Human Trichohyalin Gene Is Clustered with the Genes for Other Epidermal Structural Proteins and Calcium-Binding Proteins at Chromosomal Locus 1q21

Seung-Chul Lee, Mary Wang, O. Wesley McBride, Edward J. O'Keefe, In-Gyu Kim, and Peter M. Steinert

Skin Biology Branch (SCL, IGK, PMS), National Institute of Arthritis and Musculoskeletal and Skin Diseases and Laboratory of Biochemistry (MW, OWMcB), National Cancer Institute, National Institutes of Health, Bethesda, Maryland; and Department of Dermatology (EJO'K), University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, U.S.A.

Trichohyalin is a major differentiation product of hard keratinizing tissues such as the inner root sheath and medullary cells of the hair follicle and the filiform papillae of the tongue, as well as terminally differentiating epidermal cells. It consists largely of quasi-repeating peptide repeats and functions primarily as an intermediate filament-associated protein in these tissues. By mapping with human-rodent somatic cell hybrids and fluorescent *in situ* hybridization, we demonstrate that its gene maps to chromosomal region 1q21. Interestingly, genes encoding several other structural proteins expressed during terminal differentiation in the epidermis map to this region, as do also several members of the S-100 class of small calcium-binding proteins. *J Invest Dermatol 100:65–68, 1993*

Trichohyalin is a major protein product of terminally differentiating "hard" keratinizing tissues such as the inner root sheath and medullary cells of the developing hair follicle [1,2] and filiform papillae of tongue [3,4], as well as in the granular layer of terminally differentiating epidermis [3–5]. It is a large protein of about 200 kDa and adopts an elongated rod shape of about 85 nm with a distinct bead about 12 nm in diameter on one end [4]. In hair follicle cells, trichohyalin is a major substrate for transglutaminases, which catalyze the formation of N^{ϵ}-(γ-glutamyl)lysine isodipeptide crosslinks [6–8], as well as peptidyl arginine deiminase, which converts certain of its arginines to citrullines [9–13]. Current evidence suggests that it functions in these tissues primarily as an intermediate filament-associated protein [2], because it associates with the filaments of the cells [14] by isodipeptide crosslinks [8] in an apparent periodic manner [5]. However, its exact role in the interactions with the filaments of these various cell types remains to be clarified. Nor is it known whether the protein is involved in pathology.

A partial cDNA clone has been described for sheep trichohyalin [15]. We have recently described a 504-bp cDNA clone encoding the carboxyl-terminal end of human trichohyalin [16], and it shows significant homology, but not identity, to the sheep sequence. An antibody elicited against the carboxyl-terminal 16 amino acids cross-reacts with human and pig trichohyalins by Western blotting and recognizes human trichohyalins in the hair follicle, tongue, and epidermis by indirect immunofluorescence [16]. Further characterization of overlapping cDNA clones, which have provided the full-length nucleic acid and deduced amino acid sequence of human trichohyalin, indicate that it consists largely of tandemly arranged peptide repeats that have not been well conserved.* However, the amino-terminus contains two functional calcium-binding domains of the EF-hand type,* typically found in small S-100–like calcium-binding proteins [17]. In this paper, we have localized the human trichohyalin gene to chromosomal region 1q21, among a cluster of genes encoding other epidermally expressed structural proteins and calcium-binding proteins.

MATERIALS AND METHODS

A 504-bp cDNA clone encoding the carboxyl-terminal end of human trichohyalin generated by PCR of human genomic DNA [16] was utilized to probe DNAs from a panel of hamster-human and mouse-human somatic cell hybrids [18]. The same clone was used to screen a human genomic placental DNA library in λ phage EMBL-3 (Clontech Laboratories Inc, Palo Alto, CA) using standardized procedures [19]. A 15-kbp genomic clone was isolated and plaque purified. Restriction enzyme and Southern blotting methods indicated that a 6.5-kbp Sac I portion contained most of the coding sequences. This was then isolated, subcloned into pGEM 3z (Promega, Madison, WI), labeled with biotin, and used for regional localization of the trichohyalin gene in metaphase spreads of human chromosomes [19].

Manuscript received August 12, 1992; accepted for publication October 13, 1992.

Reprint requests to: Dr. Steinert, Skin Biology Branch, NIAMS, Building 6, Room 425, NIH, Bethesda, MD 20892.

Abbreviations:
cM: centiMorgan
Mbp, kbp, bp: mega-, kilo-, base pair
θ: recombination frequency
Z_{max}: lod score

* Lee S-C, Kim I-G, Marekov L, O'Keefe EJ, Steinert PM (manuscript in preparation).

0022-202X/93/$06.00

Table I. Segregation of the Human Trichohyalin Gene with Chromosome 1[a]

Human Chromosome Number	Gene/Chromosome +/+	+/−	−/+	−/−	% Discordancy
1	27	0	0	43	0
2	21	6	5	38	16
3	19	8	9	34	24
4	25	2	24	19	37
5	22	5	4	39	13
6	25	2	19	24	30
7	17	10	9	34	27
8	16	11	16	27	39
9	21	6	11	32	24
10	11	16	7	36	33
11	18	9	11	32	29
12	13	14	14	29	40
13	15	12	7	36	27
14	17	10	14	29	34
15	18	9	15	28	34
16	14	13	18	25	44
17	24	3	16	27	27
18	19	8	8	35	23
19	22	5	7	36	17
20	21	6	18	25	34
21	19	8	21	22	41
22	17	10	12	31	31
X	15	12	17	26	41

[a] The human trichohyalin gene was detected as a 6.8-kbp band in EcoRI digests of somatic cell hybrid DNAs after Southern blotting with the 0.5-bp cDNA probe. This band was well resolved away from a 5.3-kbp crosshybridizing band in Chinese hamster DNA digests, and there was no significant crosshybridization with mouse DNA. Detection of the human gene is correlated with the presence or absence of each human chromosome in the group of somatic cell hybrids. Discordancy indicates the presence of the gene despite absence of the chromosome (+/−) or absence of the gene despite the presence of the chromosome (−/+), and the sum of these numbers divided by the total number of hybrids examined (×100) represents the percentage discordancy. A 0% discordancy is the basis for chromosomal assignment. The human-hamster hybrids consisted of 29 primary and 13 subclones (13 positive of 42 total), and the human-mouse hybrids represented 15 primary clones and 13 subclones (14 of 28 total) [18].

RESULTS

The human trichohyalin gene† was localized by Southern blotting analyses of DNAs isolated from a panel of human-rodent somatic cell hybrids using a 504-bp cDNA clone. The gene was unambiguously localized to chromosome 1, and it segregated discordantly (>13%) with all other chromosomes (Table I). It was possible to further localize the gene to the region 1p12–1q31 by examination of several hybrids containing spontaneous breaks or translocations. One human-hamster hybrid that retained distal 1p with the break occurring between the oncogenes NRAS (1p12) and MYCL1 (1p32) did not contain human trichohyalin. Another human-hamster and one human-mouse hybrid each contained chromosome 1p with a break between NRAS (1p12) and a cluster of genes at 1q21, including a β-glucosidase pseudogene (GBAP), profilaggrin (FLG), and loricrin (LOR), and these hybrids did not contain human trichohyalin. Finally, one human-mouse hybrid contained a break between GBAP (1q21) and a calcium-activated ATPase (ATP2B2) (1q31) with a loss of distal 1q, but it retained trichohyalin. Interestingly, the trichohyalin, profilaggrin, and loricrin genes cosegregated in all hybrids.

The trichohyalin gene was then localized regionally on chromosome 1 by fluorescent *in situ* hybridization of metaphase spreads with a biotinylated 6.5-kbp genomic probe. Replication banding of the spreads was also performed to permit localization of the signal on a specific band, and the entire chromosomal DNA was lightly stained with propidium iodide. This method was chosen because it affords greater sensitivity over *in situ* hybridization with a ^{3}H-labeled probe, followed by autoradiography, to count the numbers and locations of silver grains. We found that the fluorescent signal was located in band 1q21 just below band 1q12 (Fig 1). In one metaphase spread exhibiting a very extended chromosome 1, the fluorescent signal was localized to the boundary of 1q21.1–1q21.2 (Fig 1 *insert*). The position of the hybridization signal was also determined as a fraction of the total length of chromosome 1 by measuring its location on photographically enlarged metaphase projections. In 13 copies of chromosome 1 on 10 metaphase spreads, the estimated fractional distance was 0.535 ± 0.035 (SD).

DISCUSSION

In the present paper, we show that the human trichohyalin gene maps to the chromosome 1q21 locus. Interestingly, the genes for several other proteins expressed in terminally differentiating epidermis have now been localized to this region. These include profilaggrin, the major matrix protein in the epidermis involved in interactions with the keratin intermediate filaments [20–22], as well as the cell envelope structural proteins involucrin [23], loricrin [24], and a proline-lysine–rich protein [25]. Each of these proteins is composed largely of repeating peptides of variable length and sequence that have not been precisely conserved between different species or even between different human individuals in the population. This has been taken as evidence that these genes are relatively modern and perhaps are still evolving [26]. Of these, trichohyalin* and profilaggrin [22] are unusual in that they both possess two functional calcium-binding domains of the EF-hand type at their amino termini. Interestingly, the genes encoding several small calcium-binding proteins of the S-100 class that contain two EF-hands on their amino termini have also been mapped to the 1q21 region [17], including calcyclin,‡ which is expressed in abundance in the epidermis. This raises the fascinating possibility that trichohyalin and profilaggrin are in fact "hybrid" genes, having been formed by the fusion of a repeating peptide structural gene with an ancestral S-100–like gene [22]. Whereas the precise function of these calcium-binding motifs on trichohyalin and profilaggrin are as yet unknown, it is possible that they may serve by regulation of intracellular calcium levels. Alternatively, and perhaps coincidentally, these motifs on trichohyalin may be involved in the auto-regulation of its own crosslinking by transglutaminases and modifications by peptidyl arginine deiminase, both of which are calcium-dependent enzymes.

Our preliminary genetic linkage analyses suggest that the loricrin and profilaggrin genes are tightly linked ($\theta = 0.018$, $Z_{max} = 20.0$), or within about 1.8 cM.§ Because each cM equals a physical distance of about 1 Mbp in human chromosomes, this suggested to us that members of the cluster of genes that includes trichohyalin may be located within about 1.8 Mbp of each other. Indeed, preliminary physical mapping of this locus by use of pulsed-field gel electrophoretic techniques has identified a 500-kbp fragment that contains both the profilaggrin and trichohyalin genes.¶

† The three-letter term TRH has been suggested to the Nomenclature Committee of the Human Genome Mapping Workshops to denote the human trichohyalin gene.

‡ Hardas BD, Elder JT: Physical linkage of CRABP-II and calcyclin on chromosome 1q21 by yeast artificial chromosome cloning (abstr). J Invest Dermatol 98:569A, 1992

§ Yoneda K, McBride OW, Hohl D, Steinert PM: Structure of the human loricrin gene: linkage at 1q21 with profilaggrin and involucrin genes (abstr). J Invest Dermatol 96:561A, 1990

¶ Korge BK, Gan S-Q, Yoneda K, Compton JG, Volz A, Steinert PM, Zeigler A, Mischke D: Physical mapping of the epidermal differentiation markers loricrin, profilaggrin, and involucrin on chromosome 1q21 by pulsed field gel electophoresis (abstr). J Invest Dermatol 98:569A, 1992

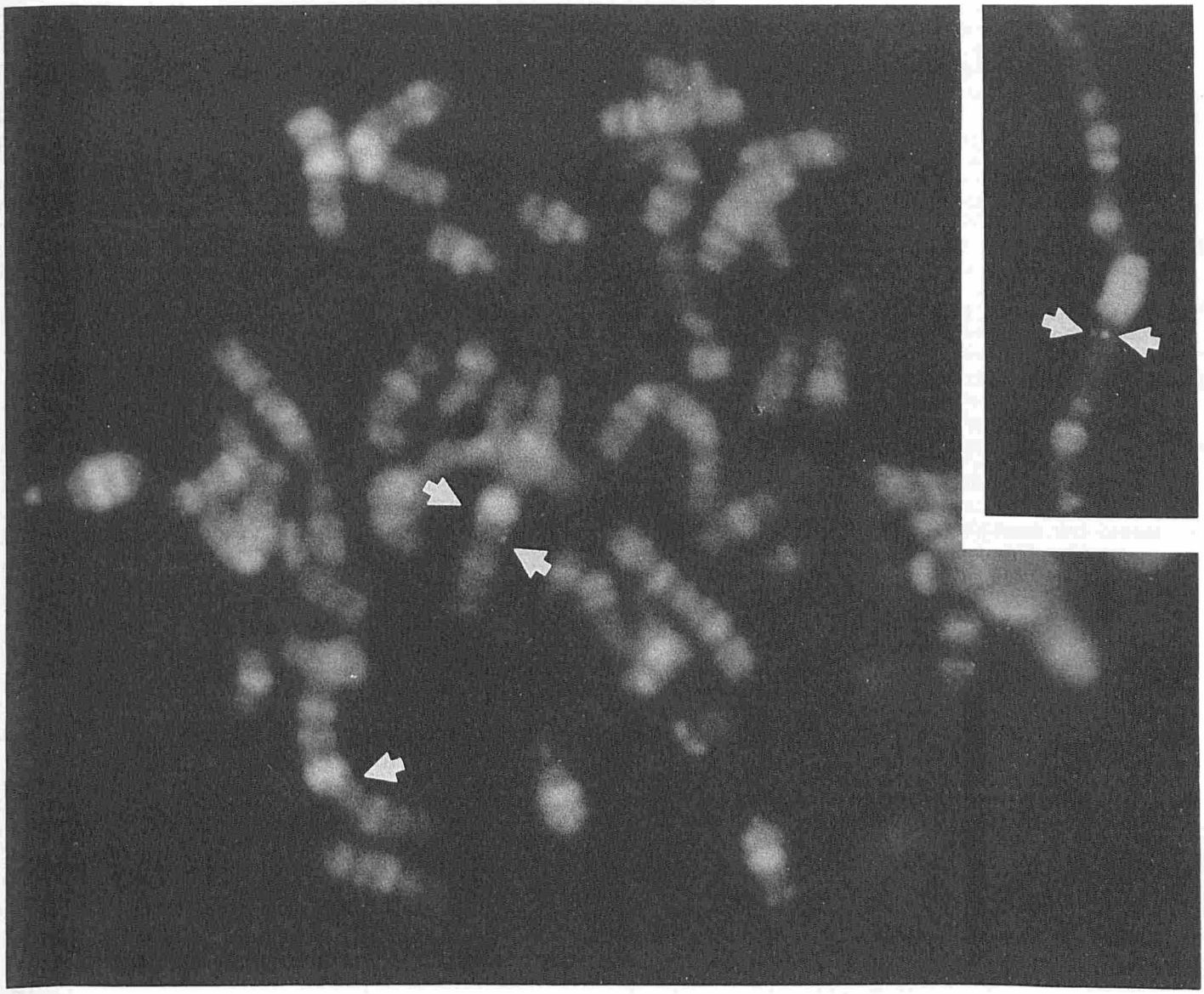

Figure 1. Chromosomal *in situ* hybridization using a genomic probe labeled with biotin. A typical metaphase spread is seen with the fluorescent signals located near 1q21.1 *(arrows)*. *Inset*, an extended chromosome 1 of another spread with the signals at 1q21.1–1q21.2 *(arrows)*. The entire region of confidence encompasses 1q21. For orientation purposes, the *bright band* encompasses 1q11 and 1q12; the fluorescent signals are located just distal to this band, in the 1q21.1–1q21.2 region.

Dr. Bernhard Korge kindly provided access to information on the physical mapping of the 1q21 locus, including data on trichohyalin. We thank Dr. Sherri Bale for valuable advice in preparation of the manuscript.

REFERENCES

1. Rogers GE: Some observations on the proteins of the inner root sheath of the hair follicle revealed by light and electron microscopy. Exp Cell Res 14:378–387, 1958
2. Rothnagel JA, Rogers GE: Trichohyalin, an intermediate filament-associated protein of the hair follicle. J Cell Biol 102:1419–1429, 1986
3. Hamilton EH, Payne RE Jr, O'Keefe EJ: Trichohyalin: presence in the granular layer and stratum corneum of normal human epidermis. J Invest Dermatol 96:666–672, 1991
4. Hamilton EH, Sealock R, Wallace NR, O'Keefe EJ: Trichohyalin: purification from porcine tongue epithelium and characterization of the native protein. J Invest Dermatol 98:881–889, 1992
5. O'Guin WM, Sun T-T, Manabe W: Interaction of trichohyalin with intermediate filaments: three immunologically defined stages of trichohyalin maturation. J Invest Dermatol 98:24–32, 1992
6. Harding HWJ, Rogers GE: $\epsilon(\gamma$-glutamyl)lysine crosslinkage in citrulline-containing protein fractions from hair. Biochemistry 10:624–630, 1971
7. Harding HWJ, Rogers GE: Isolation of peptides containing citrulline and the crosslink, $\epsilon(\gamma$-glutamyl)lysine from hair medulla protein. Biochim Biophys Acta 427:315–324, 1976
8. Steinert PM: Structural features of the α-type filaments of the inner root sheath cells of the guinea pig hair follicle. Biochemistry 17:5045-5052, 1978
9. Steinert PM, Harding HWJ, Rogers GE: The characterization of protein-bound citrulline. Biochim Biophys Acta 175:1–9, 1969
10. Rogers GE, Harding HWJ, Llewellyn-Smith IJ: The origin of citrulline-containing proteins in the hair follicle and the chemical nature of trichohyalin, an intracellular precursor. Biochim Biophys Acta 495:159–175, 1977

11. Kubilus J, Waitkus RW, Baden HP: Partial purification and specificity of an arginine-converting enzyme from bovine epidermis. Biochim Biophys Acta 615:246–251, 1980
12. Fujisaki M, Sugawara K: Properties of peptidylarginine deiminase from the epidermis of newborn rats. J Biochem (Tokyo) 89:257–264, 1981
13. Rogers GE, Rothnagel JA: A sensitive assay for the enzyme activity in hair follicles and epidermis that catalyses the peptidyl-arginine-citrulline post-translational modification. In: Seiji M, Bernstein IA (eds.). Normal and Abnormal Epidermal Differentiation. University of Tokyo Press, 1983, pp 171–183
14. Steinert PM, Dyer PY, Rogers GE: Isolation of non-keratin protein filaments from inner root sheath cells of the hair follicle. J Invest Dermatol 56:49–54, 1971
15. Fietz MJ, Presland RB, Rogers GE: The cDNA-deduced amino acid sequence for trichohyalin, a differentiation marker in the hair follicle, contains a 23 amino acid repeat. J Cell Biol 110:427–436, 1990
16. O'Keefe EJ, Hamilton EH, Lee S-C, Steinert PM, Sealock R, Wallace NR: Trichohyalin: a structural protein of hair, tongue and epidermis. J Invest Dermatol (in press)
17. Dorin JR, Emslie E, vanHeyningen V: Related calcium-binding proteins map to the same subregion of chromosome 1q and to an extended region of synteny on mouse chromosome 3. Genomics 8:420–426, 1990
18. McBride OW, Battey J, Hollis GF, Swan DC, Siebenlist, U, Leder P: Localization of human variable and constant region immunoglobulin genes on subtelomeric band q32 of chromosome 14. Nucleic Acids Res 10:8155–8170, 1982
19. Kim I-G, McBride OW, Wang M, Kim S-Y, Idler WW, Steinert PM: Structure and organization of the human transglutaminase 1 gene. J Biol Chem 267:7710–7717, 1992
20. McKinley-Grant LJ, Idler WW, Bernstein IA, Parry DAD, Cannizzaro L, Croce CM, Huebner K, Lessin SR, Steinert PM: Characterization of a cDNA clone encoding human filaggrin and localization of the gene to chromosome region 1q21. Proc Natl Acad Sci USA 86:4848–4852, 1989
21. Gan S-Q, McBride OW, Idler WW, Markova N, Steinert PM: Organization, structure and polymorphisms of the human profilaggrin gene. Biochemistry 29:9432–9440, 1990
22. Markova N, Chipev CC, Gan S-Q, Marekov L, Idler WW, Steinert PM: Profilaggrin is a major epidermal calcium binding protein. Mol Cell Biol (in press)
23. Eckert RL, Green H: Structure and evolution of the human involucrin gene. Cell 46:583–589, 1986
24. Yoneda K, Hohl D, McBride OW, Wang M, Cehrs KU, Idler WW, Steinert PM: The human loricrin gene. J Biol Chem 267:18060–18066, 1992
25. Kartasova T, van de Putte P: Isolation, characterization, and UV-stimulated expression of two families of genes encoding polypeptides of related structure in human epidermal keratinocytes. Mol Cell Biol 8:2195–2203, 1988
26. Simon M, Philips MG, Green H: Polymorphism due to variable numbers of repeats in the human involucrin gene. Genomics 9:576–580, 1991