologues display identical gene structure and consist of five coding exons with equal or similar number of nucleotides. Sequences of all virtual transcripts identified here were included as a Supplementary data.

P. troglodytes. Because no ESTs encoding for KLK6 were found, probably due to the low number of available EST clones, we searched the chimpanzee high throughput genome sequencing (HTGS) database at NCBI with the TBLASTN algorithm using the human KLK6 amino acid sequence. The clone CH251-355A20 (Genbank™: AC130782) was identified that represents a fragment of chromosome 19 encompassing the chimpanzee kallikrein locus. This sequence was retrieved and used for further analysis. Aligning of chimpanzee genomic sequence against (HUMAN)KLK6 sequence enabled prediction of the (PANTRO)KLK6 gene structure shown in Fig. 1, that however, contains a number of gaps due to incomplete genomic sequencing. The inferred partial protein structure of chimpanzee KLK6 is 100% identical to the human KLK6.

M. mulatta. For identification of the (HUMAN)KLK6 orthologue in Rhesus monkey, data were retrieved from ENSEMBL and NCBI. (MACA)KLK6 is localized on monkey chromosome 19.

B. taurus. In cow genome, kallikrein family of genes map on chromosome 18. Multiple overlapping EST clones were identified that corresponded to the complete ORF of (BOVIN)KLK6 gene including 5’ and 3’ untranslated sequences.

S. scrofa. In pig genome, a sequence (Genbank™: AC149292) was identified that contained the kallikrein locus, including the (PIG)KLK6 gene, as described for chimpanzee orthologues but its chromosomal localization is unknown. A full-length cDNA transcript for pig KLK6 was assembled as a contig of partially overlapping ESTs.

C. familiaris. In dog genome, kallikrein gene family was mapped on chromosome 1.
M. musculus and R. norvegicus. In mouse and rat, the kallikrein gene family was previously mapped on chromosome 7 and 1, respectively (Yousef and Diamandis, 2003). Here, we identified full-length KLK6 cDNAs by assembling EST clones that extended further 5′ upstream or 3′ downstream.

Xenopus tropicalis. A transcript homologous to KLK6 was identified in X. tropicalis (Genbank™; BC087753). However, it is unlikely that this represents the Xenopus orthologue of (HUMAN)KLK6, since a TBLASTN search of the inferred protein against human nucleotide sequences showed that trypsin-like proteases, such as trypsin 1/PRSS1, KLK1, KLK12, KLK13, KLK11 and KLK8 display higher homology than that trypsin-like proteases, such as trypsin 1/PRSS1, KLK1, KLK12, KLK13, KLK11 and KLK8.

Drosophila melanogaster. Using Homologene (http://www.ncbi.nlm.nih.gov/) as another tool to identify the (HUMAN) KLK6 orthologues proteins, we have identified a putative KLK6 orthologue in D. melanogaster (Genbank™; NM_166114). This mRNA encodes for a protein with 33% identity and 52% positivity to human KLK6 that has a conserved trypsin-like serine protease structure. Search of this sequence against the human nucleotide database showed that (HUMAN)KLK6 gene displayed the highest similarity. The reciprocal best BLAST hit further supports the idea that this protein may represent the Drosophila KLK6 orthologue protein.

Other mammalian species. (HUMAN)KLK6 orthologues could not be identified in cat (Felis catus) neither in sheep (Ovis aries) that is likely due to incomplete genome sequencing and the low number of available ESTs. In elephant (Loxodonta africana) genomic DNA and ab initio RNA database available by ENSEMBL, small fragments with high similarity (>60%) to (HUMAN)KLK6 were identified, but reciprocal BLAST hit showed that these sequences represented orthologues of other kallikrein-related peptidases, such as (HUMAN)KLK12, (HUMAN)KLK11 and (HUMAN) KLK8.

3.2. KLK6 transcription is regulated by alternative promoters and alternative splicing

Based on experimental data, we recently suggested that human KLK6 is coordinately regulated by three alternative promoters (P1, P2 and P3) (Pampalakis et al., 2004; Christofi et al., 2004; Pampalakis and Sotiropoulou, 2006) as schematically shown in Fig. 1. The P1 transcript is derived from P1 promoter located in the 5′ upstream region and contains two 5′ untranslated exons, the P2 transcript is derived from P2 promoter located inside intron I and contains one 5′ untranslated exon, and the P3 transcript is derived from P3 promoter located inside intron II and contains one 5′ untranslated exon. As shown in Fig. 1, EST clones derived from either P1 or P2 promoters were detected for most animals. More specifically, products of both promoters were predicted for Rhesus monkey, B. taurus, M. musculus and R. norvegicus. P1- or P2-derived transcripts are predicted to encode for full-length KLK6 protein. In B. taurus, three distinct splice variants were detected that were derived from P1 promoter (Fig. 1). All splice variants encode for wild-type bovine KLK6, since variation in sequence occurs in the 5′ untranslated region. For the other species, no splice variants were detected. In B. taurus, of 10 EST clones that extended to the 5′ end, 3 clones represented the longest P1 transcript, 4 clones represented the P1 transcript with a smaller second exon, 1 clone represented the P1 product that lacks the second untranslated exon and 2 clones represented the P2 transcript, providing a rough estimation of the relative abundance of these transcripts. Taken together, our data suggest that conserved mechanisms regulate the expression of KLK6 in mammalian species.

---

Table 1

Genbank™ accession numbers of representative EST clones encoding for KLK6 orthologues and tissue expression pattern

<table>
<thead>
<tr>
<th>Animal</th>
<th>Genbank™ accession number</th>
<th>KLK6 gene region</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pig</td>
<td>BF441391, BX671525, BX671526</td>
<td>5′ end, 3′ end, ORF</td>
<td>Mixed, Mixed, Mixed</td>
</tr>
<tr>
<td>Cattle</td>
<td>DV924321, DV871627, DV886953, DV467443, CO882996</td>
<td>P2, P1 (without un-translated Exon 2), P1 (with smaller Exon 2), Extended 3′ end, ORF</td>
<td>Skin, Hippocampus, Thalamus, Mixed, Brain</td>
</tr>
<tr>
<td>Chimpanzee</td>
<td>AY410890</td>
<td>Sequence based on alignment</td>
<td></td>
</tr>
<tr>
<td>Dog</td>
<td>CN497238, DN394359</td>
<td>Extended 3′ end, 5′ end</td>
<td>Brain, Brain</td>
</tr>
<tr>
<td>Rhesus monkey</td>
<td>BQ807957, ENSG00000167755 (ENSEMBL)</td>
<td>P1, P2</td>
<td>Hypothalamus, Hypothalamus</td>
</tr>
<tr>
<td>Rat</td>
<td>BF550821, AI45671</td>
<td>P2, Extended 3′ end</td>
<td>Mixed, Mixed</td>
</tr>
<tr>
<td>Mouse</td>
<td>BY727409, BB571443, BB180902</td>
<td>P1, P2, Extended 3′ end</td>
<td>Corpora quadrigemina, Skin, Hypothalamus</td>
</tr>
</tbody>
</table>

---

Please cite this article in press as: Pampalakis, G., et al., Identification and analysis of mammalian KLK6 orthologue genes for prediction of physiological substrates, Comput Biol Chem (2008), doi:10.1016/j.compbiochem.2007.11.002
Fig. 2. KLK6 orthologue protein sequences. Alignment of KLK6 orthologue protein sequences. The catalytic triad amino acid residues are separated by vertical white lines and are marked by an asterisk. A circle indicates an N-glycosylation site and inverted triangle an autoinactivation site (R76 in human KLK6). Autolysis sites are not conserved in rodents. The arrow indicates the signal peptide cleavage site, which was predicted by SignalP 3.0 (Bendtsen et al., 2004).

3.3. Comparison of KLK6 orthologue proteins

Based on the protein alignment shown in Fig. 2, all KLK6 orthologue proteins have six fully conserved cysteine residues that are probably engaged in the formation of conserved disulfide bonds. The amino acid residues of the catalytic triad are also fully conserved (His, coding exon 2; Asp, coding exon 3; Ser, coding exon 5). Recently, we showed that the enzymatic activity of KLK6 is self-regulated by an autocatalytic mechanism of activation and subsequent inactivation as a result of autolytic cleavage at internal R76 (Bayés et al., 2004) that is located in an exposed autolysis loop (Bernett et al., 2002; Gomis-Ruth et al., 2002). R76 is conserved in Rhesus monkey, dog, pig and bovine but not in mouse and rat. Characterization of the enzymatic properties of rat KLK6 revealed two different autoinactivation sites used with similar preference (Blaber et al., 2002). An N-glycosylation site at N132 of human KLK6 was experimentally confirmed by site-directed mutagenesis (Gomis-Ruth et al., 2002), however, the sugar moiety does not significantly affect the enzymatic activity due to its distal location from the active-site (Bernett et al., 2002; Gomis-Ruth et al., 2002). This site is conserved in most species examined here, except for rodents. Indeed, rat
KLK6 is a deglycosylated protein (Blaber et al., 2002). SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP) (Bendtsen et al., 2004) identified a signal peptide cleavage site with high similarity in all proteins that are predicted to be secreted. Table 2 shows the identities between KLK6 orthologue proteins in various animals.

3.4. Promoter analysis

Analysis of KLK6 upstream sequences predicted conserved transcription factor binding sites and CpG dinucleotide sequences, indicating common mechanisms of transcriptional regulation. The KLK6–KLK7 intergenic region that is located 5′ upstream of KLK6 and includes the putative promoter sequence was analyzed in several species in order to identify conserved regulatory regions. Multiple conserved regulatory regions were detected, as shown in Fig. 3A. Putative transcription factor binding sites (TFBS) were identified using the ECR (evolutionary conserved regions) browser at http://www.dcode.org (Ovcharenko et al., 2004). We aligned human, Rhesus monkey, mouse, rat and dog sequences, using as a cutoff detection level of ECRs a minimum length of 100 bp and a minimum identity of 70%. The identified ECRs were scanned for putative TF consensus sequences. The upstream non-coding region of the KLK6 gene is conserved within the mammalian lineage and contains several consensus sequences for nuclear receptors or basic helix-loop-helix family of transcription factors. More specifically, binding sites for HEB, E2A, PU.1, HNF4 (hepatocyte nuclear factor 4) and LXR (liver X receptor) were conserved in the proximal KLK6 promoter. HEB, E2A and PU.1 are factors involved in lymphocyte development (Murre, 2005; Hagman and Lukin, 2006), while HNF4 (Lemaigre and Zaret, 2004) is mainly a liver-specific factor. LXR is involved in regulation of genes involved in cholesterol metabolism and lipid biosynthesis (Makishima, 2005). HEB functions synergistically with myogenin, also predicted to bind to the KLK6 promoter, in order to activate the myogenin expression, thus myogenin regulates its own expression (Parker et al., 2006). Therefore, HEB and myogenin may interact for binding to the KLK6 proximal promoter. In addition, HEB forms heterodimers with E2A that bind to E-boxes to activate gene expression, an interaction involved in regulation of T-cell development (Barndt et al., 2000; Takeuchi et al., 2001).

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Human</th>
<th>Monkey</th>
<th>Pig</th>
<th>Cattle</th>
<th>Dog</th>
<th>Mouse</th>
<th>Rat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monkey</td>
<td>97</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pig</td>
<td>84</td>
<td>86</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cattle</td>
<td>80</td>
<td>80</td>
<td>86</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dog</td>
<td>86</td>
<td>87</td>
<td>91</td>
<td>86</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse</td>
<td>68</td>
<td>69</td>
<td>70</td>
<td>69</td>
<td>69</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Rat</td>
<td>66</td>
<td>67</td>
<td>68</td>
<td>68</td>
<td>67</td>
<td>88</td>
<td>100</td>
</tr>
</tbody>
</table>

% identities were obtained by pair wise alignment using blast2sequences.
We have previously shown that DNA hypermethylation underlies the observed silencing of \textit{KLK6} expression in human cancer cells (Pampalakis and Sotiropoulou, 2006). More specifically, we found that the extent of cytosine methylation in a small non-typical-CpG island region \((-72, +15)\) that spans the transcriptional start site of P1 promoter correlates well with levels of expression of the (HUMAN)\textit{KLK6} gene. Alignment of P1 \textit{KLK6} sequences of human, chimpanzee, monkey, pig, dog and cattle indicated that the genomic regions that span the P1 transcriptional start site are conserved, as shown in Fig. 3B. The highest degree of conservation is observed between the human, chimpanzee and monkey sequences. Also, conserved are the seven CpG dinucleotides (marked by asterisks) that were found methylated in human \textit{KLK6} non-expressing cells. In addition, novel CpG sites are present in \textit{KLK6} upstream sequences in the dog, cattle and pig orthologue genes indicating that epigenetic mechanisms may be an alternative conserved mechanism for regulation of \textit{KLK6} gene expression in evolutionary higher organisms. High degree of similarity of nucleotides surrounding the transcriptional start site of (HUMAN)\textit{KLK6} (depicted by arrow in Fig. 3B) indicates that all other species probably exploit this sequence as a transcriptional initiation signal. However, EST clones that extend to this region were not found.

### 3.5. The \textit{KLK6} interactome

Fig. 4 depicts a simple flowchart for the IPA function, as well as a protein–protein interaction network that involves \textit{KLK6}, named the \textit{KLK6} interactome, as designed by IPA. The input proteins (\textit{KLK6}, vimentin, calreticulin, and keratin 8) used to construct the \textit{KLK6} interactome were retrieved from previous proteomic analysis in our laboratory (unpublished data). IPA transforms a list of genes into a set of relevant networks based on extensive records maintained in the ingenuity pathways knowledge base (IPKB). This tool can help biologists and bioinformaticians to understand how their genes of interest are biologically related either directly (by interacting with each other) or indirectly (by interacting with neighboring genes).
Moreover, the users can obtain a great amount of information on a single screen such as the kind of the biological relations between their genes of interest, other genes that are closely related with the user’s ones, etc.

In order to succeed that, the developers of IPA have designed a procedure that creates quickly relevant networks. The procedure consists of six steps and the final result is a graph that is as highly connected as possible. It uses simple methods derived from graph algorithmic techniques such as triangle connectivity, best “neighbor”, etc. that have low computational cost. There is a network size limitation (by default 35) so the results can be more useful for the researchers. In many cases, the selection of some genes that are included in the final network has been performed in a random fashion. Perhaps, additional criteria should be taken into consideration in order to produce better quality networks.

The data used in these functions come from the IPKB database, which contains a large volume of information about the genes. So, the success of the IPA tool is mainly based on the quality of the IPKB data and on the structure of the IPKB. In order to increase the quality, the IPA developers have included data manually obtained from literature research and only a small percentage of them (<10%) is automatically extracted and modeled. Furthermore, it uses a well-organized ontology that consists of a hierarchical data structure and gives IPA the opportunity to be fast. Details about the development of the algorithm can be obtained from Calvano et al. (2005) or can be downloaded from the help menu of the IPA program (http://www.ingenuity.com/).

The intermediate nodes and their role(s) in signal transduction were uncovered, as following: rat KLK6 (myelencephalon-specific protease or MSP) protein rapidly degrades rat plasma fibronectin (FN) to a polypeptide with an apparent molecular mass of 200 kDa, which is subsequently degraded to numerous smaller fragments (Blaber et al., 2002). FN is a glycoprotein present in a soluble dimeric form in plasma, and in a dimeric or multimeric form at the cell surface and in extracellular matrix. FN is involved in cell adhesion and migration processes including embryogenesis, wound healing, blood coagulation, host defence and metastasis. FN treatment of vascular smooth muscle cells significantly increases the mRNA expression of transforming growth factor-beta 1 (TGF-b1) (Hu et al., 2000). TGF-b1 is an extracellular growth factor, involved in cell transformation. TGF-b1 induces the expression of FN in various cell types (Yoshida et al., 1992; Pricci et al., 1996) it binds (Mooradian et al., 1989) and activates it (Grunet et al., 1993). TGF-b1 induces TGF-b1 mRNA levels in a time- and dose-dependent manner (Baldwin and Korc, 1993). TGF-b1 treatment of pancreatic cancer cells results in reduced cytokinin 8 (KRT8) expression but increases the expression of the mesenchymal marker vimentin (VIM) (Ellenrieder et al., 2001). KRT8, a simple epithelial keratin is a component of the intracellular cytoskeleton in the cells of the single-layered sheet tissues inside the body. It is a member of the intermediate filament...
family of proteins. VIM is an intermediate filament protein, which moves bi-directionally along microtubules in the cell. The level of the transcriptional regulator factor E2F1, is down-modulated in TGF-b1/vitamin D3-treated HL-60 cells (Peiretti et al., 2001; DeGregori et al., 1995). E2F1 plays a crucial role in the control of cell cycle and action of tumor suppressor proteins. E2F1 can also modulate the transcriptional level of TGF-b1 (Thatikunta et al., 1997). The gene encoding for cardelucin (CALR) is up-regulated by E2F1 (Haing et al., 2004). CALR, a Ca^{2+} binding protein, can act as an important modulator of the regulation of gene transcription by nuclear hormone receptors. Finally, it is interesting to note that it has been found recently that human KLK6 cleaves FN as well (Ghosh et al., 2004).

4. Discussion

Mapping and sequencing of genomes from a large number of species enabled us to identify and clone in silico orthologue genes of the (HUMAN)KLK6 gene. We have identified in total 13 mammalian KLK6 orthologues including novel transcript and splice variants and the putative KLK6 orthologue from D. melanogaster. The longest KLK6 cDNA transcript was assembled for every orthologue. Similarly to humans, KLK6 co-localized with all other kallikreins in a chromosomal locus in all mammalian genomes examined. All EST clones used in this analysis were identified in tissues known in humans to express (HUMAN)KLK6. Further, all genes contained five coding exons and encoded for putative secreted serine proteases with high identity to human KLK6 (>66% in rodents and >80% in other mammals). The codons that encode for the amino acid residues of the catalytic triad were found in conserved positions. Furthermore, all intron phases were fully conserved as for (HUMAN)KLK6 (I–II–I–0). The above data confirmed our original hypothesis that the identified genes are orthologues of (HUMAN)KLK6.

Based on the high similarity observed between KLK6 orthologues, it is very likely that mechanisms regulating their expression are well conserved among species. Indeed, similarly to humans, we have demonstrated presence of multiple transcript variants in monkey, cattle, mouse and rat that all encode for the same full-length protein, since they differ only in the 5′ UTR. Splice variants resulting from splicing out of coding exons were found in cattle. Specific splice variants may have differential expression in disease, as observed for (HUMAN)KLK13 in testicular cancers (Chang et al., 2001) and (HUMAN)KLK5 and (HUMAN)KLK7 in ovarian cancer (Dong et al., 2003). In addition, differential regulation of P1 and P2 (HUMAN)KLK6 promoters in inflammatory conditions of the central nervous system was demonstrated (Christophi et al., 2004). Alignment of genomic sequences from various mammals showed the presence of conserved regulatory motifs. Also, differential DNA methylation likely underlies the tissue-specific expression of KLK6, as it was demonstrated for maspin (Futcher et al., 2002). Here, we have found that positions of regulatory CpGs (Pampalakis and Sotiropoulou, 2006) are well conserved in the other animals, suggesting that methylation may play a significant role in determining the expression profile of KLK6.

The high similarity in the structure of KLK6 orthologues indicates a conserved function at the protein level. It was reported previously that MSP, the rat orthologue of human KLK6, is implicated in multiple sclerosis (Blaber et al., 2004). Recently, human KLK6 was implicated in multiple sclerosis and inflammations of the central nervous system (Iwata et al., 2003). In spinal cord injury, MSP and KLK6 are up-regulated at sites of trauma (Scarisbrick et al., 2006) and both can degrade components of the extracellular matrix (Blaber et al., 2002; Magklara et al., 2003). MSP is a trypsin-like enzyme, able to cleave myelin and extracellular matrix components (Blaber et al., 2002). Human KLK6 was similarly shown to degrade myelin (Bennett et al., 2002), while inhibition of its enzymatic activity improved the course of experimental multiple sclerosis (Blaber et al., 2004).

To predict potential function(s) of KLK6, a protein–protein interaction network of KLK6 was constructed. Although KLK6 is a secreted protein, here, we demonstrate for the first time that KLK6 is able to regulate the levels of two intracellular cytoskeletal proteins, VIM and KRT8 via the TGF-b1 signaling pathway (Fig. 4B). KLK6 rapidly and specifically degrades both laminin (LN) and fibronectin (FN). Decreased levels of FN result in decrease of TGF-b1. This results in an even further decrease of the expression of FN and, subsequently, TGF-b1 through a loop between the two genes. As TGF-b1 plays a suppressive role in the expression of KRT8 and E2F1 and an inducing role in the expression of VIM, decreased TGF-b1 allows for an increase of KRT8 and E2F1 expression and a decrease of VIM, respectively. Increased E2F1 results in up-regulation of CALR. Differential proteomic analysis showed that KLK6 expression affects the concentration of these protein substrates as predicted (our unpublished data).

According to the KLK6 protein–protein interaction network, KLK6 reduces the expression of FN, which in turn reduces the expression of TGF-b1. Thus, levels of KLK6 and TGF-b1 are inversely correlated. Further, TGF-b1 up-regulates the expression of VIM, which is found in aggressive forms of breast cancer in contrast to KLK6 that is down-regulated in the aggressive forms. Therefore, we expect that the treatment of breast cancer cells with TGF-b1 will inhibit the expression of KLK6 mRNA. In order to test whether the KLK6 interactome that we constructed here operates in vivo, we treated the T47D breast cancer cells with TGF-b1. We found that TGF-b1 caused a time-dependent down-regulation of KLK6 mRNA (our unpublished data). In conclusion, this study for the first time predicts a pathway for KLK6 that should be tested in cellular and animal models and may potentially be exploited for the design of therapeutic strategies.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.compbiolchem.2007.11.002.


