Comparative analysis of human and bovine protein kinases reveals unique relationship and functional diversity

Kabir, Nuzhat N.; Uddin, Kazi

Published in:
Genetics and Molecular Biology

DOI:
10.1590/S1415-47572011005000035

2011

Citation for published version (APA):
Comparative analysis of human and bovine protein kinases reveals unique relationship and functional diversity

Nuzhat N. Kabir1 and Julhash U. Kazi1,2*

1Laboratory of Computational Biochemistry, KN Biomedical Research Institute, Bagura Road, Barisal, Bangladesh.
2Quality Control Section, Opsonin Pharma Limited, Bagura Road, Barisal, Bangladesh.

Abstract

Reversible protein phosphorylation by protein kinases and phosphatases is a common event in various cellular processes. The eukaryotic protein kinase superfamily, which is one of the largest superfamilies of eukaryotic proteins, plays several roles in cell signaling and diseases. We identified 482 eukaryotic protein kinases and 39 atypical protein kinases in the bovine genome, by searching publicly accessible genetic-sequence databases. Bovines have 512 putative protein kinases, each orthologous to a human kinase. Whereas orthologous kinase pairs are, on an average, 90.6% identical, orthologous kinase catalytic domain pairs are, on an average, 95.9% identical at the amino acid level. This bioinformatic study of bovine protein kinases provides a suitable framework for further characterization of their functional and structural properties.

Key words: ePK, kinome, aPK, bovine.

Received: September 25, 2010; Accepted: June 1, 2011.

The protein kinase family is one of the largest families of proteins. Protein kinases play important roles in many intracellular or intercellular signaling pathways, resulting in cell proliferation, gene expression, metabolism, motility, membrane transport, apoptosis and differentiation. Furthermore, they modulate the activity of their substrate proteins by phosphorylating serine, threonine or tyrosine residues that mediate the activation, inhibition, translocation or degradation of substrate proteins (Brognard and Hunter, 2011).

Protein kinases are subdivided into two distinct superfamilies, referred to as eukaryotic protein kinases (ePKs) and atypical protein kinases (aPKs) (Hanks and Hunter, 1995). ePKs contain a conserved catalytic domain of approximately 250 amino acids. This domain is divided into 12 subdomains with highly conserved individual amino acids and motifs (Hanks et al., 1988). Within this domain, three motifs, ‘VAIK’, ‘HRD’ and ‘DFG’, are critical for the catalytic function, even though any residue from this region is fully conserved in all family members (Manning et al., 2002b). Conservation of these typical motifs is thought to be due to selection pressure for conserving important functions, such as the interaction with ATP, and the transfer of a phosphate group to the substrate. aPKs are functionally known to have kinase activity which lack significant sequence similarity to the ePK domain.

The sequencing of several vertebrate genomes has been completed. Initial estimates of the number of protein kinases in the human genome place this at around 1000 (Hunter, 1987), with later studies identifying 518 putative protein kinases (Manning et al., 2002b). The mouse and rat genomes contain 540 and 555 protein kinases, respectively, with 509 human orthologs (Caenepeel et al., 2004; Kazi et al., 2008), thereby implying possible functional conservation across species. In the present study, and by using sensitive bioinformatics approaches we identified the near complete set of bovine protein kinases. These were further classified into groups, families and subfamilies, based on the Hanks et al. (1988) and Manning et al. (2002b) classification scheme. This classification reveals many kinases that are conserved between bovine and human, thus reflecting functional constraints of these protein kinases in the core of signaling pathway. This study provides a suitable framework for further characterization of the functional and structural properties of these protein kinases.

A search was made of bovine proteome sequences available in GenBank (Benson et al., 2010) and Ensembl (Hubbard et al., 2009) for bovine protein kinases, using various tools. A preliminary search for protein kinases was performed using PSI-BLAST (Altschul et al., 1997), against the bovine proteome with an e-value threshold of 0.0001, and an h-value of 0.1 for five iterations. Previously published human (Manning et al., 2002b), mouse (Caenepeel et al., 2004) and rat (Kazi et al., 2008) eukaryotic protein kinases and kinase catalytic domains, as well as
eukaryotic protein kinase catalytic domains from a variety of organisms available at the kinase.com database were used as query sequences. A further search for protein kinases was performed using HMMER (Eddy, 1998). A Hidden Markov profile was created and validated by means of known eukaryotic protein-kinase catalytic domains. A further search was made for atypical protein kinases, using human, mouse and rat atypical protein kinases by PSI-BLAST or HMMER. Hits identified by using the different methods were combined, and duplicate records removed. Where splice variants were encountered, the variant showing either the closest proximity to the human ortholog, or the longest protein encoding variant, was recorded. All protein kinases were then evaluated for the presence of a conserved eukaryotic protein kinase domain (Hanks et al., 1988; Manning et al., 2002b). Catalytic domains were defined using RPS-BLAST in the BLAST package (Altschul et al., 1997) against the Pfam database (Finn et al., 2010), and sequence alignments carried out with AlinX implemented in the Vector-NTI package (Lu and Moriyama, 2004). Alignments were then manually edited, and all the kinases manually evaluated. Finally, 482 eukaryotic protein kinases and 39 atypical protein kinases were identified (Table S1). The primary names of the protein kinases were derived from their respective homologs in human (Manning et al., 2002b), mouse (Caenepeel et al., 2004) and rat (Kazi et al., 2008) protein kinases. On deriving a second name and synonyms from the Entrez Gene records (Maglott et al., 2005), the full protein names were retrieved hence. Representative records in Entrez Gene, corresponding to each bovine sequence, were identified, whereupon related information was included. Based upon the human protein kinases classification scheme (Manning et al., 2002b), these protein kinases were further classified into 10 groups, 129 families and 81 subfamilies (Table 1 and Table S1).

Previous studies have shown that almost all human protein-kinase orthologs are present in mouse and rat genomes (Caenepeel et al., 2004; Kazi et al., 2008). Thus, a search among bovine kinase sequences was conducted for the orthologous kinases using BLASTP (Altschul et al., 1997). The results were parsed, the symmetrically best hits being considered as orthologous kinases. The orthology relationships were further analyzed by CLUSTALW alignment (Thompson et al., 1997), followed by phylogenetic analysis. The latter was carried out by phylogenetic tree option incorporated into the CLUSTALW program. The NJ clustering algorithm was used for drawing bootstrap trees. As almost all bovine and human protein kinases exist as orthologous pairs (Figure S1), the similar functions in both organisms give to understand their evolutionary conservation. Human and bovine genomes contain 512 common protein kinase orthologs. Our search could not identify seven human protein kinase orthologs, likely due to the incomplete nature of bovine genome sequencing data (Table 2). In these seven human protein kinases, only TAF1L is absent in the chimpanzee genome. All are present in various higher eukaryotes, such as orangutans and monkeys. Ten protein kinases were absent in the human genome, eight of which being bovine specific and the other two, PLK5 and TSSK5 found in other genomes (Table 2). EphB1L might be a retrotransposed copy of the EphB1 gene, with 89.2% of amino acid sequence identity with the EphB1 protein.

Several proteins, such as ErbB3, SCYL1 and KSR1, have an inactive catalytic domain (Citri et al., 2003; Manning et al., 2002a). These inactive kinases, besides acting mainly as adaptor proteins, or dimerizing with active kinases, have also been shown to be involved in various cellular functions (Salerno et al., 2005; Schmidt et al., 2007; Sergina et al., 2007). Three conserved motifs, ‘VAIK’, ‘HRD’ and ‘DFG’, are important for catalytic activities. Inactive kinases lack at least one of these three conserved motifs. Fifty catalytic domains and 45 protein kinases in the human genome were predicted as catalytically inactive due to the lack of at least one of the three conserved residues (Manning et al., 2002b). The bovine complement of inactive kinases is equivalent to that of the human (Table S1).

All the human and bovine orthologous protein kinase pairs and orthologous catalytic domain pairs were analyzed for the percentage of identity by AlignX incorporated into Vector-NTI (Lu and Moriyama, 2004). In protein sequence alignments of orthologous kinase pairs we observed a wide variation in local sequence conservation (Figure 1A). These were, on an average, 90.6% identical (amino acid se-

---

**Table 1** - Protein-kinases and their distribution by groups in different genomes.

<table>
<thead>
<tr>
<th>Species</th>
<th>AGC</th>
<th>CAMK</th>
<th>CK1</th>
<th>CMGC</th>
<th>Other</th>
<th>RGC</th>
<th>STE</th>
<th>TK</th>
<th>TKL</th>
<th>Atypical</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>63</td>
<td>74</td>
<td>12</td>
<td>61</td>
<td>83</td>
<td>5</td>
<td>47</td>
<td>90</td>
<td>43</td>
<td>40</td>
<td>518</td>
</tr>
<tr>
<td>Bovine</td>
<td>62</td>
<td>75</td>
<td>12</td>
<td>61</td>
<td>85</td>
<td>5</td>
<td>47</td>
<td>91</td>
<td>44</td>
<td>39</td>
<td>521</td>
</tr>
<tr>
<td>Rat</td>
<td>60</td>
<td>112</td>
<td>11</td>
<td>61</td>
<td>84</td>
<td>7</td>
<td>46</td>
<td>91</td>
<td>43</td>
<td>40</td>
<td>555</td>
</tr>
<tr>
<td>Mouse</td>
<td>60</td>
<td>97</td>
<td>11</td>
<td>60</td>
<td>84</td>
<td>7</td>
<td>47</td>
<td>90</td>
<td>43</td>
<td>39</td>
<td>538</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>30</td>
<td>32</td>
<td>10</td>
<td>33</td>
<td>45</td>
<td>6</td>
<td>18</td>
<td>32</td>
<td>17</td>
<td>17</td>
<td>240</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
<td>30</td>
<td>46</td>
<td>85</td>
<td>49</td>
<td>67</td>
<td>27</td>
<td>25</td>
<td>90</td>
<td>15</td>
<td>20</td>
<td>454</td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
<td>17</td>
<td>21</td>
<td>4</td>
<td>21</td>
<td>38</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td>130</td>
</tr>
</tbody>
</table>

---

Kabir and Kazi
### Table 2 - Lineage-specific protein kinases.

<table>
<thead>
<tr>
<th>Found in</th>
<th>Gene Name</th>
<th>Introns</th>
<th>Paralog</th>
<th>Other orthologs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>CK1α2</td>
<td>No</td>
<td>CK1α1 (91%)</td>
<td>Chimpanzee, Orangutan, Monkey</td>
</tr>
<tr>
<td></td>
<td>TAF1L</td>
<td>No</td>
<td>TAF1 (94%)</td>
<td>Orangutan, Monkey</td>
</tr>
<tr>
<td></td>
<td>PKAαg</td>
<td>No</td>
<td>PKAαa (83%)</td>
<td>Chimpanzee, Orangutan</td>
</tr>
<tr>
<td></td>
<td>LTK</td>
<td>Yes</td>
<td>-</td>
<td>Chimpanzee, Orangutan, Rat, Mouse, Horse, Chicken</td>
</tr>
<tr>
<td></td>
<td>ROS</td>
<td>Yes</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PSKH2</td>
<td>Yes</td>
<td>-</td>
<td>Chimpanzee, Monkey, Dog, Orangutan, Rabbit, Horse</td>
</tr>
<tr>
<td></td>
<td>PRKY</td>
<td>Yes</td>
<td>PRKX (92%)</td>
<td>Chimpanzee, Gorilla, Monkey</td>
</tr>
<tr>
<td>Bovine</td>
<td>AURL</td>
<td>Yes</td>
<td>AurB (61%)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CK1b</td>
<td>Yes</td>
<td>CK1α (81%)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>EphB1L</td>
<td>No</td>
<td>EphB1 (89.2%)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>FERL</td>
<td>Yes</td>
<td>FER (82%)</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MARKL</td>
<td>Yes</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>MASTL2</td>
<td>Yes</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Par-1L</td>
<td>Yes</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>PLK5</td>
<td>Yes</td>
<td>-</td>
<td>Orangutan, Monkey, Mouse, Rat</td>
</tr>
<tr>
<td></td>
<td>RYKL</td>
<td>Yes</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>TSSK5</td>
<td>Yes</td>
<td>-</td>
<td>Dog, Horse, Mouse, Rat</td>
</tr>
</tbody>
</table>

In ‘Paralog’ parentheses show % of identity (amino acid) with respective paralog.

**Figure 1** - Conservation within orthologous kinases (A) and catalytic domains (B) is family-dependent. Triangles indicate mean identity within selected families, and bars indicate the range.
quence), although some were as low as 47.7%, with four pairs presenting high levels of sequence identity throughout the protein (Table S1). Although most differences between orthologs are due to amino acid substitution, many proteins contained substantial insertions or deletions (indels) between orthologs, which may account for many of the functional differences between species. Several proteins contained insertions or deletions (indels), as shown by sequence alignment (Table S1). These comparisons are also informative within the conserved domains. Although orthologous catalytic domains were, on an average, 95.9% identical, some were as low as 63.8%. Sixty two pairs were identical across the full domain, whereas 48 differed by only one amino acid (Table S1), thus indicative of strong conservative pressure throughout the catalytic domain. Catalytic domain pairs showed clearly family-dependent variability (Figure 1B). For example, of the six casein kinase 1 (CK1) family domain pairs, three were identical, and the other three differed by two residues, an average difference of only 0.4%, thereby indicating that changes in almost any amino acid within the domain destroyed some function, and thus have been eliminated by evolution (Figure 2A). At the other extreme, PEK family catalytic domain pairs are 70-88% identical, thereby implying that the core functions of this family of kinases do not greatly constrain the domain sequence (Figure 2B).

Our study presents a bioinformatic overview and evolutionary insight into the kinases within the bovine genome. Comparison with the human kinome revealed the evolutionary conservation of the protein kinase function. The curated kinase dataset from the bovine genome, presented here, could serve as a framework for further investigation of this important gene family.

Acknowledgments

This work was supported by a KN Biomedical Research Grant. We are grateful to Nadid Ahnaf Kazi.

References


Internet Resources

Supplementary Material
The following online material is available for this article:
Table S1 - List of bovine kinases with sequences.
Figure S1 - Phylogenetic trees for human and bovine kinases.
This material is available as part of the online article from http://www.scielo.br/gmb.

Associate Editor: Alexandre Rodrigues Caetano

License information: 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 work is properly cited.