Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 6547–6551, July 1993
Biochemistry

Six *S100* genes are clustered on human chromosome 1q21: Identification of two genes coding for the two previously unreported calcium-binding proteins S100D and S100E

(gene cluster/chromosome 1/cancer)

DIETER ENGELKAMP*, BEAT W. SCHÄFER*, M. GENEVIÈVE MATTEI†, PAUL ERNE‡, AND CLAUS W. HEIZMANN*§

*Abteilung für Klinische Chemie, Kinderspital, Universität Zürich, Steinwiesstrasse 75, CH-8032 Zürich, Switzerland; †Hôpital d'enfants, Groupe hospitalier de la Timone, F-13385 Marseille, France; and ‡Abteilung für Kardiologie, Kantonsspital, CH-6000 Luzern, Switzerland

Communicated by Robert H. Wasserman, March 5, 1993

ABSTRACT **The human genome contains large regions that are highly structured. Sequence-related members of multigene families are often found in a clustered organization. Here we describe a previously unrecognized gene cluster composed of genes coding for calcium-binding proteins of the S100 family. The linkage of six *S100* genes was established by pulsed-field gel electrophoresis, and a contiguous DNA sequence of 15 kilobases containing the full coding region of four different *S100* genes was characterized. This is the tightest mammalian gene cluster discovered so far to our knowledge. Two additional *S100* genes are located within the cluster, both of which exhibit unique structural features when compared with other *S100* genes. The product of *S100E* is cysteine-rich, whereas that of *S100D* contains a long hydrophobic N-terminal tail. The gene cluster was assigned to chromosome 1q21, one of the bands showing rearrangements in neoplasms at high frequency. The deregulated expression of some *S100* genes in the cluster during tumor progression suggests that chromosomal abnormalities may influence the expression of *S100* genes in late stages of cancer, particularly in association with the formation of metastases.**

The human genome is highly organized. Global variations in G+C-content (1) and gene distribution (2) are found. Several genes are arranged in a closely linked gene order. Two classes of these gene clusters can be distinguished: a first group contains genes that are unrelated in sequence, like the mouse surfeit locus (3) and the *Drosophila* shaker complex (4); and a second group contains members of multigene families, like the α-globin and β-globin gene clusters (5), the histone gene clusters (6), and the homeobox gene clusters (7).

What is the genetic advantage that keeps these genes in a clustered organization? One motive might be the overall regulation of transcripts within gene clusters—for example, by locus control regions (8), by the maintenance of the correct gene order (9), or by the location within chromosomal domains of particular "chromatin flavor" (10). Information about clustered multigene families is rare and restricted to genes that are expressed only in a certain cell type (globin genes, α-amylase genes) and to genes that show a strictly developmental expression pattern (homeobox genes, globin genes).

S100 proteins¶ are small EF-hand calcium-binding proteins (11) of variable sequence identities (38 to 63%). Each member of this gene family displays a quite different expression pattern in human tissues. Some S100 proteins are associated with tumor development and the metastatic behavior of tumors. S100L is preferentially expressed in normal mammary epithelial cells and not in breast tumor cells (12). Calcyclin (CACY), S100α, and S100β, on the other hand, show enhanced expression with tumor progression (11). The expression of CAPL (calcium protein, murine placental homologue) increases with increasing degrees of metastasis in mouse mammary carcinoma and leukemic cells (13).

Indications that *S100* genes might form a gene cluster came from their colocalization on human chromosome 1 (12, 14–16). The *S100* gene cluster described in this work contains at least six different *S100* genes, including two previously unreported members of this family.‖ The compact organization of the genes and their diverse expression patterns suggest regulation mechanisms different from other gene clusters. The assignment of this cluster to human chromosome 1q21, a chromosome band often involved in tumor-associated rearrangements, together with their deregulated expression patterns in cancer suggests a participation of some *S100* genes in tumor progression.

MATERIALS AND METHODS

Polymerase Chain Reaction (PCR). Total RNA of different human tissues was isolated by the guanidinium isothiocyanate method (17). First-strand cDNA was synthesized by using Superscript (BRL) reverse transcriptase. Boehringer *Taq* DNA polymerase was used for PCR.

Oligonucleotides. The following primers were used in which B = C, G, or T; N = A, C, G, or T; M = A or C; Y = C or T; R = A or G; S = C or G; K = G or T; V = A, C, or G: Cons-I, TBNMYRYSTTCCACTAGTAYTCNGKSMRVGA (contains recognition sequence for *Spe* I); Cons-IV, GGAGAGCTCSBTBBSAKNAGYTCYTTSAG (contains recognition sequence for *Sst* I); RACE-I-O-$(T)_{17}$, AAGTCGACATCGATGGATCCTGCAGAAGC$(T)_{17}$; RACE-O, AAGTCGACATCGATGGATCC; RACE-I, GATGGATCCTGCAGAAGCTT; RACE-D31, GAGGGTAGCAAACTGACCCT; RACE-D32, GACCCTGAGTAGGAAGGAAC; RACE-D51, GATCTCCTGGTCGCTGTTCT; RACE-D52, TGTCCAGGCTCTTCATCAAG; RACE-E51, CGGGGTCCAGGTGGCCAGCTC; RACE-E52, CTTCTGCAGCAGCTCCTTGAG; S100D-A, GAGTAAAGTTTCTGGGATTGGG; S100D-B, AAAGAGGGTC-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: PFGE, pulsed-field gel electrophoresis; RACE, rapid amplification of cDNA ends; CACY, calcyclin; CAPL, calcium protein, murine placental homologue.

§To whom reprint requests should be addressed.

¶To avoid confusion about the different names given to S100 proteins, we will use the nomenclature of the committee of the Tenth International Workshop on Human Gene Mapping (1989). All other names are listed in Table 1.

‖The sequences reported in this paper have been deposited in the GenBank data base [accession nos. Z18954 (S100D cDNA), Z18948 (S100E cDNA), and Z18949 and Z18950 (*S100* gene cluster)].

TGTGAGAGGAGC; S100E-A, GTCTCAGATTGGTAA-ACACCCC; S100E-B, GGCAAGTCCAGATTGAA-AGGGG.

PCR with degenerated primers. Human genomic DNA was isolated from blood lymphocytes (17). PCR was performed in 100-μl reaction mixtures containing 2 nmol each of the primers Cons-I and Cons-IV, 1 μg of genomic DNA as template, and 2.5 units of *Taq* DNA polymerase. Cycle steps were as described in Engelkamp *et al.* (18).

3'-Rapid amplification of cDNA (RACE). Amplification of the S100D cDNA 3' end was performed as described by Frohman (19) with the following modifications: total RNA (3 μg) was reverse-transcribed by using 0.1 nmol of RACE-I-O-$(T)_{17}$ as primer; cycle steps were 1 min at 95°C, 1 min at 65°C, and 1.5 min at 72°C (30 cycles). The primer pairs used for the first and second round of amplification were: RACE-D31/RACE-O and RACE-D32/RACE-I.

5'-RACE. Amplifications of the S100D and S100E cDNA 5'-ends were performed by the method of Frohman (19). Total RNA (30 μg) was reverse-transcribed in a total volume of 50 μl by using 1.0 μg of random hexamers as primers. After poly(A)-tailing (terminal deoxyribonucleotidyltransferase, BRL), 1/25th of this first-strand cDNA mixture was used as template for amplification. The primer sets used for the first and second round of amplification were: Oligo-I-O-$(T)_{17}$/RACE-O/RACE-D51 (S100D) or Oligo-I-O-$(T)_{17}$/RACE-O/RACE-E51 (S100E) and RACE-I/RACE-D52 (S100D) or RACE-I/RACE-E52 (S100E). The cycle steps were as follows: 1 min at 95°C, 1 min at 65°C, 1.5 min at 72°C (30 cycles). The first amplification round was preceeded by one cycle at 37°C annealing temperature (1 min at 95°C, 10 min at 37°C, and 60 min at 55°C).

cDNA amplification. Total RNA (3 μg) was reverse-transcribed in a total volume of 20 μl with 0.1 nmol of oligo$(dT)_{15}$ as primer. cDNAs of S100E and S100D were amplified at standard conditions (17). Primer sets were: S100E-A/S100E-B and S100D-A/S100D-B.

Isolation of Genomic DNA Containing Phages. Two human genomic libraries (Clontech, nos. HL 1067J and HL 1006D) were screened with [α-^{32}P]dCTP random-primed cDNA probes of CACY, CAPL, and S100D as described by Sambrook *et al.* (17). Sequencing of DNA subcloned into Bluescript SK(+) (Stratagene) followed the Sanger protocol (20), and sequence analysis was performed with the GCG sequence analysis software package of the University of Wisconsin (21).

Pulsed-Field Gel Electrophoresis (PFGE). Human genomic DNA used for mapping was isolated from the human epithelial cell line HEp-2. Logarithmic-phase culture cells were embedded in low-gelling-temperature agarose, processed, and digested as described by Sambrook *et al.* (17). Restriction enzyme-digested DNA was resolved at 14°C and 200 V by contour-clamped homogenous electrophoresis (CHEF system, Bio-Rad) with 60-sec pulses for 15 hr, followed by 90-sec pulses for an additional 9 hr. After electrophoresis, the DNA was transferred to Hybond-N membrane (Amersham) by diffusion blotting and was hybridized (17). Exposure times were 4 and 6 days. Complete removal of the previous probe was checked by exposure for at least 2 days before rehybridization with the next probe.

***In situ* Hybridization.** *In situ* hybridization on chromosome preparations was carried out as described (22).

RESULTS

S100D and S100E Are Two Previously Unreported S100 Proteins. Initially we attempted to search for new S100 proteins using a degenerated PCR approach. Oligonucleotides were constructed against conservative regions in the first calcium loop of S100 proteins. This sequence region contains the noncanonical EF-hand (calcium loop of 14 amino acids), which is specific to S100 proteins and significantly different from the canonical EF-hand (calcium loop of 12 amino acids) (23). In all known *S100* genes, the first calcium loop is encoded by a single exon. A low-stringency PCR was performed on human genomic DNA under the same conditions as those used to screen for S100 proteins in human heart RNA (18). After subcloning and sequencing, we identified 23 clones as known human *S100* genes (S100α, nine separate clones; S100β, 3; CACY, 10; and CAPL, 1). Twelve clones revealed no sequence homology to *S100* genes. Four clones, with identical sequence within the amplified region, were homologous to *S100* genes but could not be attributed to published *S100* sequences. This suggested that we had identified an unrecognized *S100* gene, which we designated *S100D*. The 5' and 3' ends of the expected cDNA were amplified by RACE (19). To verify the linkage of the 5'- and 3'-RACE products, the entire translated part was amplified from human kidney and heart RNA and sequenced.

The complete S100D cDNA (709 base pairs, not shown) contains an unusually long 5' untranslated leader sequence, compared to other S100 cDNAs. In contrast to all other S100 cDNAs, S100D codes for additional 18 amino acids at the N terminus (Fig. 1). Hydrophilicity analysis revealed that this protein domain is highly hydrophobic. No sequence homology between this *N*-terminal sequence and other proteins was found by searching GenBank and Protein Identification Resource data bases. By comparison of the deduced amino acid sequence to other members of the S100 protein family, S100D clearly was identified as an S100 protein (the first calcium loop contains 14 amino acids; Fig. 1). Highest degrees of identities were found to CACY (54%), CAPL (53%), and S100L (52%).

In our sequence analysis of the *S100* gene cluster (described below), we discovered genomic subclones coding for a second unreported *S100* gene, termed *S100E*. The 5' end of the cDNA was amplified by RACE, and cDNA clones spanning all three exons and coding for the whole translated part were generated by PCR. Different subcloned PCR products of lung and heart RNA revealed no sequence variation when compared to the genomic DNA. The S100E cDNA (738 base pairs, not shown) contains one long open reading frame with strong sequence homology to the amino acid sequence of S100 proteins and the characteristic calcium loop spanning 14 amino acids. The highest sequence identities of the deduced S100E amino acid sequence were found to CACY, CAPL, and S100D (46%). Different from all other S100 proteins is the high content of cysteines in S100E. Ten of a total of 101 amino acids are cysteine residues. Five of them are located within the terminal 21 amino acids. We found no sequence similarities to metallothioneins or zinc-finger proteins—two protein families with increased cysteine occurrence. Like S100D, S100E is unique in the S100 protein family (Fig. 1).

On Northern blots no transcripts of S100D and S100E were detectable. We therefore applied a quantitative PCR technique to determine the expression patterns. Fragments of S100D, S100E, and human β-actin cDNAs as a control were amplified by PCR using equal amounts of reverse-transcribed RNA as template (results not shown). *S100D* and *S100E* both showed a very low general expression in diaphragm, heart muscle, skeletal muscle, stomach, lung, liver, fat tissue, and placenta. In addition, *S100D* (but not *S100E*) was also expressed in neuronal tissues. Using various primer sets, in most human tissues *S100E* was expressed at higher levels than *S100D*. Compared to other *S100* genes, both show, however, low transcription levels. In contrast, CACY and CAPL, both located in the same gene cluster (see below), are clearly detectable on Northern blots and are highly expressed in various human tissues (18).

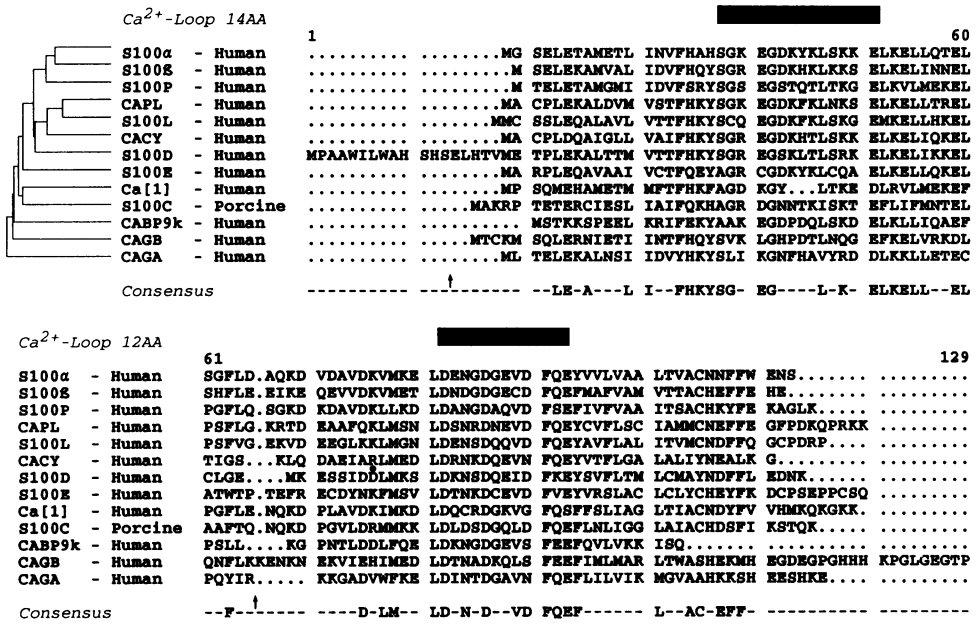

FIG. 1. Amino acid sequence comparison of S100D and S100E with all known S100 proteins. To get a maximum of homology in the alignments, gaps were introduced at those positions where an intron interrupts the first and second EF-hand. The positions of the two calcium-binding loops are indicated by black bars. The dendrogram reflects sequence relationship among the different S100 proteins. The position of an allelic sequence variation (D → G) found in S100D is indicated by a large black dot. Intron positions are marked by an arrow.

The Complete Coding Region of Four *S100* Genes Is Organized Within 15 kilobases of the Human Genome. In a search of the EMBL DNA data base with the complete cDNA sequence of S100D, we found a 100% identity of the S100D 3′ cDNA end (base pairs 494–709) to the promoter region of human *CACY* (14) (base pairs −1371 to −1175). The most simple explanation of this sequence identity is that in the human genome, the two *S100* genes are located directly one after another in a head-to-tail fashion. To prove the direct linkage of *S100* genes, we analyzed the gene region surrounding the *CACY* gene. Two genomic libraries were screened with cDNA probes corresponding to human *CAPL*, *S100D*, and *CACY*. Only phages that hybridized to at least two of the probes were analyzed. Four different, overlapping phages, corresponding to 33 kb of genomic DNA, were mapped by restriction enzyme digestion and Southern hybridization (Fig. 2). In our attempt to map the gene order of the three *S100* genes, we found an additional unknown *S100* gene localized close to the human *CAPL* gene. Isolation and characterization of the corresponding complete cDNA, designated S100E, has been described above. The genes of *S100E*, *CAPL*, and *S100D* (except parts of the third *S100D* intron) were sequenced completely. Since the *CACY* gene has been published previously (14), this genomic region was only partially sequenced and found to be identical to the published sequence. The four *S100* genes are arranged in a direct head-to-tail order: 5′-*S100E*-*CAPL*-*S100D*-*CACY*-3′ and are transcribed in the same direction (Fig. 2). By partially sequencing the 5′ region of *S100E* and the 3′ region of *CACY*, we obtained no further indications for the localization of other *S100* genes in our cloned fragments.

Like other *S100* genes studied so far, the genes of *CACY* (14), *S100E*, and *CAPL* (24, 25) are composed of three exons and two introns. In all cases the two EF-hand coding sequences are on separate exons (exons II and III). The first exon is not translated in any one of the *S100* genes. The gene structure of *S100D* is, in contrast, organized into four exons and three introns. The third and the fourth exon, coding for the two EF-hands, correspond to the second and third exon in other *S100* genes. Interestingly, exon II of *S100D* codes for 13 of the additional 18 amino acids that are not present in other S100 proteins. Exon I is not translated into protein sequence.

No direct evidence for gene duplication of parts of this cluster was seen. Intron, intergenic, and noncoding sequences of *S100E*, *CAPL*, *S100D*, and *CACY* show no sequence homologies to each another. Therefore, the division of *S100* genes by gene duplication of an ancestral gene must have occurred early in evolution. Further comparison of the human *CAPL* gene with the corresponding genes of rat

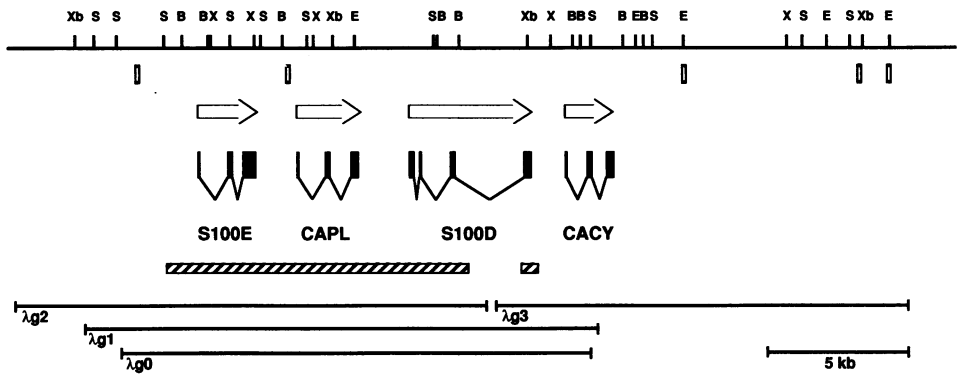

FIG. 2. Organization of the human *S100* gene cluster. Restriction enzyme maps in the top line for *Bam* I (B), *Eco*RI (E), *Sst* I (S), *Xba* I (Xb), and *Xho* I (X) were deduced by single and double digests of genomic DNA-containing phages (λg0 to λg3). Below this line, identified *Alu* family repeats are designated by open boxes. Filled boxes in the center connected by thin lines correspond to exon sequences. The direction of transcription is indicated by open arrows. The regions marked by hatched bars were sequenced completely and submitted to the GenBank data base.

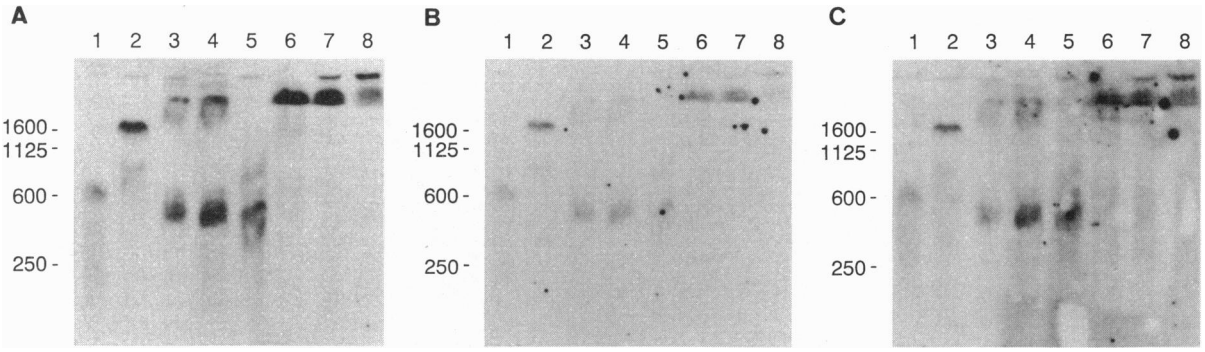

FIG. 3. Autoradiographs of a human PFGE Southern blot sequentially hybridized with S100α (*A*), S100L (*B*), and CAPL (*C*) sizes are shown in kb. Fragments in lanes: 1, *Not* I; 2, *Sal* I; 3, *Bss*HII; 4, *Bss*HII/*Nru* I; 5, *Bss*HII/*Mlu* I; 6, *Mlu* I/*Nru* I; 7, *Mlu* I; 8, *Nru* I.

(24) and mouse (25) revealed a high degree of similarity, even outside the exon sequences (exons: 76–87% identity; introns, 5′ and 3′ surrounding sequences: 64–85% identity).

The total (G+C) content of the sequenced parts approaches about 56%. This is higher than the average (G+C) content of the human genome. In addition, in all four genes the (G+C) content of the third-codon-position bases is higher than 75%. At least five different *Alu* family repeats are present in the cloned region (Fig. 2). This indicates that the gene cluster might be located within "T-bands" of the chromatin (10). The total gene complex analyzed in this work spans a region of 15 kb (5′ end of *S100E* to 3′ end of *CACY*) and contains the complete coding sequences for four different genes. As far as we know, this is the tightest human gene cluster reported up to now.

At Least Six *S100* Genes Are Linked Within 450 kb of Genomic DNA. The chromosomal colocalization of *S100* genes (12, 14–16) on the human chromosome 1 suggests that our cluster is only part of a larger *S100* gene cluster. Therefore, we analyzed a more extended gene region by PFGE. Three S100 and a β-actin cDNA probes were successively hybridized: CAPL as a member of the four-gene cluster, S100L (the cDNA was amplified by PCR and corresponds to the sequence of clone CaN19) (12) and S100α (18). We chose S100L, as this cDNA has the highest homology to the four *S100* genes of the cluster, and S100α as a cDNA with higher sequence divergence. All three S100 cDNA probes revealed an identical hybridization pattern (Fig. 3; β-actin hybridization resulted in a different pattern and is not shown). The smallest fragment that hybridized to the three probes was generated by digestion with *Bss*HII. The length of this fragment was calculated to be about 450 kb.

These results clearly demonstrate the existence of a large *S100* gene cluster containing at least six genes (*S100E, CAPL, S100D, CACY, S100L*, and *S100α*) within 450 kb.

Assignment of the *S100* Gene Cluster to Human Chromosome 1q21. To precisely localize the *S100* gene cluster on chromosome 1, *S100D* and *S100α* (not shown) were mapped by *in situ* hybridization. In 100 metaphase cells analyzed by each probe 16.1% (*S100α*) and 18.9% (*S100D*) of the silver grains were located on chromosome 1. Of these 62.3% (*S100α*) and 62.5% (*S100D*) mapped to the region 1q21. These results allowed us to map *S100α* and *S100D* (together with *S100E, CAPL*, and *CACY*) to the 1q21 band of the human genome. They are in agreement with the previous localization of *CACY* to 1q21-25 (14), *CAPL* to 1q12-22 (15), and *S100α* to chromosome 1 (16).

DISCUSSION

The tight linkage of six *S100* genes was established by gene cloning and PFGE. Two unreported *S100* genes were identified in this gene cluster. Both exhibit unique features within the *S100* gene family. *S100D* gene product contains a long hydrophobic N terminus, which may interact with hydrophobic binding partners. Ca[1] (26) and CACY (27), for example, interact with members of the annexin protein family. Alternatively, it may function as a signal peptide for secretion of the protein. Other S100 proteins display extracellular functions (11, 28, 29) but lack signal peptide-like sequences.

S100E has the highest content of cysteines of all calcium-binding proteins. Possibly, cystine bridges are formed that stabilize tertiary structures of S100E and might affect calcium-binding properties. The correct formation of disulfide-bridged dimers is thought to be important for the biological activity of S100 proteins (11). The deletion of a cysteine residue in S100β, for example, leads to a loss of function (29).

The sequenced region of the gene cluster contains four *S100* genes in a head-to-tail order (5′-*S100E-CAPL-S100D-CACY*-3′). Two additional *S100* genes (*S100L* and *S100α*) were linked to the gene cluster by PFGE within a maximal region of 450 kb in the human genome. In addition, the chromosomal colocalization of *CAGA* and *CAGB* (15) and the assignment of Ca[1] gene to a syntenic region on mouse chromosome 3 (30), implicates the existence of an even larger gene cluster (15) (Table 1).

Table 1. Chromosomal assignment of *S100* genes

Gene	Synonyms	Chromosome		Ref.
		Human	Mouse	
S100α		1q21	—	*
S100β	NEF	21q22	—	31
S100P		—	—	
CAPL	p9ka, 42A, pEL98, mts1, metastasin, calvasculin, 18A2	1q21	3	* 15
S100L	CaN19	1q21	—	*
CACY	2A9, PRA, CaBP, 5B10, calcyclin	1q21	3	* 32
S100D		1q21	—	*
S100E		1q21	—	*
Ca[1]	p11, p10, 42C, calpactin light chain	—	3	30
S100C	calgizzarin	—	—	
CaBP9k	ICaBP, Calbindin-D9k	X	—	33
CAGB	CFAg, MRP14, p14, MAC387, 60B8Ag, L1Ag, MIF, NIF	1q12-21	3	15
CAGA	CFAg, MRP8, p8, MAC387, 60B8Ag, L1Ag, CP-10, MIF, NIF	1q12-21	3	15

Synonyms of S100 proteins used by various authors. *, Results of authors; —, not determined.

This report identifies a gene cluster of calcium-binding proteins. Other EF-hand calcium-binding proteins with strong sequence relations were assigned to different human chromosomes (34). Calmodulin genes (presently three genes are known) are not clustered (35).

The overall expression pattern of the *S100* genes within this cluster is not comparable to those of other gene clusters. Most S100 proteins are expressed in more than one tissue. But each S100 protein is restricted to its own specific set of (different) cell types (11). S100α, for example, is expressed in neurons, heart muscle cells, and slow twitch muscle cells (36), whereas CACY is expressed in fibroblasts and epithelial cells (37). So far we could also not find a temporal correlation. In contrast, all genes of the globin gene clusters are expressed only in one particular cell type, the erythroid cells (5). Furthermore, the expression of globin and homeobox genes is dependent on the gene order within the clusters in either temporal or spatial manner (7, 9). Therefore, we suggest a regulatory mechanism that allows individual and independent expression of each *S100* gene in the cluster. The heterogeneity in the *S100* promoter regions supports this idea (sequence analysis data of the promoter regions are not shown).

The *in situ* assignment to chromosome 1q21 and the nucleotide composition of the sequenced region suggests that the gene cluster is located on a chromosomal "T-band." These are highly active, G+C-rich, and *Alu*-rich chromosomal regions that represent only 15% of all chromosome bands but contain 65% of mapped genes, 60% of cancer-associated breaks, and 42% of x-ray-induced breaks (10). In that respect it is intriguing that in human solid tumors, the single chromosome most frequently involved in structural alterations is chromosome 1 (38). One of the regions affected is 1q21. Chromosome abnormalities around 1q21 were found in breast carcinoma with a frequency of 30% (39, 40), gastric carcinoma (41), and leukemic B-cell clones (42). Furthermore, at least three *S100* genes located within the cluster show a deregulated expression in association with tumor progression. *CAPL* expression is increased in mouse mammary carcinoma or leukemic cells with a higher degree of metastatic potential (13). *CACY* is expressed at higher levels in cells from patients with acute myeloid leukemia than in normal cells (14). On the contrary, *S100L* is preferentially expressed in normal mammary epithelial cells but not in breast tumor cells (12). This suggests that chromosomal aberrations may alter the organization of *S100* genes during tumor progression. The occurrence of translocations, duplications, or deletions of chromosome segments might lead to a deregulated expression. One example for a gene whose expression is associated with neoplastic transformation is *HOX-11*. Overexpression of *HOX-11* in T-cell acute lymphoblastic leukemia is caused by translocations with breakpoints in the 5′ regulatory region of the gene (43).

We are grateful to Dr. P. Deloukas and Dr. A. P. G. M. van Loon (Hoffmann–La Roche, Basel) for providing technical assistance on PFGE, to E. Wey for her help in performing RACE experiments, and to M. Killen and M. Pedrocchi for critical reading of the manuscript. This work was supported by Grants no. 31-30742.91 and no. 32-029975.90 of the Swiss National Science Foundation and by the Helmut Horten Stiftung.

1. Ikemura, T., Wada, K.-N. & Aota, S.-I. (1990) *Genomics* **8**, 207–216.
2. Zerial, M., Salinas, J., Filipski, J. & Bernardi, G. (1986) *Eur. J. Biochem.* **160**, 479–485.
3. Huxley, C. & Fried, M. (1990) *Mol. Cell. Biol.* **10**, 605–614.
4. Ferrús, A., Lamazares, S., de la Pompa, J. L., Tanouye, M. A. & Pongs, O. (1990) *Genetics* **125**, 383–398.
5. Higgs, D. R., Vickers, M. A., Wilkie, A. O. M., Pretorius, I.-M., Jarman, A. P. & Weatherall, D. J. (1989) *Blood* **73**, 1081–1104.
6. Osley, M. A. (1991) *Annu. Rev. Biochem.* **60**, 827–861.
7. Krumlauf, R. (1992) *BioEssays* **14**, 245–252.
8. Jarman, A. P., Wood, W. G., Sharpe, J. A., Gourdon, G., Ayyub, H. & Higgs, D. R. (1991) *Mol. Cell. Biol.* **11**, 4679–4689.
9. Hanscombe, O., Whyatt, D., Fraser, P., Yannoutsos, N., Greaves, D., Dillon, N. & Grosveld, F. (1991) *Genes Dev.* **5**, 1387–1394.
10. Holmquist, G. P. (1992) *Am. J. Hum. Genet.* **51**, 17–37.
11. Hilt, D. C. & Kligman, D. (1991) in *Novel Calcium-Binding Proteins: Fundamentals & Clinical Implications*, ed. Heizmann, C. W. (Springer, Berlin), pp. 65–103.
12. Lee, S. W., Tomasetto, C., Swisshelm, K., Keyomarsi, K. & Sager, R. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2504–2508.
13. Ébralidze, A., Tulchinsky, E., Grigorian, M., Afanasyeva, A., Senin, V., Revazova, E. & Lukanidin, E. (1989) *Genes Dev.* **3**, 1086–1093.
14. Ferrari, S., Calabretta, B., de Riel, J. K., Battini, R., Ghezzo, F., Lauret, E., Griffin, C., Emanuel, B. S., Gurrieri, F. & Baserga, R. (1987) *J. Biol. Chem.* **262**, 8325–8332.
15. Dorin, J. R., Emslie, E. & Van Heyningen, V. (1990) *Genomics* **8**, 420–426.
16. Morii, K., Tanaka, R., Takahashi, Y., Minoshima, S., Fukuyama, R., Shimizu, N. & Kuwano, R. (1991) *Biochem. Biophys. Res. Commun.* **175**, 185–191.
17. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.
18. Engelkamp, D., Schäfer, B. W., Erne, P. & Heizmann, C. W. (1992) *Biochemistry* **31**, 10258–10264.
19. Frohman, M. A. (1990) *Amplifications* **5**, 11–15.
20. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
21. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
22. Parmentier, M., Passage, E., Vassart, G. & Mattei, M.-G. (1991) *Cytogenet. Cell Genet.* **57**, 41–43.
23. Kretsinger, R. H., Tolbert, D., Nakayama, S. & Pearson, W. (1991) in *Novel Calcium-Binding Proteins: Fundamentals and Clinical Implications*, ed. Heizmann, C. W. (Springer, Berlin), pp. 17–37.
24. Barraclough, R., Savin, J., Dube, S. K. & Rudland, P. S. (1987) *J. Mol. Biol.* **198**, 13–20.
25. Tulchinsky, E. M., Grigorian, M. S., Ébralidze, A. K., Milshina, N. I. & Lukanidin, E. M. (1990) *Gene* **87**, 219–223.
26. Gerke, V. (1991) in *Novel Calcium-Binding Proteins: Fundamentals and Clinical Implications*, ed. Heizmann, C. W. (Springer, Berlin), pp. 139–155.
27. Minami, H., Tokumitsu, H., Mizutani, A., Watanabe, Y., Watanabe, M. & Hidaka, H. (1992) *FEBS Lett.* **305**, 217–219.
28. Barger, S. W. & Van Eldik, L. J. (1992) *J. Biol. Chem.* **267**, 9689–9694.
29. Winningham-Major, F., Staecker, J. L., Barger, S. W., Coats, S. & Van Eldik, L. J. (1989) *J. Cell Biol.* **109**, 3063–3071.
30. Saris, C. J. M., Kristensen, T., D'Eustachio, P., Hicks, L. J., Noonan, D. J., Hunter, T. & Tack, B. F. (1987) *J. Biol. Chem.* **262**, 10663–10671.
31. Duncan, A., Higgins, J., Dunn, R. J., Allore, R. & Marks, A. (1989) *Cytogenet. Cell Genet.* **50**, 234–235.
32. Moseley, W. & Seldin, M. (1989) *Genomics* **5**, 899–905.
33. Howard, A., Legon, S., Spurr, N. K. & Walters, J. R. F. (1992) *Biochem. Biophys. Res. Commun.* **185**, 663–669.
34. Heizmann, C. W. & Hunziker, W. (1991) *Trends Biochem. Sci.* **16**, 98–103.
35. Fischer, R., Koller, M., Flura, M., Mathews, S., Strehler-Page, M.-A., Krebs, J., Penniston, J. T., Carafoli, E. & Strehler, E. E. (1988) *J. Biol. Chem.* **263**, 17055–17062.
36. Haimoto, H. & Kato, K. (1987) *J. Neurochem.* **48**, 917–923.
37. Kuznicki, J., Kordowska, J., Puzianowska, M. & Wozniewicz, B. M. (1992) *Exp. Cell Res.* **200**, 425–430.
38. Trent, J. M. (1985) *Breast Cancer Res. Treat.* **5**, 221–229.
39. Devilee, P., van Vliet, M., Bardoel, A., Kievits, T., Kuipers-Dijkshoorn, N., Pearson, P. L. & Cornelisse, C. J. (1991) *Cancer Res.* **51**, 1020–1025.
40. Wolman, S. R., Pauley, R. J., Mohamed, A. N., Dawson, P. J., Visscher, D. W. & Sarkar, F. H. (1992) *Cancer* **70**, Suppl. 6, 1765–1774.
41. Sano, T., Tsujino, T., Yoshida, K., Nakayama, H., Haruma, K., Ito, H., Nakamura, Y., Kajiyama, G. & Tahara, E. (1991) *Cancer Res.* **51**, 2926–2931.
42. Ghose, T., Lee, C. L. Y., Fernandez, L. A., Lee, S. H. S., Raman, R. & Colp, P. (1990) *Cancer Res.* **50**, 3737–3742.
43. Lu, M., Gong, Z., Shen, W. & Ho, A. D. (1991) *EMBO J.* **10**, 2905–2910.