Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural and biochemical analyses of the human PAD4 variant encoded by a functional haplotype gene

Naoki Horikoshi,[a] Hiroaki Tachiwana,[a] Kengo Saito,[a] Akihisa Osakabe,[a] Mamoru Sato,[b] Michiyuki Yamada,[b] Satoko Akashi,[c] Yoshifumi Nishimura,[c] Wataru Kagawa[a] and Hitoshi Kurumizaka[a]*

[a]Laboratory of Structural Biology, Graduate School of Advanced Science and Engineering, Waseda University, 2-2 Wakamatsu-cho, Shinjyuku-ku, Tokyo 162-8480, Japan, [b]Division of Macromolecular Crystallography, Graduate School of Nanobioscience, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan, and [c]Division of Structural Biology, Graduate School of Nanobioscience, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan

Correspondence e-mail: kurumizaka@waseda.jp

Received 19 November 2010
Accepted 9 December 2010

PDB References: PAD4, 3apn; PAD4^{SNP}, 3apm.

PAD4 is a peptidylarginine deiminase that catalyzes citrullination, a type of post-translational modification. In this reaction, arginine residues in proteins are converted to citrulline. PAD4 promotes the deimination of arginine residues in histones and may regulate transcription in the context of the chromatin. Single-nucleotide polymorphisms (SNP) in the gene encoding PAD4 identified it as one of the genes associated with susceptibility to rheumatoid arthritis. The PAD4 SNP involve three amino-acid substitutions: Ser55 to Gly, Ala82 to Val and Ala112 to Gly. Autoantibodies for improperly citrullinated proteins have been found in rheumatoid arthritis patients, suggesting that the PAD4^{SNP} mRNA is more stable than the conventional PAD4 mRNA and/or the PAD4^{SNP} protein possesses a higher citrullination activity than the PAD4 protein. In order to study the effects of the three amino-acid substitutions found in PAD4^{SNP}, the crystal structure of PAD4^{SNP} was determined and it was found that the amino-acid substitutions in PAD4^{SNP} only induced conformational changes within the N-terminal domain, not in the active centre for citrullination located in the C-terminal domain. Biochemical analyses also suggested that the citrullination activity of PAD4^{SNP} may not substantially differ from that of conventional PAD4. These structural and biochemical findings suggested that the improper protein citrullination found in rheumatoid arthritis patients is not caused by defects in the citrullination activity of PAD4^{SNP} but by other reasons such as enhanced PAD4^{SNP} mRNA stability.

1. Introduction

PAD proteins (protein-arginine deiminases, protein L-arginine iminohydrolases) are a family of proteins that catalyze the conversion of arginine residues to citrulline residues (deimination) in proteins in a Ca^{2+}-dependent manner (Vossenaar *et al.*, 2003). To date, five human genes encoding PAD proteins, PAD1, PAD2, PAD3, PAD4 and PAD6, have been cloned (Guerrin *et al.*, 2003; Ishigami *et al.*, 2002; Kanno *et al.*, 2000; Nakashima *et al.*, 1999; Chavanas *et al.*, 2004) and have been suggested to play essential roles in development and cell differentiation. These PAD proteins share about 50% identity in their amino-acid sequences.

Of the PAD proteins, only PAD4 has a nuclear localization signal and is localized in the nucleus (Nakashima *et al.*, 2002). PAD4 reportedly deiminates arginine residue at multiple sites in histones H3 and H4 and may regulate gene expression by

arginine citrullination of histones (Cuthbert *et al.*, 2004; Wang *et al.*, 2004; Hidaka *et al.*, 2005; Kearney *et al.*, 2005; Thompson & Fast, 2006; Raijmakers *et al.*, 2007; Denis *et al.*, 2009). The hypercitrullination of histones by PAD4 also mediates chromatin decondensation in granulocytes/neutrophils (Wang *et al.*, 2009). Structural studies revealed that PAD4 is composed of N-terminal and C-terminal domains (Arita *et al.*, 2004). The N-terminal domain (Met1–Pro300) is further divided into two immunoglobulin-like subdomains (Arita *et al.*, 2004). The C-terminal domain (Asn301–Pro663) forms an α/β-propeller structure and contains the binding site and catalytic centre for arginine citrullination of proteins (Arita *et al.*, 2004, 2006).

Single-nucleotide polymorphisms (SNPs) in PAD4 genes have been identified as being associated with susceptibility to rheumatoid arthritis (Suzuki *et al.*, 2003; Harris *et al.*, 2008), in which the improper hypercitrullination of proteins may cause them to act as autoantigens. In the PAD4 SNP variant (PAD4^{SNP}), the Ser55, Ala82 and Ala112 residues of the conventional PAD4 protein are replaced by Gly55, Val82 and Gly112, respectively. It has been proposed that these nucleotide substitutions in the PAD4^{SNP} mRNA or amino-acid substitutions in the PAD4^{SNP} protein may increase the level of protein citrullination in rheumatoid cells by augmenting the mRNA stability (Suzuki *et al.*, 2003) or enhancing the activity of the protein (Hung *et al.*, 2007), respectively.

In the present study, we purified both PAD4^{SNP} and PAD4 and determined the crystal structures of the proteins at 2.5 and 2.7 Å resolution, respectively. A structural comparison between PAD4^{SNP} and PAD4 revealed that the amino-acid substitutions in PAD4^{SNP} induced conformational changes within the N-terminal domain. However, the active centre for citrullination, which is located in the C-terminal domain, was not significantly affected in PAD4^{SNP}. PAD4^{SNP} consistently possesses similar levels of citrullination activity compared with conventional PAD4. These results support the view that the mRNA stability, but not the citrullination activity, of PAD4 may be responsible for susceptibility to rheumatoid arthritis.

2. Experimental procedures

2.1. Purification of recombinant human PAD4 and PAD4^{SNP}

Human PAD4 and PAD4^{SNP} were overexpressed in *Escherichia coli* cells as N-terminal hexahistidine (His_6)-tagged proteins. The DNA fragments encoding PAD4 and PAD4^{SNP} were ligated into the *Nde*I and *Bam*HI sites of the pET15b vector (Novagen), which harbours a His_6 tag and a PreScission protease-recognition sequence (GE Healthcare Biosciences) at the N-terminus. PAD4 and PAD4^{SNP} were produced in *E. coli* strain BL21 (DE3) cells. After overnight cultivation of *E. coli* cells carrying the PAD4 or PAD4^{SNP} expression vector on LB plates containing 100 µg ml^{-1} ampicillin at 310 K, 20–50 colonies on the LB plates were collected and inoculated into 2.5 l LB medium containing 100 µg ml^{-1} ampicillin and the cultures were incubated at 298 K. When the cell density reached an OD_{600} of 0.4, 0.5 m*M* isopropyl β-D-1-thiogalactopyranoside was added to induce the expression of PAD4 or PAD4^{SNP} and the cultures were incubated at 298 K for a further 15 h. The cells producing PAD4 or PAD4^{SNP} were harvested, resuspended in 50 m*M* sodium phosphate buffer pH 7.5 containing 2 m*M* β-mercaptoethanol, 10% glycerol, 0.5 m*M* EDTA, 0.1% Triton X-100 and 0.5 *M* NaCl and disrupted by sonication. The cell debris was removed by centrifugation (39 191*g*; 15 min) and the lysate was mixed gently with 5 ml (50% slurry) nickel–nitrilotriacetic acid (Ni–NTA) agarose resin (Qiagen) at 277 K for 1 h. The PAD4-bound or PAD4^{SNP}-bound Ni–NTA beads were then packed into an Econo-column (Bio-Rad) and washed with 125 ml 50 m*M* sodium phosphate buffer pH 7.5 containing 10% glycerol, 0.5 m*M* EDTA, 0.1% Triton X-100, 500 m*M* NaCl and 5 m*M* imidazole at a flow rate of about 0.5 ml min^{-1}. His_6-tagged PAD4 or PAD4^{SNP} was eluted with a 40 ml linear gradient of 5–500 m*M* imidazole in 50 m*M* sodium phosphate buffer pH 7.5 containing 10% glycerol, 0.5 m*M* EDTA, 0.1% Triton X-100 and 500 m*M* NaCl. The His_6 tag was removed from the PAD or PAD4^{SNP} portion by PreScission protease (6–10 U per milligram of protein). The sample was immediately dialyzed against 20 m*M* sodium phosphate buffer pH 7.5 containing 200 m*M* NaCl, 0.5 m*M* EDTA, 10% glycerol and 2 m*M* β-mercaptoethanol. After removal of the His_6 tag, the concentration of NaCl in the sample was reduced to 100 m*M* with dilution buffer (20 m*M* sodium phosphate pH 7.5, 0.5 m*M* EDTA, 10% glycerol and 2 m*M* β-mercaptoethanol) and PAD4 or PAD4^{SNP} was then purified using a Mono Q (GE Healthcare Biosciences) column by elution with a 20 ml linear gradient of 100–400 m*M* NaCl in 20 m*M* sodium phosphate buffer pH 7.5 containing 0.5 m*M* EDTA, 10% glycerol and 2 m*M* β-mercaptoethanol. The purified PAD4 and PAD4^{SNP} were concentrated and dialyzed against 10 m*M* Tris–HCl buffer pH 8.5 containing 500 m*M* NaCl, 1 m*M* dithiothreitol and 1 m*M* EDTA.

2.2. Preparation of recombinant human histones

Human H2A, H2B, H3 and H4 were overexpressed in *E. coli* cells as N-terminally His_6-tagged proteins (Tanaka *et al.*, 2004). During protein purification, the His_6 tag was removed from the histones and the H2A–H2B and H3–H4 complexes were prepared using previously described methods (Tachiwana *et al.*, 2008; Osakabe *et al.*, 2010).

2.3. Crystallization, structure determination and refinement

The crystals of human PAD4 and PAD4^{SNP} were grown by the hanging-drop method at 293 K. The hanging drop was formed by adding 1 µl human PAD4 or PAD4^{SNP} (at a concentration of 4 mg ml^{-1}) to 1 µl reservoir solution (0.1 *M* imidazole pH 8.0, 0.2 *M* lithium sulfate and 7% PEG 3350). For data collection, PAD4 and PAD4^{SNP} crystals were harvested in reservoir solution containing 28% ethylene glycol and flash-cooled in a stream of N_2 gas (100 K). Data sets were collected from the crystals on the SPring-8 BL41XU beamline (Harima, Japan). The data sets were processed and scaled using the *HKL*-2000 program suite (Otwinowski & Minor, 1997). The crystals of PAD4 and PAD4^{SNP} belonged to the

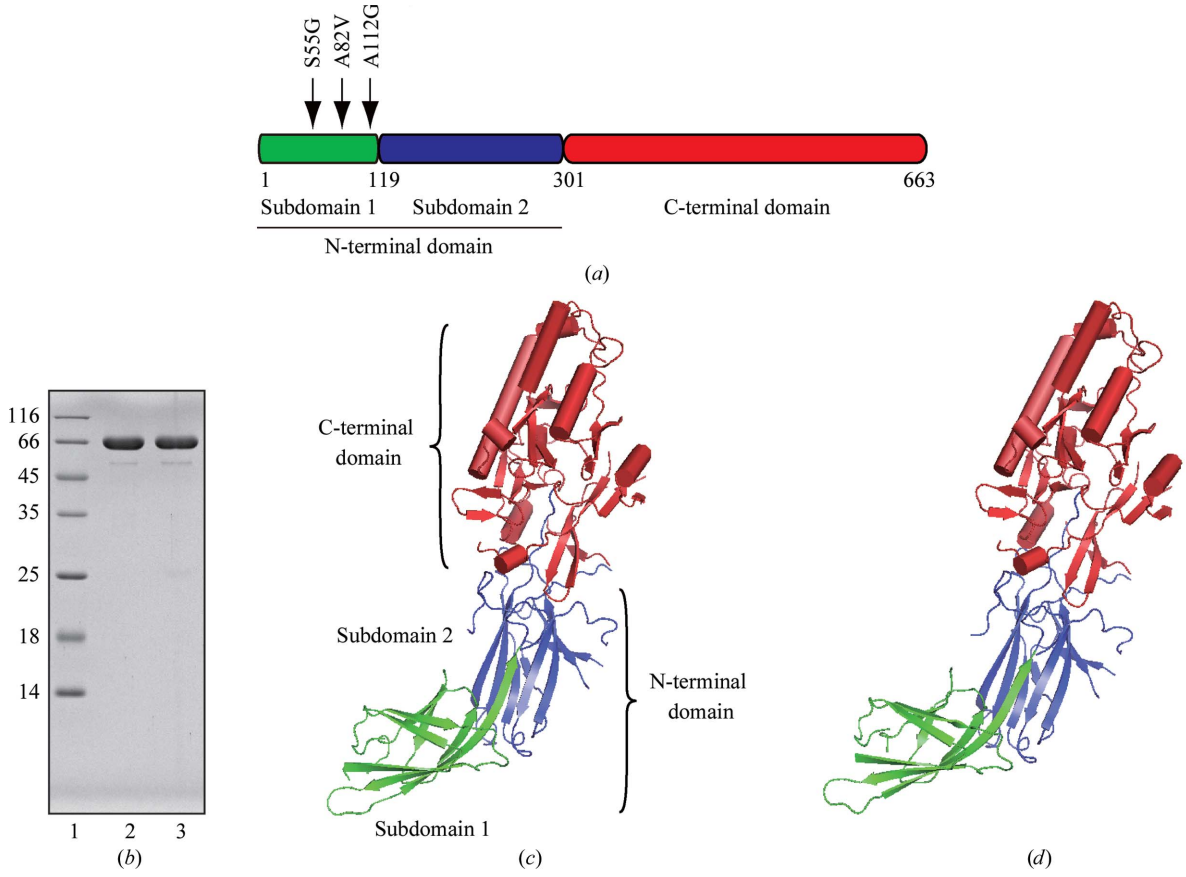

Figure 1
Purification of human PAD4 and PAD4^{SNP}. (*a*) Schematic representation of human PAD4. Subdomains 1 and 2 of the N-terminal domain are coloured green and dark blue, respectively. The C-terminal domain is coloured red. Arrows indicate the locations of the Gly55, Val82 and Gly112 residues of PAD4^{SNP}. (*b*) A 15% SDS–PAGE analysis of purified PAD4 and PAD4^{SNP} (lanes 2 and 3, respectively). These PAD4 and PAD4^{SNP} proteins lack the His_6 tag. Lane 1 contains molecular-mass markers (labelled in kDa). (*c*) Crystal structure of PAD4. (*d*) Crystal structure of PAD4^{SNP}. The colours correspond to the domains shown in (*a*).

monoclinic space group *C*2, with unit-cell parameters $a = 145.25$, $b = 61.06$, $c = 113.57$ Å for PAD4 and $a = 145.56$, $b = 61.24$, $c = 113.47$ Å for PAD4^{SNP}. The data were processed using the *CCP*4 program suite (Collaborative Computational Project 4, Number 4, 1994). The structures of PAD4 and PAD4^{SNP} were determined by molecular replacement using the *Phaser* program (McCoy *et al.*, 2007). The coordinates of the human PAD4 structure (PDB code 1wd8) were used as the search model. The prime-and-switch method in the *RESOLVE* program (Terwilliger, 2003) was used to remove model bias from the electron-density map. The structures of PAD4 and PAD4^{SNP} were subjected to rigid-body, energy-minimization and *B*-factor refinements using the *CNS* program (Brünger *et al.*, 1998) and model building using the *Coot* program (Emsley *et al.*, 2010). The Ramachandran plots of the final structures showed 84.3% of the residues in the most favourable regions for PAD4 and 83.2% of the residues in the most favourable regions for PAD4 ^{SNP}, with no residues in the disallowed region. All structure figures were created using *PyMOL* (DeLano, 2002). The atomic coordinates of the PAD4 and PAD4^{SNP} proteins have been deposited in the Protein Data Bank under accession codes 3apn and 3apm, respectively.

2.4. Citrullination assay for human histones

H2A–H2B (300 ng µl^{-1}) or H3–H4 (300 ng µl^{-1}) were incubated with PAD4 or PAD4^{SNP} at 303 K for 60 min in 10 µl 20 m*M* HEPES–NaOH buffer pH 7.5 containing 2.4 m*M* dithiothreitol, 135 m*M* NaCl, 0.5 m*M* EDTA and 10 m*M* $CaCl_2$. The reactions were terminated by adding SDS–PAGE sample buffer (twofold) and boiling the mixtures for 2 min. The proteins were then fractionated by 15% SDS–PAGE. Bands were visualized by Coomassie Brilliant Blue staining.

2.5. Citrullination assay for N^{α}-benzoyl-L-arginine ethyl ester hydrochloride

N^{α}-Benzoyl-L-arginine ethyl ester hydrochloride (BAEE; 1 m*M*; Knipp & Vasák, 2000) was incubated with PAD4 or PAD4^{SNP} (up to 0.6 µ*M*) at 310 K for 15 min in 60 µl 5 m*M* Tris–HCl buffer pH 8.5 containing 10 m*M* $CaCl_2$, 2.5 m*M* dithiothreitol, 0.5 m*M* EDTA and 250 m*M* NaCl. The reaction was stopped by flash-freezing and was followed by the addition of 200 µl colour-developing reagent, which was freshly prepared by combining solution *A* (80 m*M* 2,3-butanedione monoxime and 2 m*M* thiosemicarbazide) with solution *B* [3 *M* phosphoric acid, 6 *M* sulfuric acid and 2 *M* $NH_4Fe(SO_4)_2$] in a

1:3 ratio. The reaction mixture was heated at 368 K for 15 min and then cooled to 298 K for 10 min. The absorbance was measured at 540 nm on a Powerscan HT spectrophotometer (DS Pharma Biomedical, Osaka, Japan) and the amounts of L-citrulline generated were estimated from a standard curve.

2.6. Circular-dichroism measurements

The circular-dichroism spectra of PAD4 (25 µg ml^{-1}) and PAD4^{SNP} (25 µg ml^{-1}) were measured on a Jasco J-820 spectropolarimeter (Japan Spectroscopic Co. Ltd, Tokyo, Japan) using a 1 cm path-length cell. Experiments were performed in 0.05 m*M* Tris–HCl buffer pH 8.5 containing 2.5 m*M* NaCl, 5 µ*M* dithiothreitol and 5 µ*M* EDTA.

3. Results

3.1. Crystal structure of PAD4^{SNP}

PAD4^{SNP} was found as a protein encoded by a *PADI4* haplotype associated with susceptibility to rheumatoid arthritis (Suzuki *et al.*, 2003; Harris *et al.*, 2008). In PAD4^{SNP}, the Ser55, Ala82 and Ala112 residues of conventional PAD4 are replaced by Gly55, Val82 and Gly112 (Fig. 1*a*). In order to study the effect of these three amino-acid substitutions on the PAD4 structure, we purified both PAD4^{SNP} and PAD4 as bacterially expressed recombinant proteins (Fig. 1*b*). Recombinant PAD4^{SNP} and PAD4 were produced as His$_6$-tagged proteins and the His$_6$ tag was removed by treatment with PreScission protease during the purification procedure. PAD4^{SNP} and PAD4 were then crystallized and the PAD4^{SNP}

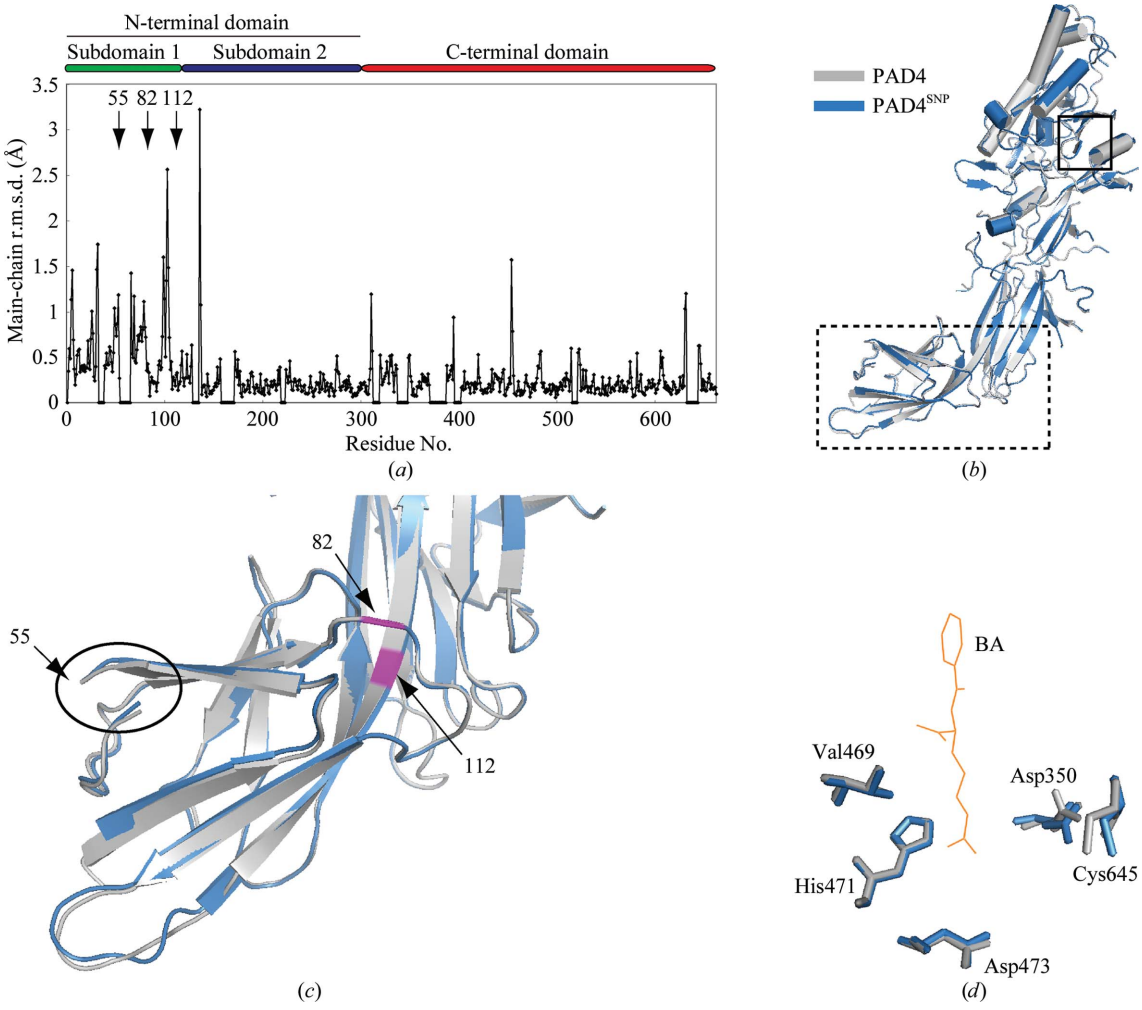

Figure 2
Comparison of the PAD4 and PAD4^{SNP} structures. (*a*) Structural differences between PAD4 and PAD4^{SNP}. The PAD4 and PAD4^{SNP} structures were superimposed and the r.m.s.d. values for each residue pair were calculated and plotted. The domain composition of PAD4 is shown at the top of the panel. Arrows indicate the locations of the PAD4^{SNP}-specific amino-acid residues Gly55, Val82 and Gly112. (*b*) Superimposition of the PAD4 (grey) and PAD4^{SNP} (blue) structures. The solid box indicates the citrullination centre. The dashed box indicates the N-terminal subdomain 1. (*c*) Enlarged view of the N-terminal subdomain 1 indicated by the dashed box in (*b*). The locations of the PAD4^{SNP}-specific amino-acid residues Val82 and Gly112 are coloured purple. The oval indicates the disordered region that contains either Gly55 for PAD4^{SNP} or Ser55 for PAD4. (*d*) Enlarged view of the citrullination centre indicated by the solid box in (*b*). The Asp350, Val469, His471, Asp473 and Cys645 residues, which constitute the catalytic centre for citrullination of PAD4 and PAD4^{SNP}, are superimposed. The benzoyl-L-arginine amide molecule (BA) deduced from the PAD4–BA complex structure (Arita *et al.*, 2004) is coloured orange.

Table 1
Crystallographic statistics.

Values in parentheses are for the highest resolution shell.

	PAD4	PAD4^{SNP}
Data collection		
Space group	*C*2	*C*2
Unit-cell parameters (Å, °)	*a* = 145.25, *b* = 61.06, *c* = 113.57, $\alpha = \gamma = 90$, $\beta = 123.74$	*a* = 145.56, *b* = 61.24, *c* = 113.47, $\alpha = \gamma = 90$, $\beta = 123.85$
Resolution (Å)	50–2.7 (2.75–2.70)	50–2.5 (2.59–2.50)
Reflections (measured/unique)	229479/22962	240257/28960
R_{merge}† (%)	6.5 (75.1)	6.5 (64.1)
Mean $\langle I/\sigma(I)\rangle$	13.9 (2.3)	12.9 (3.0)
Completeness (%)	99.7 (100)	99.5 (96.5)
Multiplicity	3.8 (3.8)	3.8 (3.6)
Overall *B* factor from Wilson plot (Å^2)	82	65
Refinement		
Resolution (Å)	30–2.7	30–2.5
R factor/free *R* factor‡ (%)	24.7/29.3	26.0/28.6
R.m.s. deviations		
Bond lengths (Å)	0.009	0.009
Bond angles (°)	1.54	1.47
Ramachandran plot, residues in		
Most favourable regions (%)	84.3	83.2
Additionally allowed regions (%)	15.1	16.1
Generously allowed regions (%)	0.6	0.6

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of an observation and $\langle I(hkl)\rangle$ is the mean value for that reflection. ‡ *R* factor = $\sum_{hkl} \big||F_{obs}| - |F_{calc}|\big| / \sum_{hkl} |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes, respectively. The free *R* factor was calculated using 5% of the data that were excluded from the refinement.

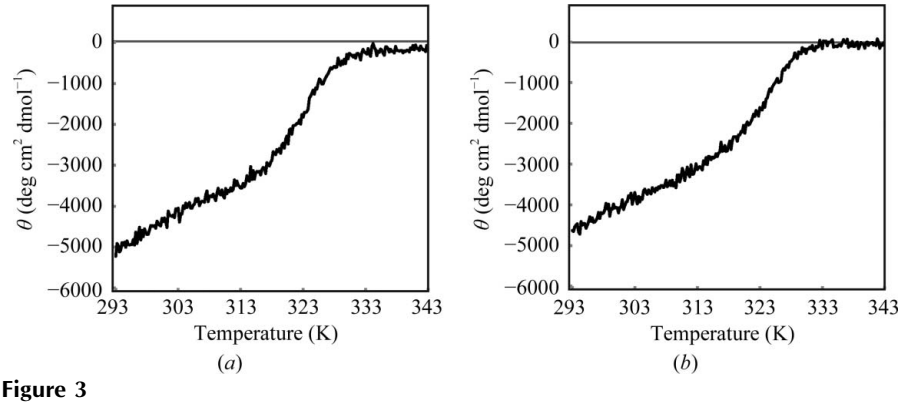

Figure 3
The CD effect at 230 nm as a function of temperature. (*a*) The CD spectrum of conventional PAD4. (*b*) The CD spectrum of PAD4^{SNP}.

and PAD4 structures were solved to 2.5 and 2.7 Å resolution, respectively, by the molecular-replacement method using the previously reported PAD4 structure (PDB code 1wd8) as a search model (Arita *et al.*, 2004, 2006; Table 1). The overall structure of PAD4^{SNP} was essentially similar to that of PAD4 (Figs. 1*c* and 1*d*).

3.2. Structural comparison between PAD4^{SNP} and PAD4

The PAD4^{SNP} and PAD4 structures were superimposed and the r.m.s.d. (root-mean-square deviation) values for each residue pair were calculated and plotted against each other (Fig. 2*a*). The PAD4 structure was divided into N-terminal and C-terminal domains and the N-terminal domain was further divided into subdomains 1 and 2 (Fig. 1*c*). The PAD4^{SNP}-specific Gly55, Val82 and Gly112 residues are located in the N-terminal subdomain 1 (Figs. 2*b* and 2*c*). The Gly55 residue is located in a disordered region of the PAD4^{SNP} structure and thus was not visible in the crystal structure (Fig. 2*c*). The Val82 and Gly112 residues are located near the interface of the N-terminal subdomains 1 and 2 (Figs. 2*b* and 2*c*). Most of the structural differences were found in the N-terminal region, particularly in the N-terminal subdomain 1, in which the three amino-acid substitutions are located (Fig. 2*a*). In contrast, the structural differences in the C-terminal region, which contains the active centre for citrullination, were minimal (Fig. 2*a*). A close-up view of the superimpositions of the Asp350, Val469, His471, Asp473 and Cys645 residues, which constitute the catalytic centre for citrullination (Arita *et al.*, 2004), revealed that these residues overlapped well with those of conventional PAD4 (Fig. 2*d*). Therefore, the structural perturbations introduced by the amino-acid substitutions in PAD4^{SNP} do not allosterically affect the citrullination centre in the distinct C-terminal domain.

3.3. Thermal stability of PAD4^{SNP}

We next tested the stability of PAD4^{SNP}. To do so, we measured the circular-dichroism (CD) spectra of PAD4 and PAD4^{SNP}. CD spectra were collected for PAD4 and PAD4^{SNP} at 230 nm while the temperature was gradually increased from 293 to 343 K (Figs. 3*a* and 3*b*). The conventional PAD4 showed a sharp increase in ellipticity around 323 K, indicating that a transition from the native state to the denatured state occurred around this temperature (Fig. 3*a*). A similar thermal denaturation profile was observed with PAD4^{SNP} (Fig. 3*b*). Therefore, the Gly55, Val82 and Gly112 residues of PAD4^{SNP} do not affect the stability of the protein.

3.4. Citrullination activity of PAD4^{SNP}

We then compared the citrullination activity of PAD4^{SNP} with that of conventional PAD4. As shown in Fig. 4(*a*), PAD4^{SNP} efficiently deiminated N^{α}-benzoyl-L-arginine ethyl ester hydrochloride (BAEE) to generate citrulline and its activity was almost the same as that of conventional PAD4. The citrullination activity of PAD4^{SNP} was also tested with core histones, which are known substrates of PAD4 (Cuthbert *et al.*, 2004; Wang *et al.*, 2004; Denis *et al.*, 2009; Shimoyama *et*

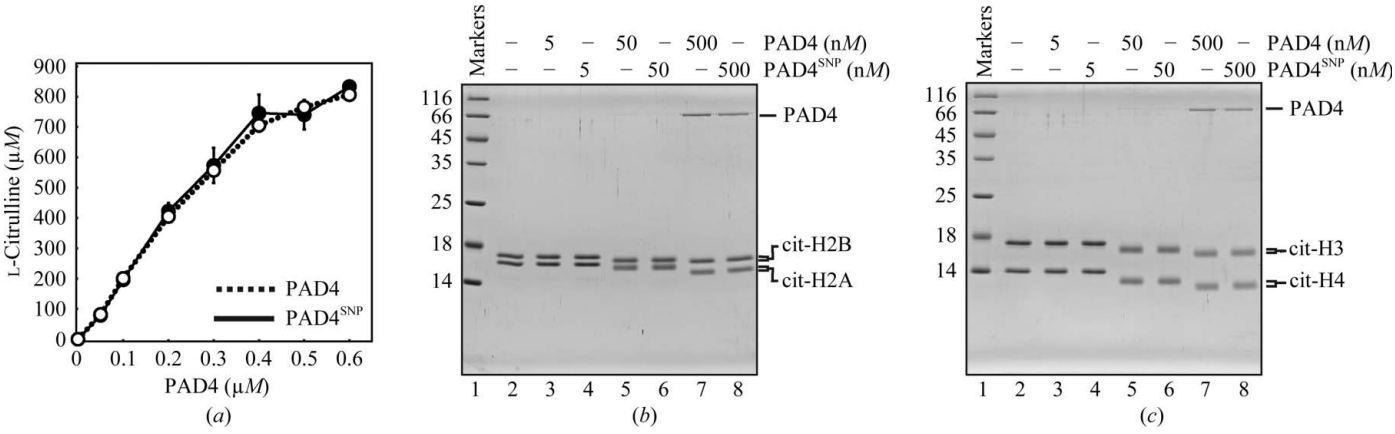

Figure 4
The citrullination activity of PAD4^{SNP}. (*a*) PAD4 or PAD4^{SNP} was incubated with N^{α}-benzoyl-L-arginine ethyl ester hydrochloride (BAEE) in the presence of Ca^{2+}. After 15 min incubation, the resulting ureido group in the modified BAEE was assayed and was plotted against the protein concentration. The solid line with closed symbols indicates the experiments with PAD4^{SNP}. The dotted line with open symbols indicates the experiments with PAD4. The averages of three independent experiments are represented by the standard deviation values. (*b*) H2A–H2B (300 ng μl^{-1}) was incubated with PAD4 (5, 50 or 500 n*M*) or PAD4^{SNP} (5, 50 or 500 n*M*) at 303 K for 60 min. The reactions were terminated by adding SDS–PAGE sample buffer (twofold) followed by boiling the mixtures for 2 min. The proteins were then fractionated by 15% SDS–PAGE. Bands were visualized by Coomassie Brilliant Blue staining. The bands that correspond to citrullinated H2A and H2B are represented as cit-H2A and cit-H2B. (*c*) H3–H4 (300 ng μl^{-1}) was incubated with PAD4 or PAD4^{SNP} at 303 K for 60 min. The reactions were terminated by adding SDS–PAGE sample buffer (twofold) followed by boiling the mixtures for 2 min. The proteins were then fractionated by 15% SDS–PAGE. The bands were visualized by Coomassie Brilliant Blue staining. The bands that correspond to citrullinated H3 and H4 are represented as cit-H3 and cit-H4.

al., 2010). To do so, we purified human histones as the H2A–H2B and H3–H4 complexes (Tanaka *et al.*, 2004). The histone H2A–H2B or H3–H4 complex was then incubated with the indicated amounts of PAD4 or PAD4^{SNP} and the resulting citrullinated histones were separated by 15% SDS–PAGE. PAD4 promoted the citrullination of all core histones, H2A, H2B, H3 and H4, and the citrullinated histones migrated faster than the unmodified histones in the gel (Figs. 4*b* and 4*c*; lanes 3, 5 and 8). As shown in Figs. 4(*b*) and 4(*c*), no substantial difference in histone citrullination activity was observed for PAD4^{SNP} and PAD4. We therefore concluded that the amino-acid substitutions found in PAD4^{SNP} do not cause hypercitrullination in rheumatoid cells.

4. Discussion

PAD4^{SNP} was found as a protein encoded by a *PADI4* haplotype associated with susceptibility to rheumatoid arthritis and may improperly produce citrullinated proteins that act as autoantigens (Suzuki *et al.*, 2003; Harris *et al.*, 2008). Two possible mechanisms for inducing improper protein citrullination by PAD4^{SNP} have been proposed: augmented stability of the PAD4^{SNP} mRNA and enhanced citrullination activity of the PAD4^{SNP} protein (Suzuki *et al.*, 2003; Hung *et al.*, 2007). We note that these two possibilities are not mutually exclusive.

In the present study, we determined the crystal structures of the PAD4^{SNP} and PAD4 proteins. PAD4 is composed of distinct N-terminal and C-terminal domains. The three amino-acid residues that differ between PAD4^{SNP} (Gly55, Val82 and Gly112) and conventional PAD4 (Ser55, Ala82 and Ala112) are located in the N-terminal domain. We hypothesized that the enhanced citrullination observed in rheumatoid arthritis patients is the result of hypercitrullination activity by PAD4^{SNP}. If this is true, then the amino-acid substitutions in PAD4^{SNP} may induce conformational changes in the active centre for citrullination located in the C-terminal domain, endowing PAD4^{SNP} with a higher citrullination activity than conventional PAD4. However, our structural comparison revealed that structural deviations were only found in the N-terminal domain and not in the C-terminal domain. Consistently, we did not detect any substantial differences in citrullination activity between PAD4^{SNP} and PAD4 using N^{α}-benzoyl-L-arginine ethyl ester hydrochloride and core histones as substrates. These results strongly suggested that the amino-acid substitutions in PAD4^{SNP} do not affect the structure of the citrullination centre and the citrullination activity of PAD4^{SNP}. On the other hand, the N-terminal domain of PAD4 contains a classical monopartite nuclear localization signal (NLS; 56-PPAKKKST-63; Nakashima *et al.*, 2002) and the NLS is located on the solvent-accessible surface in a loop region between β5 and β6 in subdomain 1 (Arita *et al.*, 2004). Therefore, the amino-acid substitutions found in PAD4^{SNP} may affect its transportation to the nucleus.

SNPs may be associated with mRNA stability in scleroderma and melanoma cells (Zhou *et al.*, 2002; Yang *et al.*, 2003). The susceptible-haplotype mRNA of PAD4^{SNP} has also been reported to be stabilized as compared with that of conventional PAD4 (Suzuki *et al.*, 2003). Our structural and biochemical analyses indicated that the structure of the citrullination centre and the citrullination activity of PAD4^{SNP} are indistinguishable from those of conventional PAD4. Therefore, the improper protein hypercitrullination by PAD4^{SNP} in rheumatoid arthritis patients is not a consequence

of enhanced activity of PAD4^{SNP}. Larger amounts of the susceptible-haplotype mRNA may accumulate owing to its higher stability and may result in excess amounts of PAD4^{SNP}.

We thank the beamline scientists Drs N. Shimizu, Y. Kawano and M. Makino for their assistance with data collection on the BL41XU beamline of SPring-8. This work was supported in part by Grants-in-Aid from the Japanese Society for the Promotion of Science (JSPS) and the Ministry of Education, Culture, Sports, Science and Technology, Japan. HK is a Research Fellow in the Waseda Research Institute of Science and Engineering.

References

Arita, K., Hashimoto, H., Shimizu, T., Nakashima, K., Yamada, M. & Sato, M. (2004). *Nature Struct. Mol. Biol.* **11**, 777–783.
Arita, K., Shimizu, T., Hashimoto, H., Hidaka, Y., Yamada, M. & Sato, M. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 5291–5296.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Chavanas, S., Méchin, M. C., Takahara, H., Kawada, A., Nachat, R., Serre, G. & Simon, M. (2004). *Gene*, **330**, 19–27.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cuthbert, G. L., Daujat, S., Snowden, A. W., Erdjument-Bromage, H., Hagiwara, T., Yamada, M., Schneider, R., Gregory, P. D., Tempst, P., Bannister, A. J. & Kouzarides, T. (2004). *Cell*, **118**, 545–553.
DeLano, W. L. (2002). *PyMOL*. http://www.pymol.org.
Denis, H., Deplus, R., Putmans, P., Yamada, M., Métivier, R. & Fuks, F. (2009). *Mol. Cell. Biol.* **29**, 4982–4993.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Guerrin, M., Ishigami, A., Méchin, M. C., Nachat, R., Valmary, S., Sebbag, M., Simon, M., Senshu, T. & Serre, G. (2003). *Biochem. J.* **370**, 167–174.
Harris, M. L., Darrah, E., Lam, G. K., Bartlett, S. J., Giles, J. T., Grant, A. V., Gao, P., Scott, W. W. Jr, El-Gabalawy, H., Casciola-Rosen, L., Barnes, K. C., Bathon, J. M. & Rosen, A. (2008). *Arthritis Rheum.* **58**, 1958–1967.
Hidaka, Y., Hagiwara, T. & Yamada, M. (2005). *FEBS Lett.* **579**, 4088–4092.
Hung, H.-C., Lin, C.-Y., Liao, Y.-F., Hsu, P.-C., Tsay, G. J. & Liu, G.-Y. (2007). *Apoptosis*, **12**, 475–487.
Ishigami, A., Ohsawa, T., Asaga, H., Akiyama, K., Kuramoto, M. & Maruyama, N. (2002). *Arch. Biochem. Biophys.* **407**, 25–31.
Kanno, T., Kawada, A., Yamanouchi, J., Yosida-Noro, C., Yoshiki, A., Shiraiwa, M., Kusakabe, M., Manabe, M., Tezuka, T. & Takahara, H. (2000). *J. Invest. Dermatol.* **115**, 813–823.
Kearney, P. L., Bhatia, M., Jones, N. G., Luo, Y., Glascock, M. C., Catchings, K. L., Yamada, M. & Thompson, P. R. (2005). *Anal. Biochem.* **286**, 257–264.
Knipp, M. & Vasák, M. (2000). *Anal. Biochem.* **286**, 257–264.
McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
Nakashima, K., Hagiwara, T., Ishigami, A., Nagata, S., Asaga, H., Kuramoto, M., Senshu, T. & Yamada, M. (1999). *J. Biol. Chem.* **274**, 27786–27792.
Nakashima, K., Hagiwara, T. & Yamada, M. (2002). *J. Biol. Chem.* **277**, 49562–49568.
Osakabe, A., Tachiwana, H., Matsunaga, T., Shiga, T., Nozawa, R., Obuse, C. & Kurumizaka, H. (2010). *J. Biol. Chem.* **285**, 11913–11921.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Raijmakers, R., Zendman, A. J., Egberts, W. V., Vossenaar, E. R., Raats, J., Soede-Huijbregts, C., van Rutjes, F. P., Veelen, P. A., Drijfhout, J. W. & Pruijn, G. J. (2007). *J. Mol. Biol.* **367**, 1118–1129.
Shimoyama, S., Nagadoi, A., Tachiwana, H., Yamada, M., Sato, M., Kurumizaka, H., Nishimura, Y. & Akashi, S. (2010). *J. Mass Spectrom.* **45**, 900–908.
Suzuki, A. *et al.* (2003). *Nature Genet.* **34**, 395–402.
Tachiwana, H., Osakabe, A., Kimura, H. & Kurumizaka, H. (2008). *Nucleic Acids Res.* **36**, 2208–2218.
Tanaka, Y., Tawaramoto-Sasanuma, M., Kawaguchi, S., Ohta, T., Yoda, K., Kurumizaka, H. & Yokoyama, S. (2004). *Methods*, **33**, 3–11.
Terwilliger, T. C. (2003). *Methods Enzymol.* **374**, 22–37.
Thompson, P. R. & Fast, W. (2006). *ACS Chem. Biol.* **1**, 433–441.
Vossenaar, E. R., Zendman, A. J. W., Venrooij, W. J. & Pruijn, G. J. M. (2003). *Bioessays*, **25**, 1106–1118.
Wang, Y., Li, M., Stadler, S., Correll, S., Li, P., Wang, D., Hayama, R., Leonelli, L., Han, H., Grigoryev, S. A., Allis, C. D. & Coonrod, S. A. (2009). *J. Cell Biol.* **184**, 205–213.
Wang, Y., Wysocka, J., Sayegh, J., Lee, Y.-H., Perlin, J. R., Leonelli, L., Sonbuchner, L. S., McDonald, C. H., Cook, R. G., Dou, Y., Roeder, R. G., Clarke, S., Stallcup, M. R., Allis, C. D. & Coonrod, S. A. (2004). *Science*, **306**, 279–283.
Yang, T., McNally, B. A., Ferrone, S., Liu, Y. & Zheng, P. (2003). *J. Biol. Chem.* **278**, 15291–15296.
Zhou, X., Tan, F. K., Reveille, J. D., Wallis, D., Milewicz, D. M., Ahn, C., Wang, A. & Arnett, F. C. (2002). *Arthritis Rheum.* **46**, 2990–2999.