ELSEVIER

Biochimica et Biophysica Acta 1448 (1998) 254–263

BIOCHIMICA ET BIOPHYSICA ACTA

BBA

# Clustered organization of S100 genes in human and mouse

Katrin Ridinger [a], Evelyn C. Ilg [b], Felix K. Niggli [b], Claus W. Heizmann [a], Beat W. Schäfer [a,*]

[a] *Division of Clinical Chemistry and Biochemistry, Department of Pediatrics, University of Zürich, Steinwiesstr. 75, CH-8032 Zürich, Switzerland*

[b] *Division of Oncology, Department of Pediatrics, University of Zürich, Steinwiesstr. 75, CH-8032 Zürich, Switzerland*

Accepted 13 October 1998

**Abstract**

S100 $Ca^{2+}$-binding proteins became of major interest because of their differential expression in tissues and their association with human diseases. Earlier studies showed that 13 S100 genes are located as a cluster on human chromosome 1q21. Since a number of mouse S100 genes, such as S100A4 and S100A6, have been localized to a syntenic region on mouse chromosome 3, we investigated if the S100 gene cluster exists in mouse and is structurally conserved during evolution. First we identified the cDNA sequences of mouse S100A1, S100A3 and S100A5. Then we isolated a 490 kb mouse YAC clone which gives a specific signal by FISH most likely on chromosome 3. Hybridization studies with different mouse S100 cDNAs revealed that eight mouse S100 genes are arranged in a clustered organization similar to that in human. The linkage relationships between the genes S100A8–S100A9 and S100A3–S100A4–S100A5–S100A6 were conserved during divergence of human and mouse about 70 million years ago. However, the separation of the mouse S100 genes S100A1 and S100A13 in comparison to the human linkage group suggests rearrangement processes between human and mouse. Our data demonstrate that the S100 gene cluster is structurally conserved during evolution. Further studies on the genomic organization of the S100 genes including various species could generate new insights into gene regulatory processes and phylogenetic relationships. 

*Keywords:* S100 gene; Gene cluster; Comparative mapping; Yeast artificial chromosome; Linkage relationship; Rearrangement process

## 1. Introduction

S100 proteins are a family of small acidic $Ca^{2+}$-binding proteins containing one canonical EF-hand in the C-terminal half and a S100 specific motif in the N-terminal half of the protein. For these proteins, functions in various cellular activities such as cell cycle progression, signal transduction and cell differentiation have been proposed, which are achieved by modulating the activity of other proteins, termed target proteins, in a calcium-dependent manner [1,2]. In general, S100 family members dis-

Abbreviations: PCR, polymerase chain reaction; SPE, S100 response element; SPRR, small proline-rich protein; cM, centiMorgan; CD2, lymphocyte adhesion molecule; ATP1A1, $\alpha_1$-subunit of Na,K-ATPase; NGFB, β-subunit of nerve growth factor; TSHB, β-subunit of thyrotropin; AMPD1, muscle adenylate deaminase; YAC, yeast artificial chromosome; FISH, fluorescence in situ hybridization; EST, expressed sequence tag; TBE, Tris–borate–EDTA electrophoresis buffer; SSC, sodium chloride/sodium citrate; SDS, sodium dodecyl sulfate; DIG, digoxigenin; DAPI, 4′,6-diamidino-2-phenylindole; FITC, fluorescein isothiocyanate; HOX, homeobox-containing genes; LCR, locus control region

* Corresponding author. Fax: +41-1-266-7169; E-mail: schafer@kispi.unizh.ch

0167-4889/98/$ – see front matter 
PII: S0167-4889(98)00137-2

play a cell- and tissue-specific expression pattern, have a distinct subcellular localization, and are associated with different human diseases such as cancer, neurodegenerative diseases and cardiomyopathies. Down-regulated expression of S100A2 in tumor cells, for example, suggests an involvement in tumorigenicity [3,4] whereas S100A1 shows altered expression in human cardiomyopathies [5].

Up to now, 18 different human S100 family members are known. Certain members have been identified from a variety of species such as rat, mouse, bovine, porcine, rabbit, avian and *Xenopus* [6]. Regarding the interspecies homology between individual S100 proteins, especially the S100A10 sequence available from six different species, there is a tight coincidence with the evolutionary development of species divergence. The sequence homology between rat and mouse as well as between human and bovine is higher than that between the mammalian and *Xenopus*/avian S100A10 sequence [6].

The general organization of S100 gene loci within family members as well as individual members of various species is remarkably conserved. Indeed, interspecific matches of human and rat S100A8, S100B and S100A4 exhibit almost no alterations in gene structure. With the exception of S100A4, S100A5, trichohyalin and profillagrin, the human S100 genes consist of three exons separated by two introns whereby the first exon is untranslated. The length of the three exons and the first intron is well conserved between species, as for example in human and rat S100A1. However, the length of the second intron is more diverse. In some cases, also the 5′ flanking regions of the genes are characterized by the appearance of homologous regulatory elements such as the proposed SPE (S100 response element) [6]. Nevertheless, additional nucleic acid sequences of other non-human S100 members will be required to confirm the degree of species conservation within the coding region as well as in possible regulatory elements.

Earlier studies showed that most human S100 genes are located in a gene cluster on human chromosome 1q21 [7]. The only known exceptions are calbindin D9k (Xp22), S100B (21q22) and S100P (4p16). Within the 1q21 gene cluster, subgroups of closely arranged S100 genes exist, e.g., a contiguous stretch of 15 kb genomic sequence contains four S100 genes (S100A3–S100A4–S100A5–S100A6). In addition, S100A1 neighbors S100A13 [8], and S100A8, S100A12, S100A9 could be mapped within a short distance. Interestingly, S100A10 and S100A11 are positioned 1.5 Mb away from the S100 core gene cluster, separated by epidermal differentiation genes (e.g. involucrin, SPRR3, SPRR1B, SPRR2A, loricrin) and next to the two fusion genes profilaggrin/trichohyalin containing an N-terminal S100 domain. The co-localization of the S100 gene cluster with another gene family, the epidermal differentiation genes, suggests the existence of a gene complex with functional interacting genes [9]. Because of a high gene density in this region it cannot be excluded that additional S100 genes will be identified in the future. Interestingly, this region is also a sensitive region for a number of rearrangements (deletions, translocations, duplications) associated with tumorigenicity. For example, the translocation t(X;1) (p11;q21) generating a gene fusion between TFE3 on the X-chromosome and PRCC on chromosome 1 is associated with papillary renal cell carcinomas [10]. Possibly, the altered expression of S100A2 [4], S100A4 [11] and S100A6 [12,13] in tumor cells is influenced by other rearrangements on chromosome 1q21.

Gene mapping of several mammalian species makes rapid progress, especially in the case of human and mouse sequences. The knowledge of homology relationships can be a fundamental tool in identifying disease-associated gene loci mapped just in the other species. Up to now, several hundred different conserved linkage groups in human and mouse have been described [14]. Earlier studies have demonstrated the existence of a large linkage group conserved between human chromosome 1q21–1p22 and the distal mouse chromosome 3 [15–17] Indeed, S100A4, S100A6 and S100A10 belong to this syntenic region and have been mapped to mouse chromosome 3 [14,16,18,19].

Since a conserved gene order on mouse chromosome 3 has been found for genes such as CD2 (murine Ly-37), ATP1A1, NGFB, TSHB and AMPD1 [17], we wished to investigate if the S100 gene cluster might also be evolutionary conserved. In addition, a comparison of the genomic organization of the S100 genes in evolution might elucidate hot spots of possible rearrangement processes. Furthermore, the

availability of mouse S100 genes will allow to address future functional questions through genetic manipulations in the animal. Here, we present the isolation of a 490 kb YAC clone using PCR screening with primers derived from the mouse S100A3 gene. On this YAC, which was non-chimeric as demonstrated by FISH, we were able to localize eight mouse S100 genes. Our study suggests that the S100 gene cluster is structurally conserved during evolution, whereas some differences might be explained by rearrangements such as inversion processes.

## 2. Materials and methods

### 2.1. Screening of a YAC mouse clone

A YAC library from mouse genomic DNA (mouse Super Pool YAC 8540, Genome Systems) was screened by PCR using primers specific for S100A3 (MA3-1: 5′-GCAGGCAGTAGCTGCCATC-3′ and MA3-2: 5′-TTGAAGTACTCGTGGCAGTAG-3′) under the following conditions: 95°C 1 min, 58°C 1 min, 72°C 2 min, 30 cycles, followed by 72°C 10 min extension.

### 2.2. DNA probes

The *Sal*I–*Bam*HI and the *Cla*I–*Bam*HI fragments from pBR322 which flank the cloning site of pYAC4 were used after gel purification (QIAquick Gel Extraction Kit, Quiagen) to establish the restriction map of the YAC 13088. EST clones containing the different S100 cDNAs (Table 1) were supplied by the HGMP Resource Center, Cambridge, UK; cDNA inserts, released after digestion with *Eco*RI and *Not*I restriction enzymes, followed by gel purification, were used for the hybridization studies. A S100A8 probe [20] and a S100A4 probe [21] were generous gifts from Dr. Robert Passey and Dr. Gajanan Sherbet. A probe specific for S100A3 was generated by PCR using primers from mouse genomic DNA (Promega) as described above.

Table 1
Mouse S100 family members

| | EST-clone (accession no.) | cDNA (accession no.) |
|---|---|---|
| S100A1 | AA109939 | AF 087687 |
| S100A3 | AA015155 | AF 087470 |
| S100A5 | W09299 | AF 087469 |
| S100A6 | W18305 | X66449 |
| S100A9 | W18902 | M83219 |
| S100A10 | W34395 | M16465, J02779 |
| S100A11 | W33683 | U41341 |
| S100A13 | AA033478 | X99921 |

### 2.3. DNA preparation and pulse field electrophoresis

Yeast containing the YAC 13088 was cultured in AHC Medium (Current Protocols in Molecular Biology, Suppl. 20, 6.10.16) for 2 days at 30°C, 250 rpm and DNA prepared using the Chef Genomic DNA Plug Kit (BioRad). Thirty to 60 μl agarose blocks containing 2–4 μg YAC DNA were digested with the appropriate restriction enzyme for 2–10 h. After equilibration in 0.5% TBE they were resolved with the Chef-DR III System (Biorad) using a 1% agarose gel (pulse field certified agarose, Biorad).

### 2.4. Southern blot analysis

After blotting to Nytran Nylon membranes (Schleicher and Schuell) the filters were hybridized (QuickHyb Hybridization Solution, Stratagene) with [$^{32}$P]dATP-labeled (Prime-a-gene, Promega) probes and washed with 2×SSC, 0.1% SDS at room temperature for 2×15 min, 0.1×SSC, 0.1×SDS for 30 min. The S100A10 probe was also labeled using the DIG-system (DIG High Prime; DIG Luminescent Detection Kit, Boehringer).

### 2.5. Fluorescence in situ hybridization

Chromosome spreads of C2C12 mouse myoblasts were prepared according to standard cytogenetic methods. Twenty-five μg of total yeast DNA containing approximately 1 μg of the mouse YAC 13088 was labeled by nick-translation with biotin-14-dATP (Life Technologies). After co-precipitation with 20 μg of mouse Cot-1 DNA (Life Technologies) and 25 μg of salmon sperm DNA (Sigma), and denaturing for 10 min at 72°C, repetitive sequences present in the YAC were pre-annealed for 2 h at 37°C in 40 μl hybridization buffer (50% deionized formamide, 10% dextran sulfate, 2×SSC,

**A**

```
human 1  MGSELETAMETLINVFHAHSGKEGDKYKLSKKELKELLQTELSGFLDAQKDVDAVDKVMKELD 63
Mouse          S                              D           V   A

human 64 ENGDGEVDFQEYVVLVAALTVACNNFFWENS 94
mouse             K                   T
```

**B**

```
human 1  MARPLEQAVAAIVCTFQEYAGRCGDKYKLCQAELKELLQKELATWTPTEFRECDYNKFMSVLD 63
Mouse     T                          I  S          P    S

human 64 TNKDCEVDFVEYVRSLACLCLYCHEYFKDCPSEPPCSQ 101
mouse             G       S          E  P    P
```

**C**

```
human 1  MPAAWILWAHSHSELHTVMETPLEKALTTMVTTFHKYSGREGSKLTLSRKELKELIKKELCLG 63
Mouse    ------------------                                        T  S A

human 64 E-MKESSIDDLMKSLDKNSDQEIDFKEYSVFLTMLCMAYNDFFLEDNK 111
mouse     K       N                        T
```

Fig. 1. Comparison between the amino acid sequences of human and predicted mouse S100A1 (A), S100A3 (B) and S100A5 (C). The upper sequence shows the complete human amino acid sequence; in the lower sequence the exchanges of amino acids in the mouse sequences are given. The underlined amino acids near the C-terminus and next to the N-terminus build the N-terminal and C-terminal EF-hand. (C) The prolonged N-terminus of S100A5 was not confirmed in the mouse EST clone Wo9299 and is depicted by dashes.

0.5 mM phosphate buffer, pH 7.0). Hybridization and visualization of the YAC was performed as described by Nacheva [22]. The slides were mounted in anti-fade medium containing 125 ng/ml DAPI (Vysis) for counterstaining of the chromosomes. Fluorescence signals were visualized on a Olympus BX60 microscope equipped with appropriate filters. Fluorescence images were captured by Perceptive Scientific International (Chester, UK) imaging software.

## 3. Results

### 3.1. Identification of mouse S100 family members

The S100 family has been predominantly defined on the basis of amino acid and nucleotide sequence homology. The divergency between different S100 members of one species ranges on the cDNA level from 61% for S100A2 and S100A4 to 37% for S100A8 and S100A10 [6]. In contrast, the interspe-

cies homology between the presently known mouse and human individual S100 cDNAs is in the range of 85–95%. To identify as many mouse S100 family members as possible we searched the accessible nucleotide databases. For mouse S100A4, S100A6, S100A8, S100A9, S100A10, S100A11 and S100A13 the respective cDNA sequences could immediately be retrieved and the corresponding EST clones identified (Table 1).

In addition, we identified different ESTs coding for mouse cDNAs with high homology to human S100A1, S100A3 and S100A5. In order to confirm that these EST clones represent indeed the mouse S100A1, S100A3 and S100A5 cDNAs, we compared the interspecies identity of the coding regions in percentage (Table 2). Human S100A1 and the EST clone (AA109939) displayed an identity which is similar to the homologies found for bovine (94%) and rat (88%) S100A1. In conclusion, the homologies are consistent with the general order of species divergence. Comparable to S100A1, the mouse EST clones AA015155 and W09299 are both 90% homologous to human S100A3 and S100A5 counterparts, respectively. The interspecies identity on the amino acid level reaches up to 93% for S100A1, 90% for S100A3, and 93% for S100A5.

All EST clones were subsequently resequenced and some minor sequencing errors corrected allowing for comparison with the human protein sequences. Comparing mouse and human S100A1 amino acid sequences (Fig. 1A), only a few exchanges are observed which are mostly found in the central region that is thought to be responsible for the interaction with target proteins [23,24]. However, most amino acids are only replaced with other conserved residues like E→D, A→V, V→A (Fig. 1A), suggesting that the specificity of interaction is probably retained. In the S100A3 amino acid sequence exchanges from human to mouse S100A3 are found at the N-terminus (A→T), in the N-terminal EF-hand (A→S), in the central hinge region (A→P) and in the C-terminal EF-hand (V→G), which might result in a less hydrophobic character of the mouse protein (Fig. 1B). Comparing the sequence divergency of mouse and human S100A5, exchanges are particularly striking in the hinge region (Fig. 1C) with nonconserved substitutions (K→T), (C→S), (G→A), which might change the functional properties of the mouse protein. In summary, because of high homology and structural conservation to human S100A1, S100A3 and S100A5 we conclude that we have determined the corresponding mouse cDNA sequences.

Table 2
Interspecies identity of S100A1, S100A3, and S100A5

| Human | Mouse | |
|---|---|---|
| | mRNA (coding region) | Protein |
| S100A1 | 87% | 93% |
| S100A3 | 90% | 90% |
| S100A5 | 90% | 93% |

Likewise, we searched for mouse S100A2, S100A7 and S100A12 in the EST databases, but failed to identify these cDNAs. One explanation could be that the interspecies divergence is much higher. However, given the homologies observed with other S100 proteins, this seems to be rather unlikely. Alternatively, these genes might not exist in mouse or their expression level in cells or tissues is low. However, so far the mouse projects are still ongoing and it seems likely that the remaining S100 genes will be identified through this approach in the future.

### 3.2. *Isolation of a YAC clone containing the genomic locus of mouse S100A3*

An interesting question is whether the clustered organization of the S100 genes on human chromosome 1q21 [7] is evolutionary conserved in other organisms. To investigate this, we screened a mouse YAC genomic DNA library with primers derived from the mouse S100A3. S100A3 was chosen because it is a central gene in the human S100 cluster which might thereby enhance the probability to identify a YAC with a high density of other S100 mouse genes. The YAC clone 13088 was established to contain S100A3 by subsequent hybridization with the S100A3 PCR-product. Its size was estimated to be 490 kb (Fig. 2A). A common problem in working with YAC clones could be their chimeric composition from different chromosomal origins generated during library construction. To exclude this for YAC 13088, we applied FISH on normal mouse chromosomes. As shown in Fig. 3, hybridization was detected on only one chromosome pair, most likely on chromosome 3. Hence,

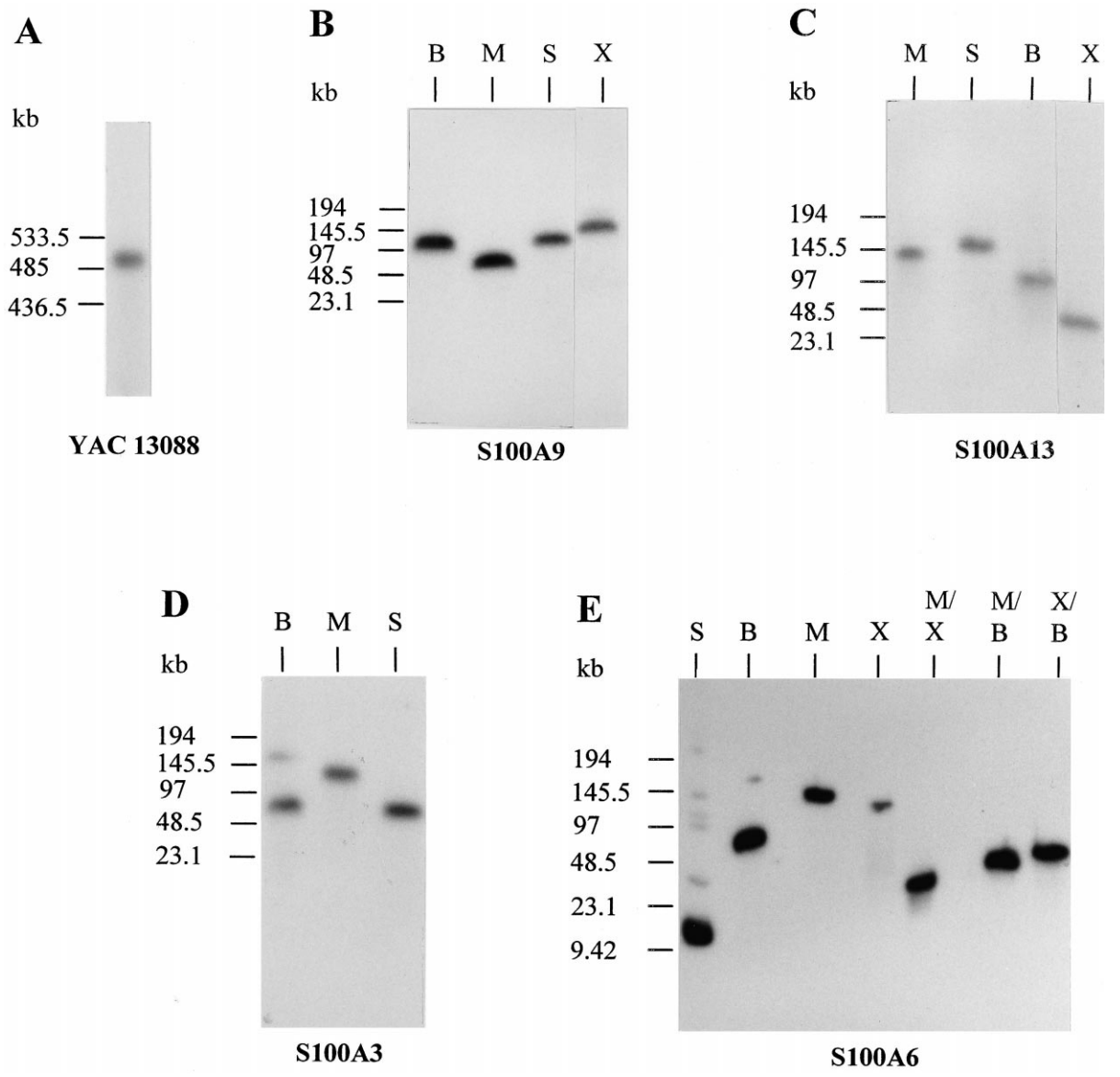


Fig. 2. Hybridization of gene-specific DNA probes to yeast DNA containing YAC 13088 separated by pulse field gel electrophoresis. In the audioradiogram the different lanes show the hybridization of gene-specific DNA probes to the YAC 13088 (A) or to the indicated restriction enzyme digests (B–E). In B–E, rare cutting enzymes such as *Mlu*I = M, *Bss*HII = B, *Sal*I = S, *Xho*I = X are used. Additionally in E, double digestions are shown. The gene-specific DNA probes were labeled with $^{32}P$, followed by hybridizations. The sizes were calculated from the co-electrophoresed molecular mass marker.

YAC 13088 was indeed derived from one chromosomal region.

### 3.3. Mouse S100 genes are organized in a cluster

To map the mouse S100 genes we first established a restriction map of YAC 13088 by hybridizing single, partially and completed restriction digests of the yeast clone with pBR332 fragments flanking the left and right YAC arms (Fig. 4A). Next, we localized the different mouse S100 genes by hybridizing the different cDNAs and genomic probes with informative restriction digests (*Bss*HII, *Mlu*I, *Sal*I and *Xho*I). Examples of hybridizations are shown in Fig. 2B–E. A distinct pattern of hybridization was obtained for S100A9, S100A13, S100A3 and S100A6. Using the cDNA probes for S100A3, S100A4 and S100A5, hybridization with the same 70 kb *Bss*HII-, 125 kb

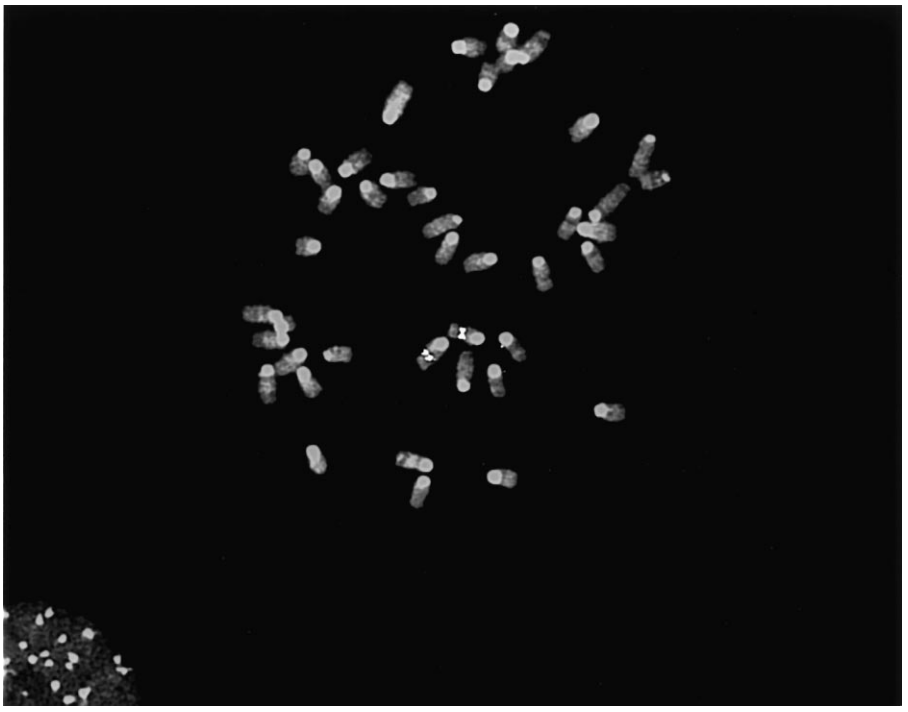

Fig. 3. FISH (fluorescence in situ hybridization) of mouse YAC 13088 on chromosomes derived from C2C12 mouse myoblasts. Chromosomes from colcemid-treated cells were prepared and hybridized with biotin-14-dATP-labeled mouse 13088 YAC DNA. Specific positive signals were detected in one chromosome pair by using FITC-conjugated avidin in combination with biotinylated goat anti-avidin. Counterstaining of the chromosomes was performed by DAPI.

*Mlu*I- and 60 kb *Sal*I-fragments was observed (Fig. 2D). Thus S100A3, S100A4 and S100A5 are mapped within a minimal distance of 25 kb. S100A6 yielded the same 70 kb *Bss*HII- and 125 kb *Mlu*I-fragment as S100A3; however, the *Sal*I-fragment was only 10 kb (Fig. 2E). Because of this, S100A6 could be orientated close to S100A3–A5 versus the left arm YAC. In order to confirm this localization, different double digestions with *Mlu*I/*Xho*I, *Mlu*I/*Bss*HII and *Bss*HII/*Xho*I were performed. In conclusion, mouse S100A3–S100A4–S100A5–S100A6 are localized in a small stretch of maximal 35 kb. Further analyses to resolve the order of S100A3–S100A4–S00A5 have so far been unsuccessful.

Furthermore, we found that S100A8 and S100A9 (Fig. 2B) hybridized to the same 100 kb *Bss*HII-, 70 kb *Mlu*I-, 120 kb *Sal*I- and 155 kb *Xho*I-fragment. This allowed to place S100A8 and S100A9 as neighbors in a distance of maximal 45 kb. Similarly, the localization of the S100A1 gene could be determined on the first 100 kb *Bss*HII-, 70 kb *Mlu*I-, and 120 kb *Sal*I-fragment versus the left arm of the YAC (data not shown). The hybridization of S100A13 to a 130 kb *Mlu*I-, 145 kb *Sal*I-, 100 kb *Bss*HII- and 25 kb *Xho*I-fragment (Fig. 2C) and further double digestions (data not shown) allowed us to assign S100A13 at least 170 kb away from S100A1, S100A8 and S100A9 and about 110 kb apart from S100A6. In our studies, we failed to find a positive hybridization signal for S100A11 and Calb3 probes. The S100A10 probe gave contradictory results in that a DIG-labeled probe was weakly hybridizing whereas a radioactively labeled S100A10 probe repeatedly failed to give a positive signal on the YAC.

Hence, our results demonstrate that S100A1, S100A8, S100A9, S100A13, S100A6, S100A3, S100A4 and S100A5 are clustered on a mouse chromosome within a distance of 410 kb. Since S100A4 and S100A6 have been previously assigned to the mouse chromosome 3 [14,16,19], we conclude that the S100 cluster identified in this study is located on mouse chromosome 3. This is consistent with the result obtained by FISH analysis (Fig. 3).

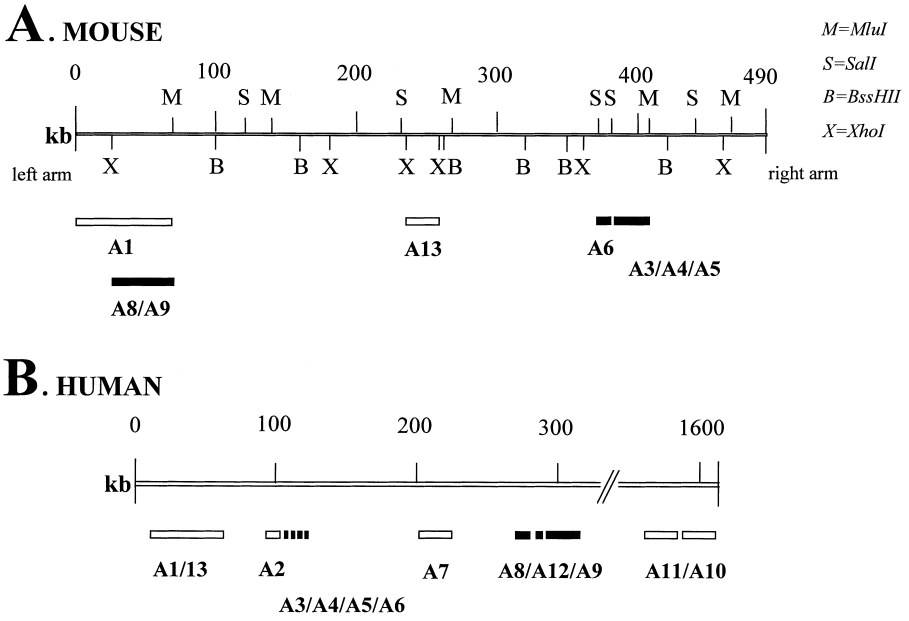


Fig. 4. Comparison between the organization of the mouse and human S100 gene cluster. (A) A 490 kb genomic mouse sequence of the YAC 13088 shows the location of the S100 genes depicted by boxes. A restriction map was established by hybridizing single partially and completed restricted genomic yeast DNA with pBR322 fragments flanking the left and right arms. S100A3–S100A4–S100A5–S100A6 and S100A8–S100A9 are defined by filled boxes, S100A1 and S100A13 are defined by open boxes. Considering the assignment of S100A4 and S100A6 [14,16,19] to the mouse chromosome 3, the cluster is located on this mouse chromosome. (B) A region of 1600 kb of the human chromosome 1q21 is drawn schematically with the genes and their location depicted by boxes [7].

## 4. Discussion

In this article we describe the isolation of a mouse YAC clone containing at least eight mouse S100 genes. The identification of this mouse S100 cluster now allows to perform some evolutionary comparisons with the known human cluster. Interestingly, the correlation reveals that the linkage relationship is structurally conserved, but also shows some divergency (Fig. 4). Similar to the 15 kb stretch of human genomic sequence containing four genes (S100A3–S100A4–S100A5–S100A6), the corresponding mouse genes are arranged in a small sequence of 35 kb at maximum. So far, we were not able to determine the complete gene order in mouse, but it seems likely to be the same as in human. Likewise, S100A8 and S100A9 are neighbors in both mouse and human. Up to now, no cDNAs for mouse S100A2 and S100A12 could be identified. Because of the close linkage of S100A2 to S100A3–S100A4–S100A5–S100A6 and S100A12 to S100A8–S100A9 in human, a similar arrangement in mouse might be postulated. In contrast, we could observe the separation of the mouse S100A1 and S100A13 in comparison to the homologous human genes. Concerning mouse S100A10, we obtained contradictory results. As mentioned above, only a DIG-labeled S100A10 probe but not a radioactive labeled probe gave a hybridization signal. Since the epidermal differentiation gene SPRR1A was not localized on this YAC clone (data not shown) and since S100A10 in human maps at a distance of 1.5 Mb from the main S100 cluster, it seems reasonable to assume that mouse S100A10 is not localized on the YAC. This is further supported by the notion that we also failed to detect a positive signal for S100A11 which neighbors S100A10 in human.

Taking into account the previous assignment of S100A4 and S100A6 to mouse chromosome 3 [14,16,19] and our FISH results, the cluster of eight mouse genes is located on chromosome 3. In contrast to genes such as CD2, ATP1A1, NGFB, TSHB, and AMPD1 which show an identical gene order between human and mouse in the linkage segment of human

chr1/mouse chr3, some rearrangements occurred within the S100 gene cluster. These rearrangements could be explained by an inversion event requiring breakpoints between S100A1–S100A13 and S100A7–S100A8 that would result in juxtapositioning of S100A1 to S100A8–S100A9. In this model, during the separation of S100A1 and S100A13, S100A13 might have been inserted between the subgroups S100A8–S100A9 and S100A6–S100A3–S100A4–S100A5. It would be of great interest to see if these breakpoint regions implicated by evolution might also be relevant in rearrangement processes explaining the involvement of chromosome 1q21 genes in tumorigenicity.

Presumably, the S100 gene clusters in human and mouse evolved from mechanisms such as gene duplication and diversification processes. An expansion of S100 genes within a primordial gene cluster prior to the divergence of mouse and human about 70 million years ago is likely. After this initial expansion, at least two rearrangement processes including an altered gene order and a diversity in the gene sequence occurred. A common origin of mouse and human clusters could also be shown, e.g., in the case of the Antennapedia-class HOX genes [25,26] and the globin gene clusters [27]. Hence, it would be of great interest to study the S100 gene organization in more unrelated species, e.g., in *Xenopus* and avian. Moreover, the search for S100 members in non-vertebrate animals could elucidate the evolutionary origin of S100 genes. It will also be intriguing to see if the relationship in sequence similarity and genomic organization of S100 genes in mammalians and more unrelated species are under strong conservational constraints like, e.g., in the HOX gene clusters [25]. It has been proposed that S100 proteins interact with target proteins via their specific 'hinge' region [23,24] which was recently supported by the three-dimensional structure of S100B [28–30]. Since other specific parts of S100 proteins such as helices III and helices IV are also important in target protein recognition [28], the availability of S100 proteins from more primitive species and their structures would allow a comparison between characteristic binding surfaces. Perhaps the three-dimensional structures of S100 homologues from different species might reveal if alterations in the functions of proteins took place during evolution.

Less is known about the transcriptional regulation of S100 genes. Up to now, a few regulatory factors of S100 genes have been identified, e.g., a single positive regulatory enhancer has been described in the S100A2 gene [31]. Furthermore, the S100A6 promoter is known to be regulated by serum-inducible sequences [32] and through an AP-1 like region during cell differentiation [33]. In the case of S100A4 an enhancer element was identified in the first intron [34]. The clustered organization of S100 genes in mouse and human raises the question if there are DNA segments acting as locus control regions (LCR) or other superior control elements that are required for the activation of cluster subgroups in the S100 cluster. Given that locus control regions are located at the far 5′ end of the mammalian β-globin gene cluster [35], comparisons between genomic sequences within the human and mouse S100 gene cluster might be useful to identify candidate regions containing regulatory elements. However, because of the rearrangements between the human and the mouse gene cluster, we propose that *cis*-acting regulatory elements might not be common for the entire gene cluster but more likely for the conserved subgroups such as S100A3–S100A4–S100A5–S100A6. In this regard it would be of great interest to see if the S100 gene cluster in mouse is co-localized with the epidermal differentiation genes as in human. The physical linkage of these gene families suggests a functional dependence during keratinocyte differentiation [9]. Hence, locus control regions might be present which regulate the optimal gene expression within the context of the entire gene complex such as in the major histocompatibility (HLA) complex [36,37]. Our study provides a first step in a phylogenetic approach to elucidate the genomic organization of S100 genes in a variety of species, yielding important insights into regulation and function of S100 genes and proteins.

## Acknowledgements

This study was supported by BIOMED2 (European Union, Grant no. BMH4CT950319/BBW, Grant no. 95.0215-1, Switzerland). We thank Dr. David Betts for valuable suggestions and discussion on FISH analysis, Dr. Gajanan Sherbet for the

18A2/mts (murine S100A4) gene probe, and Dr. Robert Passey for the specific mouse probe to CP-10 (S100A8).

## References

[1] B.W. Schäfer, C.W. Heizmann, Trends Biochem. Sci. 21 (1996) 134–140.
[2] C.W. Heizmann, J.A. Cox, Biometals (1998) in press.
[3] E.C. Ilg, B.W. Schäfer, C.W. Heizmann, Int. J. Cancer 68 (1996) 325–332.
[4] S.W. Lee, C. Tzomasetto, K. Swisshelm, K. Keyomarsi, R. Sager, Proc. Natl. Acad. Sci. U.S.A. 89 (1992) 2504–2508.
[5] A. Remppis, P. Greten, B.W. Schäfer, P. Hunziko, P. Erne, H.A. Katus, C.W. Heizmann, Biochim. Biophys. Acta 1313 (1996) 253–257.
[6] D.B. Zimmer, J. Chessher, W. Song, Biochim. Biophys. Acta 1313 (1996) 229–238.
[7] B.W. Schäfer, R. Wicki, D. Engelkamp, M.G. Mattei, C.W. Heizmann, Genomics 25 (1995) 638–643.
[8] R. Wicki, I. Marenholz, D. Mischke, B.W. Schäfer, C.W. Heizmann, Biochim. Biophys. Res. Commun. 227 (1996) 594–599.
[9] D. Mischke, B.P. Korge, I. Marenholz, A. Volz, A. Ziegler, J. Invest. Dermatol. 106 (1996) 989–992.
[10] M.A.J. Weterman, M. Wilbrink, A. Geurts van Kessel, Proc. Natl. Acad. Sci. U.S.A. 93 (1996) 15294–15298.
[11] B.R. Davies, M.P.A. Davies, F.E.M. Gibbs, R. Barraclough, Oncogene 8 (1993) 999–1008.
[12] M.A.J. Weterman, G.M. Stoppen, G.N.P. Van Muijen, J. Kuznicki, D.J. Ruiter, Cancer Res. 52 (1992) 1291–1296.
[13] M. Pedrocchi, B.W. Schäfer, H. Mueller, U. Eppenberger, C.W. Heizmann, Int. J. Cancer 57 (1994) 684–690.
[14] R.W. DeBry, M.F. Seldin, Genomics 33 (1996) 337–351.
[15] A.G. Searle, J. Peters, M.F. Lyon, E.P. Evans, J.H. Edwards, V.J. Buckle, Genomics 1 (1987) 3–18.
[16] W.S. Moseley, M.F. Seldin, Genomics 5 (1989) 899–905.
[17] S.F. Kingsmore, W.S. Moseley, M.L. Watson, R.L. Sabina, E.W. Holmes, M.F. Seldin, Genomics 7 (1990) 75–83.
[18] C. Saris, T. Kristensen, P. D'Eustachio, L. Hicks, D. Noonan, T. Hunter, B. Tack, J. Biol. Chem. 262 (1987) 10663–10671.
[19] J.R. Dorin, E. Emslie, V. Van Heyningen, Genomics 8 (1990) 420–426.
[20] C. Geczy, Biochim. Biophys. Acta 1313 (1996) 246–252.
[21] A. Ebralidze, E. Tulchinsky, M. Grigorian, A. Afanasyeva, V. Senin, E. Revazova, E. Lukanidin, Genes Dev. 3 (1989) 1086–1093.
[22] E. Nacheva, L. Kearney, M. Bower, T. Chaplin, E. Douek, E. Das, B.D. Young, Genes Chromosomes Cancer 7 (1993) 123–127.
[23] D. Kligman, D.C. Hilt, Trends Biochem. Sci. 13 (1988) 437–443.
[24] M. Lackmann, P. Rajasekariah, S.E. Iismaa, G. Jones, C.J. Cornish, S. Hu, R.J. Simson, R.L. Moritz, C.L. Geczy, J. Immunol. 150 (1993) 2981–2991.
[25] C. Kappen, K. Schughart, F.H. Ruddle, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 5459–5463.
[26] J. Zhang, M. Nei, Genetics 142 (1996) 295–303.
[27] A. Jeffreys, in: G.A. Dover, R.B. Flavell (Eds.), Genome Evolution, Academic Press, London, 1982, pp. 157–176.
[28] A.C. Drohat, D.M. Baldisseri, R.R. Rustandi, D.J. Weber, Biochemistry 37 (1998) 2729–2740.
[29] S.P. Smith, G.S. Shaw, Structure 6 (1998) 211–222.
[30] H. Matsumura, T. Shiba, T. Inoue, S. Harada, Y. Kai, Structure 6 (1998) 233–241.
[31] R. Wicki, C. Franz, F.A. Scholl, C.W. Heizmann, B.W. Schäfer, Cell Calcium 22 (1997) 243–253.
[32] F. Ghezzo, E. Lauret, S. Ferarri, R. Baserga, J. Biol. Chem. 263 (1988) 4758–4763.
[33] F. Bottini, K. Mazzocco, T. Abbondi, G.P. Tonini, Neurosci. Lett. 181 (1994) 35–38.
[34] E. Tulchinsky, D. Kramerov, H.L. Ford, E. Reshetnyak, E. Lukanidin, S. Zain, Oncogene 8 (1993) 79–86.
[35] R. Hardison, J.L. Slightom, D.L. Gumucio, M. Goodman, N. Stojanovic, W. Miller, Gene 205 (1997) 73–94.
[36] S. Carson, M.V. Wiles, Nucleic. Acids Res. 21 (1993) 2065–2072.
[37] N. Dillon, F. Grosfeld, Trends Genet. 9 (1993) 134–137.