Cloning of a novel transcription factor-like gene amplified in human glioma including astrocytoma grade I

Ulrike Fischer1, Dirk Heckel1, Armin Michel1, Michael Janka2, Theo Hulsebos3 and Eckart Meese1,*

1Department of Human Genetics, Bau 68, Medical School, University Hospital, University of Saar, D-66421 Homburg/Saar, Germany, 2Department of Neurosurgery, Medical School, University Hospital, Josef-Schneider-Straße 11, 97080 Würzburg, Germany and 3Department of Human Genetics, AMC, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands

Received November 6, 1996; Revised and Accepted July 9, 1997

Gene amplification, which is generally considered to occur late in tumor development, is a common feature of high grade glioma. Up until now, there have been no reports on amplification in astrocytoma grade I. In this study, we report cloning and sequencing of a cDNA termed glioma-amplified sequence (GAS41) which was identified recently in a glioblastoma cell line by microdissection-mediated cDNA capture. This technique is tailored to isolate amplified genes from human tumors. An increased copy number of GAS41 was found in glioblastoma multiforme and astrocytoma III, and at a high frequency in astrocytoma grades I and II. Sequence comparison indicates a high homology between the GAS41 protein, the yeast and human AF-9 and the human ENL proteins. Both AF-9 and ENL belong to a new class of transcription factors, indicating that GAS41 might also represent a transcription factor. With GAS41 being the first gene found with increased copy number in low grade glioma, this study provides the first evidence that gene amplification can occur in early tumor development.

INTRODUCTION

Gene amplification occurs frequently in human tumors and was proposed to contribute to tumorigenesis. Specifically, there is ample evidence for a prevalent role of amplification in human gliomas. Astrocytic gliomas are the most common human tumors of the central nervous system. According to the World Health Organization (WHO), astrocytomas were divided into four groups of different grades: pilocytic astrocytoma WHO grade I, astrocytoma WHO grade II, anaplastic astrocytoma WHO grade III and glioblastoma multiforme (GBM) WHO grade IV. Southern blot analysis revealed amplification of several genes including EGFR, TGFRα, PDGFRα, GLI, MET, MDM2, SAS, CDK4 and N-MYC in human glioma (1–6). Comparative genomic hybridization identified amplifications at chromosome regions 1q32.1, 5p15.1, 9p2, 9q12–13, 7q21, 12p and 22q12 (7–10).

As of yet, however, gene amplifications in gliomas seem to be restricted mostly to GBM and, at a lower frequency, to anaplastic astrocytoma. Independent studies on pilocytic astrocytoma and astrocytoma WHO grade II failed to reveal amplification of EGFR, PDGFRα and genes from chromosome region 12q13–15, including MDM2, CDK4, SAS, GLI, A2MR, WNT1, GADD153 and ERBB3 (1,2,5,11). Previously, we reported a single case of MET amplification in a grade II astrocytoma (12).

To gain further insight into the biology of amplifications in glioma, we recently identified several cDNAs from a homogeneously staining region (HSR) at 12q13–15 by microdissection-mediated cDNA capture (13). cDNA was generated from tumor cells, hybridized against metaphases of the tumor cells and recovered by microdissection of the HSR. Bound cDNAs were amplified by PCR and mapped to the amplified domain at 12q13–15.

Here, we set out to isolate and characterize the corresponding cDNA (GAS41) of a microdissected cDNA fragment which maps within the amplicon at 12q13–15. Sequence comparison was performed and the gene GAS41 was tested for amplification in high and low grade gliomas, including pilocytic astrocytoma.

RESULTS

Cloning and sequencing GAS41 cDNA

Plasmid clone pGAS41 containing a 220 bp insert (GenBank accession No. U61384) was isolated previously from a homogeneously staining region of glioblastoma TX3868 (13). To isolate the corresponding cDNA, we established a λZAP Express (Stratagene) cDNA library from glioblastoma cell line TX3868. Using the partial cDNA fragment of pGAS41 (220 bp) as probe, positive phage clones (two positive clones per 10 000 p.f.u.) were identified and in vivo excised into plasmid clones. The newly isolated cDNAs were digested with EcoRI and XhoI and reconfirmed by Southern hybridization using the pGAS41 insert as probe. Sequencing of clone GAS41 identified the 220 bp
Figure 1. (A) Alignment of the putative protein sequence of GAS41 with homologous database sequences. Identical amino acids are inverted, similar amino acids are shaded. Gaps inserted by the sequence alignment software are indicated by ‘–’. GAS41, amino acid sequence of the putative protein deducted from the GAS41 cDNA sequence; g1183966, amino acid sequence of a *Saccharomyces cerevisiae* protein similar to human AF-9 (GenBank accession No. gi1183966); AF9, amino acid sequence of the human AF-9 protein. (B) Predicted amino acid sequence of GAS41 with acidic amino acids shown in bold letters and regions with a high probability of forming α-helical structures underlined.

Sequence used for hybridization and revealed an open reading frame (ORF) coding for a protein of 227 amino acids.

**GenBank homology search and protein sequence analysis**

Homology searches of *GAS41* cDNA were performed using both BLASTX and BLASTN algorithms. BLASTN search revealed homology (62%) to the *Saccharomyces cerevisiae* AF-9 gene. BLASTX search revealed high homologies to several proteins including *S.cerevisiae* AF-9 protein, with 53% identity and 80% similarity, human AF-9 protein, with 39% identity and 63% similarity, and human ENL protein, with 39% identity and 63% similarity. AF-9 and ENL proteins were supposed to belong to a new class of transcription factors, suggesting a possible role for the newly isolated GAS41 (14) in transcription regulation.

The function of GAS41 as a transcription factor is supported by the finding of an enrichment of acidic amino acids (27%) found within a stretch of 60 amino acids in the C-terminal region. Protein structure-predicting programs indicate a high probability of this region forming α-helices. Negatively charged α-helical structures were the first defined activation domains in eukaryotic transcription factors. (15). The predicted amino acid sequence of GAS41 is shown in Figure 1B.

**Amplification analysis**

To gain further insight into the role of GAS41 amplifications in human glioma, we investigated 64 gliomas for gene amplification including 52 glioblastomas, four anaplastic grade III astrocytomas, two grade II astrocytomas and six pilocytic grade I astrocytomas. The tumors were analyzed preferentially by comparative PCR (16) using primers located in the 5′–untranslated region of *GAS41* cDNA and MUC2 primers (18) as a single copy control. To confirm the data obtained by comparative PCR, we analyzed selected tumors by Southern hybridization using *GAS41*, *SAS*, *CDK4* and *WNT1* as probes. In the majority of cases, however, the limited amounts of DNA were not sufficient for Southern blot analysis. Furthermore, probes for *GAS41* did not hybridize efficiently to single or low copy targets.

Representative amplification analysis by comparative PCR is shown in Figure 2. As positive control in our comparative PCR screening, we used DNA from glioblastoma TX3868 (Fig. 2A) which contains a *GAS41* amplification as previously demonstrated by Southern hybridization (13). Comparative PCR detected *GAS41* amplifications in gliomas of all grades, including glioblastoma as shown in Figure 2A, anaplastic astrocytoma as shown in Figure 2B and in pilocytic astrocytoma as shown in Figure 2C. Interestingly, the *MDM2* gene, which is also localized at 12q13–15, was not amplified in pilocytic astrocytoma as determined by comparative PCR (Fig. 2D).

As demonstrated for a subset of glioblastomas, the *GAS41* amplifications were also identifiable by Southern blot analysis. The same subset of glioblastomas showed co-amplification of the genes *SAS* and *CDK4* but did not reveal *WNT1* amplification. Examples of the amplification analysis by Southern blot hybridization are given in Figure 3.

In total, our study revealed *GAS41* amplification in 20 of 64 tumors. Specifically, *GAS41* amplification was detected in 12 out of 52 glioblastomas (23%), two out of four grade III anaplastic astrocytomas (50%), one out of two grade II astrocytomas (50%) and five out of six (80%) pilocytic astrocytomas. Although comparative PCR does not allow a quantitative comparison of
Figure 2. Amplification analysis of GAS41 in gliomas of different origin using comparative PCR. Three different concentrations of tumor and blood DNA were each amplified by PCR using primers specific for GAS41 (lower panel) and MUC2 (upper panel) as single copy control. PCR product sizes are indicated by arrows. (A) DNA from peripheral blood lymphocytes (PB), from glioblastoma multiforme (GBM) TX3868 with known amplification as positive control and from glioblastoma multiforme GBM9. Amplification of GAS41 was detected in TX3868 and GBM9. (B) DNA from PB and from anaplastic astrocytoma AA40. Amplification of GAS41 was detected in AA40. (C) DNA from PB and from pilocytic astrocytomas PA11, PA18, PA23, PA31, PA43 and PA61. Amplification of GAS41 was detected in PA11, PA18, PA31, PA43 and PA61. PA23 shows no GAS41 amplification. (D) Amplification analysis of MDM2 in pilocytic astrocytomas. DNA from PB and from pilocytic astrocytomas PA61 and PA11 was amplified by PCR using primers specific for MUC2 and MDM2. Amplification of MDM2 was not detectable in PA11 and PA61.

amplification levels, our data indicate that the level of GAS41 gene amplification is lower in pilocytic astrocytoma than in anaplastic astrocytoma and glioblastoma (Fig. 2). Results of the amplification analysis are summarized in Table 3.

Table 3. GAS41 amplification in gliomas of different grades

<table>
<thead>
<tr>
<th>Tumor</th>
<th>Total no.</th>
<th>No. of amplified cases</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBM</td>
<td>52</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>AIII</td>
<td>4</td>
<td>2</td>
<td>50</td>
</tr>
<tr>
<td>AII</td>
<td>2</td>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>PA</td>
<td>6</td>
<td>5</td>
<td>80</td>
</tr>
<tr>
<td>Total</td>
<td>64</td>
<td>20</td>
<td>31</td>
</tr>
</tbody>
</table>

DISCUSSION

Sequence analysis and comparison of GAS41 with known protein sequences revealed high homology of GAS41 protein to yeast and human AF-9 protein and human ENL protein. The AF-9 and ENL genes encode proteins which are involved frequently in translocation events and were suggested to belong to a new class of transcription factors (14). Both proteins share extensive sequence homologies in the N-terminal region and several common protein motifs including a proline-rich domain in the N-terminal overlapping sequence. Interestingly, the region of GAS41 with the highest sequence homology corresponds to the N-terminal proline-rich regions of AF-9 and ENL. This finding indicates a possible similar biological function of GAS41 and the transcription factors AF-9 and ENL.

The amino acid sequence of GAS41 contains a much higher percentage (27%) of acidic amino acids in the C-terminal region than on average found in proteins (11.7%). Significant negatively charged acidic α-helical structures were the first defined activation domains in eukaryotic transcription factors. Acidic activation domains were proposed to interact in a relatively non-specific manner with components of the initiation complex (15). In GAS41, a stretch of 40 amino acids starting at position 136 meets the criteria for an acidic activation domain. Besides 22% of acidic...
amino acids, the region is negatively charged, and protein structure prediction programs indicate α-helical structures.

In contrast to AF-9 and ENL, GAS41 lacks a typical DNA-binding domain for transcriptional activation (15). It is conceivable that GAS41 protein activates transcription in combination with a second protein containing a DNA-binding domain. Herpes simplex virus trans-activator VP16 is an example of transcription regulation by protein–protein interactions. The VP16 protein activates transcription of early viral genes by attachment to host-encoded proteins that recognize DNA sequences of viral promoters (17). Interestingly, VP16 also lacks a DNA-binding domain but contains a C-terminal acidic activation domain. These data are consistent with a function of GAS41 as a transcriptional activator.

Using the complete sequence of GAS41 presented above in in situ hybridization experiments, we were able to confirm the localization (data not shown) of GAS41 within chromosome region 12q13–15, a region which frequently is involved in gene amplifications in gliomas. Recently, we demonstrated amplification of 12 different genes within a single amplified domain at 12q13–15 in glioblastoma (19). Reifenberger and co-workers showed frequent amplification of the genes CDK4 and SAS in the 12q13–15 ampicon in human malignant glioma. However, these and additional studies failed to reveal 12q13–15 gene amplifications in low grade gliomas. Likewise, amplification of the EGFR gene has only been detected in glioblastoma and astrocytoma grade III. Here, we demonstrate the amplification of GAS41 in pilocytic astrocytoma and astrocytoma grade II.

As for the amplification level, pilocytic astrocytomas contain a relatively high number of normal brain cells including oligodendrocytes, astrocytes and neurons, complicating the detection of amplification. This histological feature of low grade gliomas may account for low levels of amplification which were only detectable by comparative PCR but not by Southern blot analysis. An enrichment of tumor cells prior to amplification analysis will allow a more adequate assessment of the amplification level in the tumor cells. Alternatively, the amplification level in single cells could be analyzed by in situ hybridization.

The amplification frequencies of GAS41 in glioblastoma and anaplastic astrocytoma are similar to those found for CDK4 and SAS. While the amplification frequency of other genes appears to increase during tumor progression, our data do not show a comparable correlation between tumor progression and GAS41 amplification frequency. GAS41 amplification frequency was detected in 23% of glioblastomas and 80% of grade I astrocytomas. Although a very limited number of low grade gliomas have been analyzed in this study, our results clearly indicate amplification of GAS41 not only in late tumor progression but also in early tumor development. This finding is in contrast to a generally held belief of amplification as a late event in tumor progression. To verify further the idea of gene amplification as an early event in tumorigenesis, additional genes need to be tested for amplification in different tumor types. Specifically, genes which are localized in the vicinity of GAS41 can be analyzed for co-amplification. As demonstrated for SAS and CDK4, both mapping at 12q13–15, co-amplification has been detected in several glioblastomas; however, no amplification of MDM2, also localized at 12q13–15, could be detected in pilocytic astrocytomas.

A final conclusion about the biological role of GAS41 in early tumor development cannot be drawn unless all amplified genes in the vicinity of GAS41 have been identified and analyzed in tumors of different grades. The directed identification and isolation of these genes will be facilitated greatly by methods such as microdissection-mediated cDNA capture. Up until now, the majority of amplified genes have been identified by screening known genes for amplification. The increasing use of methods tailored to identify amplified genes will help to analyze early amplification events and, in the long run, might alter our perspectives on the role of gene amplification in tumorigenesis.

In conclusion, GAS41, which most probably represents a transcription factor, is the first gene to be amplified in all grades of human glioma, including pilocytic astrocytoma.

MATERIALS AND METHODS

cDNA library construction and screening

cDNA was synthesized from mRNA of TX3868 glioblastoma cells using oligo(dT) primers. Second strand synthesis was performed according to the manufacturer’s instructions (Stratagene). Following EcoRI adaptor ligation and digestion with XhoI, the cDNA fragments were ligated into ZAP Express™ vector. Packaging was done using Gigapack III Gold packaging extract from Stratagene. The library was amplified once prior to cDNA library screening. Phages were plated for screening at a density of 5000 p.f.u. per 132 mm plate. Phage lift hybridization was performed using 32P-labeled GAS41 insert DNA. Positive phage clones were in vivo excised into pBK-CMV phagemids according to the manufacturer’s protocols (Stratagene), and plasmid DNA was digested with EcoRI and XhoI to release the insert DNA. In a second screening round, positive clones were identified by Southern hybridization. In brief, digested plasmid DNA was separated by gel electrophoresis and blotted onto nylon membrane. Subsequently, the membrane was hybridized with 32P-labeled GAS41 insert DNA.

Sequencing and sequence analysis

Plasmid DNA of GAS41 was isolated for sequencing using the Quiagen-Mini kit. Sequencing was performed using vector-specific primers close to the cloning site (T7 vector site: 5′ ACC CCG GTG GAA AA T CGA TGG 3′, T3 vector site: 5′ ACA AAA GCT...
**Southern blot analysis**

Isolation of high molecular weight DNA from cell cultures and frozen tissue samples was performed according to standard protocols. In brief, DNA (5 μg) from peripheral blood lymphocytes and from tumor samples was digested with EcoRI, separated on a 0.8% agarose gel and blotted onto a nylon membrane (GeneScreen). Southern hybridization was carried out in a 500 mM phosphate buffer (pH 7.2) as described previously (9,10). Insert DNA (50 ng) was labeled with 32P by the random primer method of Feinberg and Vogelstein (21).

**Comparative PCR**

Concentrations of tumor DNA and peripheral blood DNA were determined by optical density measurement. A set of dilutions (0.5, 1.0 and 2 ng/μl) was made available through the BCM launcher service.

**RT-PCR**

Prior to RT-PCR, 1.5 μg of RNA of each sample were treated with DNase I (RNase-free). After DNase I treatment, 1.5 ng of each sample were used for Alu-PCR to exclude any genomic DNA contamination. The remaining DNA-free RNA was divided into sample were used for Alu-PCR to exclude any genomic DNA contamination. The remaining DNA-free RNA was divided into

**REFERENCES**