Kagawa, Japan; [7]Department of Dermatology, Nagoya University Graduate School of Medicine, Nagoya, Japan.

***Correspondence and reprint requests to**:

Ken Natsuga, M.D., Ph.D.

Department of Dermatology,

Hokkaido University Graduate School of Medicine,

North 15 West 7, Sapporo 060-8638, Japan

Telephone: +81-11-716-1161, ext. 5962

Fax: +81-11-706-7820

e-mail <natsuga@med.hokudai.ac.jp>

Abstract

Numerous inherited diseases develop due to missense mutations leading to an amino acid substitution. Whether an amino acid change is pathogenic depends on the level of deleterious effects caused by the amino acid alteration. We show an example of different structural and phenotypic consequences caused by two individual amino acid changes at the same position. Epidermolysis bullosa simplex (EBS) is a genodermatosis resulting from *KRT5* or *KRT14* mutations. Mutation analysis of an EBS family revealed that affected individuals were heterozygous for a novel mutation of c.1237G>C (p.Ala413Pro) in *KRT14*. Interestingly, two of 100 unrelated normal controls were heterozygous and one of the 100 was homozygous for a different mutation in this position, c.1237G>A (p.Ala413Thr). *In silico* modeling of the protein demonstrated deleterious structural effects from proline substitution but not from threonine substitution. *In vitro* transfection studies revealed a significantly larger number of keratin clumped cells in HaCaT cells transfected with mutant *KRT14* cDNA harboring p.Ala413Pro than those transfected with wild-type *KRT14* cDNA or mutant *KRT14* cDNA harboring p.Ala413Thr. These results show that changes in two distinct amino acids at a locus are destined to elicit different phenotypes due to the degree of structural distortion resulting from the amino acid alterations.

Keywords: vesiculobullous skin diseases/ keratin/ dominant negative effect/ heterodimer/ coiled-coil

Introduction

Keratins are the largest group of intermediate filament (IF) proteins, which are expressed in epithelial cells (Schweizer *et al.*, 2006). The prominent IFs consist of keratins K1–K20, which are further classified into types I (K9–K20) and II (K1–K8). Type I and type II keratins form noncovalent type I/II keratin heteropolymers (Moll *et al.*, 1982). Unique keratin expression serves as specific markers that characterize different epithelial cell types. Of the many kinds of epithelial cells that exist, basal epidermal keratinocytes preferentially express K5/K14 heteropolymers (Moll *et al.*, 1982; Nelson and Sun, 1983). These, in turn, form predominantly heterodimers *in vivo* (Coulombe and Fuchs, 1990; Hatzfeld and Weber, 1990; Steinert, 1990) with chains parallel to one another and in axial register (Conway and Parry, 1990; Steinert *et al.*, 1993).

Epidermolysis bullosa (EB) comprises a group of heterogeneous disorders in which congenital skin fragility leads to dermal-epidermal junction separation. EB has been subdivided into three major groups (EB simplex, junctional EB, dystrophic EB) and one minor subtype (Kindler syndrome) based on the level of blister formation (Fine *et al.*, 2008). So far, mutations in 14 different genes have been identified as underlying EB

subtypes (Fine *et al.*, 2008; Groves *et al.*, 2010). Among them, mutations in either the *KRT5* or *KRT14* gene, which encode K5 and K14, respectively, lead to EB simplex (EBS) (Coulombe *et al.*, 1991; Lane *et al.*, 1992; Yasukawa *et al.*, 2006). According to the clinical severity of blister formation, EBS can be further subdivided into three major subtypes (Fine *et al.*, 2008). The mildest variant is "EBS, localized" (EBS-loc), with blistering confined to the hands and feet; the more moderate variant is "EBS, other generalized" (EBS-gen-nonDM), with generalized blister formation; and the most severe variant is Dowling-Meara type (EBS-DM), which is characterized by severe herpetiform blistering (Coulombe *et al.*, 2009; Fine *et al.*, 2008).

In this study, we identified a novel missense mutation in *KRT14* from a family with EBS-loc. We also detected a different nucleotide substitution at the same position in *KRT14* in normal control individuals. To clarify whether each nucleotide substitution is pathogenic, we used molecular modeling to predict the effect on the structure that results from a single amino acid change, and we examined cultured cells transfected either with wild-type or with mutated *KRT14* cDNA. Here we show that two kinds of amino acid changes at the same position of *KRT14* lead to totally different cell function *in vitro* and phenotypes *in vivo*.

Results

Case description.

The proband was a 10-year-old male with blistering on his soles (**Fig. 1a**). Blisters on his hands and feet were observed during infancy. New blisters tended to emerge more in the summer. Nail deformity was also noted (**Fig. 1b**). Ultrastructural features of skin specimens from the proband revealed many vacuoles scattered between the nucleus and basal lamina in the basal keratinocytes (**Fig. 1c**). His family had several affected members (**Fig. 1d**).

Detection of a novel *KRT14* mutation in a family with EBS and another nucleotide substitution at the same position of *KRT14* in normal controls.

Direct sequencing of the *KRT5/KRT14* gene revealed that the proband (III-2, **Fig.1d**) was heterozygous for the novel mutation of c.1237G>C transversion (p.Ala413Pro) in the helix termination motif of *KRT14* (**Fig. 2a**). No other mutations were detected in any of the exons or exon-intron boundaries of *KRT5* and *KRT14*. Affected family members were also heterozygous for the same mutation in *KRT14* (I-1, II-1, II-3, III-1 and III-5, **Fig. 1d**) and unaffected family members did not have that mutation (I-2, II-2 and III-3, **Fig. 1d**). Mutant allele-specific amplification analysis demonstrated that a

300-bp fragment derived from the mutant allele was amplified from the genomic DNA of affected family members (**Fig. 1d, 2f**). This mutation was not found in 200 normal unrelated alleles (100 Japanese individuals) by direct sequencing. Unexpectedly, 2 of the 100 normal Japanese controls were heterozygous for c.1237G>A transition (p.Ala413Thr) in *KRT14* (**Fig. 2b**), and 1 of the 100 normal controls was homozygous for the same nucleotide transition (**Fig. 2c**). These 3 individuals with the c.1237G>A transition (p.Ala413Thr) in *KRT14* had no history of skin fragility or nail dystrophy. Interestingly, the proband's unaffected grandmother (I-2) and affected uncle (II-3) were also heterozygous for the c.1237G>A transition (p.Ala413Thr) in *KRT14*. The proband's affected uncle (II-3) was compound heterozygous for p.Ala413Thr and p.Ala413Pro (**Fig. 2d**). However, his clinical manifestations were similar to those of the other affected family members. The proband's unaffected cousin (III-4) is expected to be heterozygous for p.Ala413Thr - mutation analysis could not be performed without her consent.

Molecular dynamics predicts the deleterious structural change resulting from p.Ala413Pro substitution in keratin 14 but not from p.Ala413Thr.

An initial structure of the native K5/K14 heterodimer and the p.Ala413Thr and

p.Ala413Pro mutants, representing the C-terminal 35 residues of each chain, was generated by comparative modeling. The first 29 residues of each chain form a classical coiled-coil in which generally hydrophobic residues occupy the heptad-repeat positions *"a"* and *"d"* (**Fig. 3a**) – a lysine at position 405 in K14 and 460 in K5, conserved amongst human IF proteins (Strelkov *et al.*, 2002), interact with each other through the hydrophobic portion of their side-chains.

The structure of the native K5/K14 and the two mutants were subjected to molecular dynamics (MD) simulations to explore the structural stability. The secondary structure content in each of the chains throughout the MD simulations is presented in **Fig. 3b-d** – in this figure α-helix conformation is represented in blue. In this native heterodimer (**Fig. 3b**), the α-helix content remains essentially unchanged throughout the simulation – both peptides maintain helical geometry across the coiled-coil domains (residues 449-474 in K5, and 394-419 in K14). At the C-termini of K5 and K14 a stable bend (green), flanked on the N-terminal side by a stable short turn (yellow), is observed. In K5 a short 3_{10} helix (grey) forms after ~16 ns and is stable for the remainder of the simulation. Thus, the structure, and particularly the secondary structure conformation, remains constant throughout the simulation in the native complex.

The p.Ala413Thr mutation introduces a slight instability in the helix at the site of the mutation (black triangle), represented by an infrequent change in conformation to 'turn' during the simulation (shown in yellow) (**Fig. 3c**). However, the overall α-helix content in this mutant is very similar to that observed for the native (**Fig. 3b**). In contrast, in the p.Ala413Pro peptide predominantly adopts a non-helical conformation at the position of the mutation throughout the simulation (**Fig. 3d**). This change in conformation is also associated with additional instability (turn and coil conformation) in the helical conformation of residues C-terminal of the mutation site in both the K5 and K14 peptides (**Fig. 3d**). Thus, the p.Ala413Thr mutation introduces a slight instability into the structure of the complex, whereas the p.Ala413Pro mutation causes significant disruption to the secondary structure in the C-terminal region of both peptides.

An alanine to proline substitution at the 413 locus of KRT14 protein disrupts the KIF network in HaCaT cells, while alanine to threonine does not.

We examined whether the clinical heterogeneity between individuals with wild-type, p.Ala413Thr and p.Ala413Pro mutations in *KRT14* could be demonstrated in a cell culture system. Three mammalian expression vectors containing the wild-type *KRT14*

cDNA (K14WT), the mutated *KRT14* cDNA correspondent with the p.Ala413Thr substitution (K14A413T) and the p.Ala413Pro substitution (K14A413P) were transiently transfected into HaCaT cells. Detection of K14 was performed by probing the V5 epitope tag. Immunoblot analysis confirmed that each construct was successfully transferred into the HaCaT cells (**Fig. 4a**). Transfection efficiency was up to 50% as judged from the immunofluorescence of the V5 epitope tag in transfected cells. The cells transfected with either K14WT or K14A413T showed fine bundles of keratin filaments extending throughout the cytoplasm without disturbing the cells' morphology (**Fig. 4b, 4c**), whereas cells transfected with K14A413P exhibited small ball-like filament aggregates indicating a disruption in the keratin network (**Fig. 4d**). The percentage of the transfected cells harboring keratin clumping in each transfection assay was as follows: K14WT (mean, 17%), K14A413T (6%) and K14A413P (49%) (**Fig. 4e**). Statistical analysis showed that significantly more clumped cells were observed in HaCaT cells transfected with K14A413P than in those transfected with K14WT or K14A413T ($p < 0.05$) (**Fig. 4e**).

Although not statistically significant ($p > 0.05$), fewer keratin clumped cells were observed in HaCaT cells transfected with K14A413T than those transfected with

K14WT (**Fig. 4e**). To clarify whether p.Ala413Thr mutation has a protective effect on keratin aggregation, three different transfections using HaCaT cells were performed, including K14A413P alone, a combination of equal amounts of K14A413P and K14A413T (K14A413P/K14A413T) and a combination of equal amounts of K14A413P and K14WT (K14A413P/K14WT). The percentage of the transfected cells harboring keratin clumping in each transfection assay was as follows: K14A413P (mean, 42%), K14A413P/K14A413T (32%) and K14A413P/K14WT (30%) (**Supplementary Fig. 1**). The differences in the percentage of keratin clumped cells between K14A413P/K14A413T and K14A413P/K14WT were not statistically significant ($p \geq 0.05$), indicating that the p.Ala413Thr mutation is unlikely to have a protective effect on keratin aggregation compared with the wild-type sequence.

We further performed three different transfections (K14WT, K14A413T and K14A413P) into HeLa cells, in which endogenous K5 and K14 were not expressed but K8 was present as a potential partner of recombinant K14 (Albers and Fuchs, 1987; Hatzfeld and Franke, 1985). The percentage of the transfected cells harboring keratin aggregation in each transfection assay was as follows: K14WT (mean, 14%), K14A413T (10%) and K14A413P (77%) (**Supplementary Fig. 2**). Statistical analysis

showed that significantly more clumped cells were observed in HeLa cells transfected with K14A413P than in those transfected with K14WT or K14A413T ($p < 0.05$) (**Supplementary Fig. 2**), confirming the deleterious effect of p.Ala413Pro mutation in K14 on keratin filament network. In contrast, no keratin clumping was observed in normal human epidermal keratinocytes (NHEK) transfected with any of three plasmids (K14WT, K14A413T and K14A413P) (**Supplementary Fig. 3**).

Discussion

A single-nucleotide change in the protein-coding region that leads to an amino acid substitution can be assigned either as a non-synonymous coding SNP (nsSNP) or as a missense mutation. Numerous single-gene diseases have been attributed to missense mutations. However, it is sometimes difficult to demonstrate the effects of an amino acid substitution on protein function and disease phenotype (Thusberg and Vihinen, 2009) – it has been recently shown that the p.Met119Thr and p.Met119Val mutations in K14 result in EBS-DM and EBS-gen-nonDM, respectively (Cummins *et al.*, 2001). Our study provides a good model to study the pathogenicity of a single amino acid substitution, since replacement of Ala^{413} in K14 with proline results in an EBS-family whereas replacement with threonine results in normal controls.

Ala^{413} is located in the helix termination motif of K14 and corresponds to position "*b*" of the heptad repeat (a*bcdefg*) that is conserved among keratin polypeptides (**Fig. 3e**), where "*a*" and "*d*" are usually non-polar amino acids and the others are polar or charged. Ala^{413} is highly conserved among species (**Fig. 3f**).

Proline lacks an amide hydrogen atom and is unable to form a hydrogen bond with the

carbonyl oxygen atom of the amide four residues N-terminal of it in an alpha helix – thus, proline residues act as alpha-helix disruptors (although the stability of alpha helices harboring a proline may be environment-specific (Li *et al.*, 1996)). Previous reports using crystal X-ray analysis showed marked structural perturbations by proline residues in polypeptides (MacArthur and Thornton, 1991). Keratin has a conserved structure with helix motifs, and proline substitutions in these motifs have been reported to cause marked structural perturbations (Letai *et al.*, 1992). As for EBS, proline substitutions in K5/K14 have been described in 34 cases in the Human Intermediate Filament Database (http://www.interfil.org/) (Szeverenyi *et al.*, 2008), although no nsSNPs have been detected in the database that cause proline substitutions. Clinical manifestations in EBS with proline substitutions vary between the localized and Dowling-Meara types, although proline substitutions in helix motifs tend to lead to EBS-DM, and those substitutions in domains outside helix motifs are more often observed in EBS-gen-nonDM.

A Taiwanese patient with EBS-loc has been described as heterozygous for p.Ala413Thr in K14 (Chao *et al.*, 2002). However, p.Ala413Thr was recently found in 6 of 112 alleles in normal Japanese individuals (Hattori *et al.*, 2006). Our study confirmed the

presence of normal control individuals who are heterozygous for p.Ala413Thr (**Fig. 2b**). Furthermore, a single normal control was homozygous for p.Ala413Thr (**Fig. 2c**), which lowers the possibility of pathogenic effects from threonine substitution at this amino acid. Threonine and alanine substitute for one another on a frequency commensurate with their occurrence in structured proteins (Henikoff and Henikoff, 1992). Thus, structurally they can be considered interchangeable. Proline, on the other hand, is seldom observed to substitute for any residue, including alanine, highlighting its unique structural characteristics. Our *in vitro* study using HaCaT cells confirmed that it is not threonine substitution but proline substitution that causes keratin aggregation. Nevertheless, it is possible that p.Ala413Thr is a phenotypical mutation in certain environmental conditions as many contributing factors including temperature, trauma and location of blister formation can affect the development of blistering phenotypes (Coulombe *et al.*, 2009).

Modeling of K5/K14 mutations in coiled-coil structures provides evidence that a correlation exists between the clinical severity of EBS and the degree of structural distortion caused by the underlying amino acid change (Smith *et al.*, 2004). Our study has also demonstrated that molecular dynamics simulations of keratin mutations

accurately correlate with the pathogenicity of an amino acid substitution.

As HaCaT cells are not normal keratinocytes, their keratin expression level is different from that of NHEK (Boukamp *et al.*, 1988). No keratin clumping was seen in NHEK transfected with any of K14WT, K14A413T or K14A413P (**Supplementary Fig. 3**). This result indicates that either overexpressed recombinant K14 is not enough to disrupt keratin network in NHEK due to a much higher expression of endogenous K14 in NHEK than in HaCaT cells (Sorensen *et al.*, 2003). On the other hand, the balance of keratin network may be substantially altered when recombinant K14 is overexpressed in HaCaT cells, in which endogenous K14 is reduced compared with NHEK (Sorensen *et al.*, 2003). The absence of keratin aggregation in the proband's skin keratinocytes (**Fig. 1c**), compared to what was observed in HaCaT cells transfected with K14A413P, may reflect the greater effect of K14 mutant introduction in HaCaT cells compared with NHEK.

In summary, this study gives insight into consequences of two different amino acid substitutions at the same codon. The biological effects of one amino acid substitution are hard to predict. *In silico* and *in vitro* analyses are useful for confirming the

pathogenicity of missense mutations.

Material and Methods

Mutation analysis. Genomic DNA extracted from peripheral blood was used as a template for polymerase chain reaction (PCR) amplification. The *KRT5* and *KRT14* genes were amplified by methods previously described (Hut *et al.*, 2000; Stephens *et al.*, 1997). DNA sequencing of the PCR products was carried out with an ABI 3100 sequencer (PerkinElmer Life Sciences-ABI, Foster City, CA). The mutation nomenclature follows published mutation nomenclature guidelines (www.hgvs.org/mutnomen) according to the reference sequence NM_000424.3 for *KRT5* and NM_000526.3 for *KRT14*, with +1 as the A of the ATG initiation codon.

Mutant allele-specific amplification analysis. To verify the mutation mutant allele-specific amplification (MASA) analysis was performed with mutant allele-specific primers carrying the substitution of two bases at the 3'-end mutant allele-specific primers (Linard *et al.*, 2002; Sapio *et al.*, 2006): forward, 5'-ACGCGGCTGGAGCAGGAGATTC-3'; reverse, 5'-GACAGCACTAGAGCTCAGCC-3'. PCR conditions were as follows: 94°C for 5 min, followed by 30 cycles of 94°C for 30 sec, 65°C for 30 sec, 72°C for 30 sec and extension at 72°C for 7 min.

Plasmid construction. cDNA containing the entire coding region of *KRT14* (K14WT) subcloned into the pcDNA 3.1/V5-His vector (Invitrogen) was employed (Yoneda *et al.*, 2004). The point-mutated *KRT14* cDNAs corresponding to the p.Ala413Pro (K14A413P) and p.Ala1237Thr (K14A413T) mutations were generated with the use of the GeneTailor Site-Directed Mutagenesis System (Invitrogen). Sense primers used for the PCR reactions to generate the K14A413P and the K14A413T fragments were 5'-CGCGGCTGGAGCAGGAGATC**c**CCACCTACCGC-3' and 5'-CGCGGCTGGAGCAGGAGATC**a**CCACCTACCGC-3', respectively (lower-case letters in bold denote mutations introduced). The anti-sense primer 5'-GATCTCCTGCTCCAGCCGCGTCTTCACGTC-3' was used to generate both of the mutant *KRT14* cDNAs.

Molecular modeling. The structure of segment 2B in human vimentin (PDB 1GK6) (Herrmann *et al.*, 2000; Strelkov *et al.*, 2002; Strelkov *et al.*, 2004) was used as a template in comparative modelling of the K5/K14 heterodimer, and the two mutations (K14-A413P and K14-A413T: 2B-108,*b*) using the MODELLER (9v7) program (Fiser and Sali, 2003). The nomenclature K14-A413P: 2B-108,*b* specifies residue 413 in chain

K14, its position 108 within segment 2B and its position *b* within the heptad repeat. From 25 models of each heterodimer the structure with the lowest Modeller Objective function was subjected to molecular dynamics simulation.

Molecular dynamics (MD) simulations using the GROMACS (v4.0) package of programs (Hess *et al.*, 2008) were performed using the OPLS-aa force field (Jorgensen and Tirado-Rives, 1988). Ionizable residues were assumed to be in their charged state, while the amine and carboxyl termini were assumed to be in their neutral form. Each molecule was solvated in a 75 x 75 x 75 $Å^3$ water box; sodium and chloride ions were added to neutralize the system and provide a final ionic strength of 0.154 M. Protein and water (with ions) were coupled separately to a thermal bath at 300 K using velocity rescaling (Bussi *et al.*, 2007) applied with a coupling time of 0.1 ps, while the pressure was coupled to an isotropic barostat using a time constant of 1 ps and compressibility of 4.5×10^{-5} bar^{-1}. All simulations were performed with a single non-bonded cut-off of 10 Å, and applying a neighbour-list update frequency of 10 steps (20 fs). The particle-mesh Ewald method was used to account for long-range electrostatics (Essman *et al.*, 1995), applying a grid width of 1.2 Å, and a fourth-order spline interpolation. Bond lengths were constrained using the LINCS algorithm (Hess, 2008; Hess *et al.*, 2008). All