The genomic structure and chromosomal localization of the mouse STAT3 gene

Wei Shi, Masahiro Inoue, Masashi Minami, Kiyoshi Takeda, Makoto Matsumoto, Yoichi Matsuda, Tadamitsu Kishimoto and Shizuo Akira

Division of Immunology, Institute for Molecular and Cellular Biology, Osaka University, 1-3 Yamadaoka, Suita, Osaka 565, Japan

1Department of Internal Medicine III, Osaka University Medical School, 2-2 Yamadaoka, Suita, Osaka 565, Japan

2Genome Research Group, National Institute of Radiological Sciences, 4-9-1 Anagawa, Inage-ku, Chiba, Chiba 263, Japan

3Present address: Laboratory of Animal Genetics, School of Agricultural Sciences, Nagoya University, Chikusa-ku, Nagoya, Aichi 464-01, Japan

4Present address: Department of Biochemistry, Hyogo College of Medicine, 1-1 Mukogawacho, Nishinomiya, Hyogo 663, Japan

Keywords: cytokine, genomic DNA, IL-6, signal transduction, promoter region

Abstract

A variety of cytokines induce the tyrosine phosphorylation of signal transducers and activators of transcription (STATs). Activation of the same STAT proteins by distinct cytokines and activation of different STAT proteins by each cytokine are thought to contribute to redundancy and pleiotropy of cytokine actions respectively. STAT3 is rapidly tyrosine phosphorylated in response to IL-6, ciliary neurotrophic factor, oncostatin M, leukemia inhibitory factor, IL-11, granulocyte colony stimulation factor and epidermal growth factor. In this report we have isolated and characterized the mouse genomic structure of STAT3. The mouse STAT3 gene consisted of 24 exons which spanned >37 kb. The structure of the mouse STAT3 gene was almost identical to that of the human STAT2 gene, including the number and size of exons, indicating that the exon–intron organization had already been accomplished before these two genes duplicated, and then these genes evolved to respond to different ligands. By molecular linkage analysis with interspecific backcross mice the STAT3 gene mapped at 1.4 cM proximal to D11Mit59 on mouse chromosome 11. The promoter region contained potential regulatory elements such as GATA, NF-IL-6, PEBP2, Sp-1, AP-2 binding sites, cAMP response element, CAAT box and E-box. Transient expression of constructs harboring the 5' flanking region of the STAT3 gene fused to the luciferase gene showed that a 160 bp sequence upstream of the transcription start site conferred a basal and an IL-6-inducible promoter activity.

Introduction

STAT proteins (signal transducers and activators of transcription) are latent transcription factors that contain src homology 2 (SH2), SH3-like domains and a C-terminal tyrosine phosphorylation site (1). These proteins become activated by phosphorylation on a specific tyrosine residue in response to a variety of cytokines and growth factors (2,3). The tyrosine phosphorylation of cytoplasmic STATs is mediated by the Janus protein kinase (JAK) family, which currently consists of JAK1, JAK2, JAK3 and Tyk2. In the case of IL-6 and ciliary neurotrophic factor (CNTF) which share signal transducing receptor component gp130, JAK1, JAK2 and Tyk2 can be activated, although the selection of the kinases varies depending on the cell line. In the absence of JAK1, the phosphorylation of gp130 and the activation of STAT3 were greatly reduced, suggesting that JAK1 is essential for the gp130-mediated signaling pathway (4).
Mouse STAT3 gene

A 2.0 kb
Exon 1 2 3 4 5-11 12-15 16-21 22-24

B
E X X X E H E E X X H H

C 0.2 kb
ATG LZ DNA SH3 SH2 Y TGA
Exon 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Fig. 1. Physical map of the mouse STAT3 gene (A) The structure of the STAT3 gene is represented with regard to the organization of the exons and introns. (B) The restriction map shows cleavage sites for restriction enzymes, XbaI (X), EcoRI (E) and HindIII (H) (C) The exons are scaled in mouse STAT3 cDNA. Translation initiates in exon 2. STAT3 leucine zipper domain (between amino acids 203 and 224) is scattered in exon 7 and 8, SH3 domain is in exon 17, 18 and 19, SH2 domain is split into exon 19, 20 and 21. Termination codon is in exon 24.

Fig. 2. Comparison of exon length of STAT3 and STAT2. Amino acid numbers in each exon of mouse STAT3 and human STAT2 were compared.

Unstimulated STAT proteins are present in monomeric forms. When cells are stimulated with ligands, STATs associate with phosphorylated tyrosine residues of an activated receptor complex through the SH2 domain of STATs and become phosphorylated on the tyrosine residues by JAK kinases. The phosphorylated STATs then form homo- or heterodimers, and rapidly translocate to the nucleus, where they bind specific DNA sequences and regulate the transcription of the target genes (5). At present, six different members of STATs (STAT1 to STAT6) have been cloned and are involved in specific cytokine regulation (6-12).

In response to IFN-α/β, STAT1 and STAT2 are tyrosine phosphorylated, whereas STAT1 is tyrosine phosphorylated in response to IFN-γ (13,14). STAT3 is predominantly phosphorylated in IL-6, leukemia inhibitory factor (LIF), CNTF, oncostatin M and IL-11 that share gp130 (15). In addition, granulocyte colony stimulating factor (G-CSF) (16), which utilizes a receptor highly related to gp130, also induces phosphorylation of STAT3. STAT4 and STAT6 are phosphorylated by IL-12 and IL-4 respectively. STAT5 is phosphorylated by a variety of cytokines including prolactin, IL-2, IL-3, IL-5 and granulocyte macrophage colony stimulating factor.

Recently it has been shown that selectivity of STAT activation is determined by a specific docking interaction between STATs and the phosphorylated tyrosine-containing sequences in the signal-transducing receptor components. STAT1 docks to the sequence YDKPH in the IFN-γ receptor, while STAT3 docks to a consensus sequence YxxQ in the gp130 and LIF receptor cytoplasmic domains (17,18). In addition to the obligatory requirement of tyrosine phosphorylation for dimerization and DNA binding, phosphorylation of a C-terminal serine residue is required for maximal transcriptional activation of STATs (19).

The complete genomic structure of human STAT2 and
Table 1. Exon–intron and junction sequences of the mouse STAT3 gene

<table>
<thead>
<tr>
<th>Exon No.</th>
<th>Size (base)</th>
<th>Junction (cDNA position)</th>
<th>Intron No.</th>
<th>Size (kb)</th>
<th>Junction (cDNA position)</th>
<th>Exon No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>235</td>
<td>GACAGgtgaa</td>
<td>1</td>
<td>&gt;15</td>
<td>tacagTGCAG</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>152</td>
<td>GACTGgtgag</td>
<td>2</td>
<td>2.50</td>
<td>cctagGGCAG</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>144</td>
<td>TGCAGgtgcc</td>
<td>3</td>
<td>0.54</td>
<td>tacagAGCAG</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>99</td>
<td>CCCAGgtgccc</td>
<td>4</td>
<td>1.88</td>
<td>ggcaGCAAGG</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>96</td>
<td>TGCAGgtaat</td>
<td>5</td>
<td>0.20</td>
<td>tgcagGATCCT</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>72</td>
<td>AGGAGgttaag</td>
<td>6</td>
<td>0.30</td>
<td>cttagACATG</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>95</td>
<td>GGAGAgtaag</td>
<td>7</td>
<td>0.14</td>
<td>aacagAGCAT</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>152</td>
<td>AACTGgttaag</td>
<td>8</td>
<td>0.51</td>
<td>ggcaGATAAA</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>159</td>
<td>AAGAGgttaac</td>
<td>9</td>
<td>0.10</td>
<td>tccagTGCTCT</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>93</td>
<td>GTCAAGgtgagg</td>
<td>10</td>
<td>0.16</td>
<td>aataGTTGTC</td>
<td>11</td>
</tr>
<tr>
<td>11</td>
<td>59</td>
<td>GATAAGtaag</td>
<td>11</td>
<td>4.50</td>
<td>tacaagAGACT</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>31</td>
<td>AGAGGgttaaa</td>
<td>12</td>
<td>0.09</td>
<td>ttaagGTCTCT</td>
<td>13</td>
</tr>
<tr>
<td>13</td>
<td>94</td>
<td>ACCTGgtcgg</td>
<td>13</td>
<td>0.11</td>
<td>tccagACCTCT</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>48</td>
<td>GTGATgtaag</td>
<td>14</td>
<td>0.43</td>
<td>tgcagGCTCT</td>
<td>15</td>
</tr>
<tr>
<td>15</td>
<td>84</td>
<td>TAGAGgtgag</td>
<td>15</td>
<td>2.10</td>
<td>tacaagGTCGGA</td>
<td>16</td>
</tr>
<tr>
<td>16</td>
<td>99</td>
<td>CCAAGgtgag</td>
<td>16</td>
<td>0.30</td>
<td>aataAAGCCT</td>
<td>17</td>
</tr>
<tr>
<td>17</td>
<td>136</td>
<td>CCTAGgtttcc</td>
<td>17</td>
<td>0.50</td>
<td>tccagGCCCT</td>
<td>18</td>
</tr>
<tr>
<td>18</td>
<td>53</td>
<td>GCAAGgtgaa</td>
<td>18</td>
<td>0.14</td>
<td>tccagGAAA</td>
<td>19</td>
</tr>
<tr>
<td>19</td>
<td>95</td>
<td>GAAGGgttaag</td>
<td>19</td>
<td>0.07</td>
<td>tgcagGTCAC</td>
<td>20</td>
</tr>
<tr>
<td>20</td>
<td>140</td>
<td>CAGTGgttaag</td>
<td>20</td>
<td>0.48</td>
<td>gacagGCAAG</td>
<td>21</td>
</tr>
<tr>
<td>21</td>
<td>213</td>
<td>AGGTGgtttg</td>
<td>21</td>
<td>1.34</td>
<td>tgcagGTCGCT</td>
<td>22</td>
</tr>
<tr>
<td>22</td>
<td>43</td>
<td>ACACGgttaag</td>
<td>22</td>
<td>0.48</td>
<td>tacaAGACGA</td>
<td>23</td>
</tr>
<tr>
<td>23</td>
<td>113</td>
<td>GTTGTgtqaq</td>
<td>23</td>
<td>0.40</td>
<td>tgcagAGTCG</td>
<td>24</td>
</tr>
<tr>
<td>24</td>
<td>137</td>
<td>ATATCgttccc</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

partial genomic structure of human STAT1 have been demonstrated (20). Here we report the complete mouse genomic organization of STAT3 and chromosomal location as well as the analysis of the promoter region of STAT3.

Methods

Genomic library screening

A mouse genomic DNA library derived from 129Sv liver in λFIX II (Stratagene, La Jolla, CA) was screened with the 32p-labeled mouse STAT3 cDNA probe. Four different parts of STAT3 cDNA (5'-untranslated region, 5'-untranslated region, and amino acids 1-28, 280-440 and 626-770) were used for probe. Filters blotted with recombinant phage DNA were incubated in hybridization buffer (2.5xDenharsts, 1% SDS, 10 mM EDTA, pH 8.0, 100 µg/ml of salmon sperm DNA, 6xSSC) with the labeled probe for 16 h at 65°C and then washed with 0.1XSSC/0.1% SDS, and autoradiographed.

Subcloning and sequencing

To identify genomic fragments containing exons, bacteriophage DNA purified from each clone was digested by several restriction enzymes, separated by electrophoresis on 0.6% agarose gels and blotted on GeneScreen Plus (NEN Research Products, Boston, MA). The blot was hybridized with 32p-labeled STAT3 cDNA. Appropriate restriction fragments were subcloned into the plasmid vector, pBluescript KS+ (Stratagene, La Jolla, CA). Sequencing was performed by using the Sequenase Version 2.0 DNA System (Amersham, Arlington Heights, IL) or The ABI 373A automated DNA sequencer (Foster City, CA) according to the manufacturer’s instructions.

Primer extension analysis

The transcription initiation site was determined by a primer extension system (Promega, Madison, WI). The primer used for the reaction was a 35-mer oligonucleotide (5'-GGAA-GCCACGGTTCCCGCTGCTCAGTCA-3') complementary to the sequence in exon 1 of the mouse STAT3 gene, and was radiolabeled at the 5' end with T4 polynucleotide kinase and [γ-32P]ATP (3000 Ci/mmol). The labeled primer was added to 25 µg of total RNA isolated from HepG2 cells or mouse liver in the presence of 50 mM Tris–HCl, pH 8.3, 50 mM KCl, 10 mM MgCl2, 10 mM dithiothreitol and 0.5 mM spermidine. The reaction mixtures were incubated at 60°C for 30 min and annealed at 30°C for 2 h. Avian myeloblastosis virus reverse transcriptase (1 U) was added and primer extension was performed at 42°C for 105 min. The products were electrophoresed on a 6% polyacrylamide gel containing 7 M urea and the length of the primer-extended product was determined by comparison with a dideoxynucleotide sequencing of the promoter region performed with an oligonucleotide complementary to exon 1.
Within the carboxy region a potential mitogen-activated protein kinase site is found and phosphorylation of this site has been shown to be involved in the transcriptional activation. In the case of STAT1, a splice variant (STAT1P) lacks the C-terminal domain of the STATs is highly divergent. Further studies are needed to elucidate the mechanism of STAT3 gene expression.

In conclusion, we succeeded in the isolation and characterization of the mouse STAT3 gene. Information of the exact exon–intron boundary of STAT3 obtained in this study will be helpful for screening novel splicing isoforms of STAT3 messenger in certain physiological and pathological conditions.

Discussion

We have isolated and characterized the mouse STAT3 gene as well as the chromosomal localization. The mouse STAT3 gene consisted of 24 exons which spanned >37 kb long. The STAT3 gene mapped on mouse chromosome 11. When compared between the human STAT2 and mouse STAT3 genes, the exon–intron junction and the length of exons were almost identical, inspite of 38% amino acid homology between these proteins. As conserved exonic organization of STAT1 and STAT2 genes suggested (20), this result may indicate that the exon–intron organization had already been accomplished before these two genes duplicated, and then these genes evolved to respond to different ligands. During preparation of this manuscript, Copeland et al. (24) showed that the mouse STAT3 gene mapped to the distal region of chromosome 11 as demonstrated in our paper.

The C-terminal domain of the STATs is highly divergent. Within the carboxy region a potential mitogen-activated protein kinase site is found and phosphorylation of this site has been shown to be involved in the transcriptional activation. In the case of STAT1, a splice variant (STAT1P) lacks the C-terminal 38 amino acids including the MAPK site (25). This variant can be phosphorylated, translocate to the nucleus and bind to DNA, but cannot activate gene transcription. Similar variants have been identified for STAT2, STAT3 and STAT5 (26–28). In the case of STAT3, a variant named STAT3P lacks the C-terminal 55 amino acids which are replaced by seven other amino acids. Our present exon–intron boundary analysis show that this variant is an alternative splicing form in which exon 22 and an as yet unidentified exon are combined.

Although the nucleotide sequence around the transcription initiation site does not contain a TATA box-like sequence, the 5′-flanking region of the mouse STAT3 gene contained several potential cis-acting regulatory sequences such as binding sites for GATA, CRE, NF-IL-6, E-box, PEBP2, Sp-1, Pit-1 and AP-2. We previously observed that STAT3 gene expression was augmented in hepatocytes in response to IL-6. We demonstrated in this study that the 160 bp region from +38 to −165 conferred the IL-6 responsive promoter activity in a hepatoma cell line. At present two IL-6 responsive elements, a C/EBP binding site and a STAT3 binding site, are identified from the analysis of acute phase protein gene regulation (29). However, within the 160 bp region we could not find these binding motifs. This may indicate that other transcription factors are involved in the regulation of the IL-6-mediated STAT3 induction. Since STAT3 is expressed in embryonic stem cells which do not contain many transcription factors, it is also interesting to examine if there are any differences in the transcription factors involved in STAT3 expression in embryonic stem cells and in hepatocytes. Further studies are needed to elucidate the mechanism of STAT3 gene expression.

In conclusion, we succeeded in the isolation and characterization of the mouse STAT3 gene. Information of the exact exon–intron boundary of STAT3 obtained in this study will be helpful for screening novel splicing isoforms of STAT3 messenger in certain physiological and pathological conditions.
Acknowledgements
We thank Ms K. Kubota for her assistance and Dr T. Sugiyama for providing us with total RNA of HepG2.

Abbreviations
- CNTF: ciliary neurotrophic factor
- EGF: epidermal growth factor
- G-CSF: granulocyte colony-stimulating factor
- GRE: glucocorticoid responsive elements
- G-CSF: granulocyte colony stimulating factor
- CNTF: ciliary neurotrophic factor
- SH2: Src homology 2
- JAK: Janus protein kinase
- RFLV: restriction fragment length variant
- SHP-2: Src homology 2
- STAT: signal transducers and activators of transcription

References