Structural characterization of human peptidyl-arginine deiminase type III by X-ray crystallography

Othman Rechiche,[a]* T. Verne Lee[b] and J. Shaun Lott[b]*

[a]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, Althouse Laboratory, Science Drive, State College, PA 16801, USA, and [b]School of Biological Sciences, The University of Auckland, 3a Symonds Street, Auckland 1142, New Zealand. *Correspondence e-mail: othman.rechiche@psu.edu, s.lott@auckland.ac.nz

The Ca^{2+}-dependent enzyme peptidyl-arginine deiminase type III (PAD3) catalyses the deimination of arginine residues to form citrulline residues in proteins such as keratin, filaggrin and trichohyalin. This is an important post-translation modification that is required for normal hair and skin formation in follicles and keratocytes. The structure of apo human PAD3 was determined by X-ray crystallography to a resolution of 2.8 Å. The structure of PAD3 revealed a similar overall architecture to other PAD isoforms: the N-terminal and middle domains of PAD3 show sequence and structural variety, whereas the sequence and structure of the C-terminal catalytic domain is highly conserved. Structural analysis indicates that PAD3 is a dimer in solution, as is also the case for the PAD2 and PAD4 isoforms but not the PAD1 isoform.

1. Introduction

Deimination (or citrullination) is a post-translational modification catalysed by peptidyl-arginine deiminases (PADs), the consequences of which are not yet fully understood (Rogers *et al.*, 1977; Kubilus *et al.*, 1979, 1980; Kubilus & Baden, 1983; Rothnagel & Rogers, 1984; Mondal & Thompson, 2019). Five paralogous human genes encoding highly conserved PAD enzymes (numbered PAD1–4 and PAD6) have been identified (Dong *et al.*, 2006; Chavanas *et al.*, 2008). These genes are clustered in a single chromosomal locus at 1p35–36 in the human genome. Peptidyl-arginine deiminase activity has been detected in most organs, tissues and cells of all vertebrates. The PAD isoforms are controlled at the transcriptional, translational and activity levels, and they each have particular substrate specificities (Kizawa *et al.*, 2008; Knuckley *et al.*, 2010; Assohou-Luty *et al.*, 2014).

PADs deiminate arginine residues in proteins, converting them into citrulline residues, in a Ca^{2+}-dependent manner (Shirai *et al.*, 2001; Das *et al.*, 2004; Knuckley *et al.*, 2007). Although the exact physiological functions of protein deimination have not been clarified, the modification of the positively charged guanidino group of arginine to the neutral ureido group of citrulline dramatically alters the charge of the targeted substrates, which decreases the electrostatic interactions (both salt bridge and cation–π) between arginine and other amino-acid residues, impairing the formation of functional tertiary structures (Vossenaar *et al.*, 2003).

The human PAD isoforms are highly conserved, with between 45% and 58% pairwise amino-acid sequence identity between them. The structures of the PAD1, PAD2 and PAD4 proteins have been determined, and they share a common architecture (Arita *et al.*, 2004; Slade *et al.*, 2015; Saijo *et al.*, 2016). Their structure is composed of three domains: an
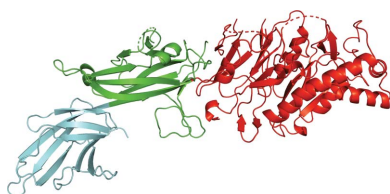
N-terminal cupredoxin-like domain (also referred to as an IgG1 domain), a central IgG-like domain and a C-terminal catalytic domain. The N-terminal and central domains have a lower shared sequence identity than the catalytic C-terminal domain.

PAD3 is specifically expressed in skin tissue and hair follicles (Steinert *et al.*, 2003; Méchin *et al.*, 2005; Kizawa *et al.*, 2008) and several *in vivo* substrates of the enzyme have been identified, in particular structural proteins: intermediate filament (IF) proteins such as keratins and IF-associated proteins such as filaggrin (Senshu *et al.*, 1996), trichohyalin (Tarcsa *et al.*, 1996) and S100A3 (a cysteine-rich Ca^{2+}-binding protein present in hair cuticles; Kizawa *et al.*, 2008). Generally, they are abundant proteins and have a high percentage of arginine in their primary sequence: trichohyalin contains 23% arginine and filaggrin contains 11% arginine. *In vitro*, PAD3 specifically deiminates a single arginine residue (Arg51) in S100A3, but the origin of this target specificity is unclear. Defects in PAD3 activity have been linked to hair and skin diseases such as uncombable hair syndrome (Basmanav *et al.*, 2016), psoriasis (Ishida-Yamamoto *et al.*, 2000), bullous congenital ichthyosiform erythroderma (Chavanas *et al.*, 2006) and central centrifugal cicatricial alopecia (Malki *et al.*, 2019).

Understanding the detailed molecular role of PAD3 in skin and hair physiology has been hindered by the lack of structural information about this isoform and its specific recognition of protein substrates. Crystals of PAD3 that diffracted to 2.95 Å resolution have previously been reported by Unno *et al.* (2012). However, a structure of this protein has not previously been published or submitted to the Protein Data Bank. As an initial step to understand the mechanism of specific protein deimination in the hair follicle and the skin, we here describe the first crystal structure of PAD3.

2. Materials and methods

2.1. Macromolecule production

The synthetic gene coding for human PAD3 (UniProt Q9ULW8) was cloned into the pDEST17 vector (Gateway system, Invitrogen, USA). The plasmid was transformed into *Escherichia coli* BL21(DE3)pLysS cells for expression (Table 1). The cells were grown at 310 K in Luria–Bertani broth containing 50 mg ml^{-1} ampicillin and 35 mg ml^{-1} chloramphenicol until mid-log phase (OD_{600} = 0.5–0.7). Protein expression was induced by the addition of 0.25 mM isopropyl β-D-1-thiogalactopyranoside, and the cells were then grown at 291 K. The cells were harvested by centrifugation (3500g for 20 min) and resuspended in lysis buffer (50 mM HEPES pH 7.8, 500 mM NaCl, 10 mM imidazole) before being lysed using a cell disruptor. Cell debris was spun out of the lysate at 40 000g for 40 min at 277 K. His$_6$-tagged PAD3 was purified from the soluble lysate fraction by Ni^{2+}-affinity chromatography using a 5 ml HisTrap HP column (GE Healthcare). To remove any nonspecific binding, the column was washed with (i) ten column volumes (CV) of binding buffer (50 mM HEPES pH 7.8, 500 mM NaCl, 10 mM

Table 1
Macromolecule-production information.

Source organism	*Homo sapiens*
Expression vector	pDEST17
Expression host	*E. coli* BL21(DE3)pLysS
Complete amino-acid sequence of the construct produced	MSYYHHHHHHLESTSLYKKAGFMSLQRIVR VSLEHPTSAVCVAGVETLVDIYGSVPEG TEMFEVYGTPGVDIYISPNMERGRERAD TRRWRFDATLEIIVVMNSPSNDLNDSHV QISYHSSHEPLPLAYAVLYLTCVDISLD CDLNCEGRQDRNFVDKRQWVWGPSGYGG ILLVNCDRDDPSCDVQDNCDQHVHCLQD LEDMSVMVLRTQGPAALFDDHKLVLHTS SYDAKRAQVFHICGPEDVCEAYRHVLGQ DKVSYEVPRLHGDEERRFVEGLSFPDAG FTGLISFHVTLLDDSNEDFSASPIFTDT VVFRVAPWIMTPSTLPPLEVYVCRVRNN TCFVDAVAELARKAGCKLTICPQAENRN DRWIQDEMELGYVQAPHKTLPVVFDSPR NGELQDFPYKRILGPDFGYVTREPRDRS VSGLDSFGNLEVSPPVVANGKEYPLGRI LIGGNLPGSSGRRVTQVVRDFLHAQKVQ PPVELFVDWLAVGHVDEFLSFVPAPDGK GFRMLLASPGACFKLFQEKQKCGHGRAL LFQGVVDDEQVKTISINQVLSNKDLINY NKFVQSCIDWNREVLKRELGLAECDIID IPQLFKTERKKATAFFPDLVNMLVLGKH LGIPKPFGGPIINGCCCLEEKVRSLLEPL GLHCTFIDDDFTPYHMLHGEVHCGTNVCR KPFSFKWWNMVP

β-mercaptoethanol, 5% glycerol, 10 mM imidazole), (ii) 5 CV wash buffer 1 (50 mM HEPES pH 7.8, 1 M NaCl, 10 mM β-mercaptoethanol, 20 mM imidazole, 5% glycerol), (iii) 5 CV wash buffer 2 (50 mM HEPES pH 7.8, 1 M NaCl, 10 mM β-mercaptoethanol, 20 mM imidazole, 5% glycerol, 10 mM ATP, 20 mM MgCl$_2$, 20 mM KCl) and (iv) 5 CV wash buffer 3 (50 mM HEPES pH 7.8, 1 M NaCl, 10 mM β-mercaptoethanol, 20 mM imidazole, 10 mM ATP, 20 mM MgCl$_2$, 150 mM KCl).

The protein was eluted with a 100 ml gradient from 20 mM to 1 M imidazole in 50 mM HEPES pH 7.8, 500 mM NaCl, 10 mM β-mercaptoethanol, 5% glycerol. Fractions containing His$_6$-tagged PAD3 were pooled, dialyzed into ion-exchange chromatography (IEX) binding buffer (50 mM Tris pH 7.8, 50 mM NaCl, 2.5% glycerol, 2 mM DTT, 1 mM EDTA) and loaded onto a Source 30Q column (GE Healthcare) pre-equilibrated with IEX binding buffer for the anion-exchange chromatography purification step. The column was washed with 10 CV IEX binding buffer. The protein was then eluted in 100 ml IEX binding buffer with a gradient from 0 to 1 M NaCl. Fractions containing His$_6$-tagged PAD3 were pooled and dialysed into size-exclusion chromatography buffer (20 mM Tris pH 7.8, 100 mM NaCl, 2 mM DTT, 1 mM EDTA). The protein was concentrated, spun at 20 000g for 10 min and filtered before loading onto an S200 16/600 size-exclusion column (GE Healthcare) for the final size-exclusion chromatography purification step. Fractions containing His$_6$-tagged PAD3 were pooled and concentrated to 10 mg ml^{-1} for crystallization. Protein purity was verified using SDS–PAGE. Our expression and purification methods are similar to those previously reported by Unno *et al.* (2012), with differences in the choice of expression vector and the strain of *E. coli* used.

Table 2
Crystallization.

Method	Vapour diffusion, sitting drop
Plate type	Intelli-Plates (Hampton Research)
Temperature (K)	291
Protein concentration (mg ml^{-1})	10
Buffer composition of protein solution	20 mM Tris pH 7.8, 100 mM NaCl, 2 mM DTT, 1 mM EDTA
Composition of reservoir solution	Na HEPES + MOPS (acid), ethylene glycols, PEG 8000
Volume and ratio of drop	0.5 μl (1:1 protein:reservoir), 0.45 μl (2:1 protein:reservoir)
Volume of reservoir (μl)	50

2.2. Crystallization

All crystallization trials were carried out using the sitting-drop vapour-diffusion method in 96-well Intelli-Plates (Hampton Research) containing 50 μl reservoir solution. For each condition, two crystallization drops, one with 0.25 μl protein solution plus 0.25 μl reservoir solution and the other with 0.3 μl protein solution plus 0.15 μl reservoir solution, were set up. Crystallization trays were maintained at 291 K. Precipitant screening was initially carried out using the MORPHEUS screen (Gorrec, 2009) as well as the JCSG-*plus* and PACT *premier* screens (Newman *et al.*, 2005). Crystals of PAD3 were grown in condition E6 of the MORPHEUS screen (Gorrec, 2009), which comprises 0.12 M ethylene glycols, 0.1 M Buffer System 2 [1.0 M sodium HEPES; MOPS (acid); pH 7.5] and 50%(v/v) Precipitant Mix 2 [40%(v/v) ethylene glycol, 20%(w/v) PEG 8000] (Table 2). As a comparison, Unno and coworkers previously reported the crystallization of PAD3 at a concentration of 2–8 mg ml^{-1} in 0.1 M HEPES buffer pH 7.5, 0.2 M NaCl, 25%(v/v) PEG 400 (Unno *et al.*, 2012).

2.3. Data collection and processing

Individual crystals were harvested directly from the MORPHEUS screen (Gorrec, 2009) and flash-cooled at 110 K. X-ray diffraction data were collected from native crystals on beamline MX2 at the Australian Synchrotron at a wavelength of 0.979 Å. The crystals diffracted to a maximum resolution of 2.8 Å. Data were collected with an oscillation angle of 0.1° and an exposure time of 0.1 s per frame, with a total rotation range of 360° (Table 3). X-ray diffraction data were integrated using the *XDS* suite (Kabsch, 2010) and data reduction was performed using *AIMLESS* (Evans & Murshudov, 2013).

2.4. Structure solution and refinement

The structure was determined by molecular replacement using *Phaser* in the *Phenix* package (version 1.13_2998; Liebschner *et al.*, 2019). Human PAD4 (PDB entry 2dew; Arita *et al.*, 2006) was used as the search model. The crystals were indexed in space group *H*32, contained one monomer of protein in the asymmetric unit and seem to be related to the crystal form previously reported by Unno *et al.* (2012), which had similar cell lengths and was indexed in space group *H*3 with two protein monomers in the asymmetric unit. PAD3 is

Table 3
Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	Beamline MX2, Australian Synchrotron
Wavelength (Å)	0.979
Temperature (K)	100
Detector	Dectris EIGER X 16M
Space group	*H*32
a, *b*, *c* (Å)	115.02, 115.02, 328.49
α, *β*, *γ* (°)	90, 90, 120
Resolution range (Å)	49.24–2.80 (2.95–2.80)
Total No. of reflections	390648 (53222)
No. of unique reflections	21054 (3035)
Completeness (%)	100 (100)
Multiplicity	18.6 (17.5)
$\langle I/\sigma(I)\rangle$	8.0 (0.8)†
$CC_{1/2}$	0.993 (0.334)
R_{meas}	0.575 (9.11)
$R_{p.i.m.}$	0.132 (2.16)
Overall B factor from Wilson plot (Å^2)	66.14

† The resolution cutoff was chosen based on the point at which the $CC_{1/2}$ value dropped to 0.3, resulting in $\langle I/\sigma(I)\rangle$ being <2.0 in the outer shell of data.

Table 4
Structure refinement.

Values in parentheses are for the outer shell.

Resolution range (Å)	49.24–2.80 (2.95–2.80)
Completeness (%)	99.8
No. of reflections, working set	19978 (2769)
No. of reflections, test set	1032 (155)
Final R_{cryst}	0.234 (0.354)
Final R_{free}	0.274 (0.336)
No. of non-H atoms	
Protein	4253
Ligand	0
Water	0
Total	4253
R.m.s. deviations	
Bonds (Å)	0.004
Angles (°)	0.076
Average B factors (Å^2)	
Protein	61.9
Ramachandran plot	
Most favoured (%)	93.4
Allowed (%)	100

664 amino acids in length, but the final model contained only 595 residues, as several surface loops could not be modelled due to weak electron density. The model was refined to a final R_{work} and R_{free} of 0.234 and 0.274, respectively, with good geometry (Table 4). The structure was built and refined with iterative cycles of model building in *Coot* (Emsley *et al.*, 2010) and refinement using *phenix.refine* (version 1.19.2-4158; Afonine *et al.*, 2012; Liebschner *et al.*, 2019). The final rounds of maximum-likelihood refinement were carried out using TLS using TLS groups determined by the *TLSMD* server (version 1.4.0; Painter & Merritt, 2006a,b). Model validation was performed using *MolProbity* (Williams *et al.*, 2018).

3. Results and discussion

The PAD3 monomer adopts an overall elongated and curved shape, approximately 130 × 45 × 45 Å in size (Fig. 1), and like

Table 5
Dimer interfaces of the PAD enzyme structures.

	PAD3	PAD4, apo	PAD2, apo	PAD4, Ca^{2+}	PAD2, Ca^{2+}	PAD1, Ca^{2+}
PDB code	6ce1	1wd8	4n20	1wd9	4n2b	5hp5
No. of residues						
Dimer interface	78 (13.1%)	77 (13.8%)	81 (12.7%)	72 (11.9%)	78 (11.9%)	25 (3.7%)
Surface	570 (96%)	507 (90.7%)	587 (91.9%)	541 (89.7%)	599 (91.7%)	613 (91.6%)
Total	594 (100%)	559 (100%)	639 (100%)	603 (100%)	653 (100%)	669 (100%)
Solvent-accessible area (Å^2)						
Dimer interface	2388 (8.7%)	2314 (8.5%)	2552 (8.8%)	2106 (7.8%)	2480 (8.5%)	1159 (3.9%)
Total	27582 (100%)	27227 (100%)	28985 (100%)	26948 (100%)	29152 (100%)	29656 (100%)
Complex-formation significance score	1	1	0.189	1	0.135	0

other PAD isoforms consists of three domains: an N-terminal domain (residues 1–113; shown in cyan), a central domain (residues 115–273; shown in green) and a C-terminal catalytic domain (residues 293–664; shown in red). The N-terminal domain has a cupredoxin-like fold: a β-sandwich composed of eight distorted parallel and antiparallel β-strands ($\beta1$–$\beta8$) connected by seven loops. The central domain has an immunoglobin-like (IgG) fold (Bruschi *et al.*, 2017) and is composed of nine β-strands ($\beta9$–$\beta17$), arranged in two sheets, and two small α-helices. One of the helices connects $\beta10$ to $\beta11$ in a loop–helix–loop motif and the other links $\beta11$ to $\beta12$. The C-terminal domain is composed of five $\beta\beta\alpha\beta$ modules that are arranged circularly in a pseudo-fivefold-symmetric α/β propeller. Together, the first β-strand of each individual module forms the active-site cleft that binds the substrate at the core of the domain in other isoforms.

As in other PAD enzymes of known structure (Arita *et al.*, 2004; Saijo *et al.*, 2016; Slade *et al.*, 2015), the crystal packing of the PAD3 molecules shows that two molecules from adjacent asymmetric units form a head-to-tail dimer with an extensive interaction interface of 2388 Å^2 as calculated using *PISA* (Krissinel & Henrick, 2007; Fig. 2). This dimeric arrangement is similar to that seen in PAD2 and PAD4, and the extent of the interaction interface is similar to that in PAD2 (2551 Å^2; PDB entries 4n20 and 4n2b) and PAD4 (2314 Å^2; PDB entries 1wd8 and 1wd9). PAD1 is the exception in this protein family, crystallizing as a monomer (PDB entry 5hp5; Table 5). A previous study (Saijo *et al.*, 2016) confirmed using SAXS experiments that PAD3 also forms a dimer in solution, like PAD2 and PAD4, whereas PAD1 is monomeric in solution. Saijo and coworkers also reported that the N-terminal loop of PAD1 appears to be the feature that prevents the formation of a native homodimer. Previous mutagenesis experiments have shown that PAD4 dimer-interface mutants can produce a monomeric form of this protein, but that it retains only 25–50% of the activity of the native dimeric form (Liu *et al.*, 2011). The monomers in the PAD3 crystal are related by a crystallographic twofold axis. The calculated electrostatic surface potential of the PAD3 monomer shows that the protein has two distinct faces: one is acidic and solvent-exposed, whereas the other is mostly basic and hydrophobic in character, forming a mainly hydrophobic dimerization interface flanked by a conserved salt bridge between Arg8 of subunit *A* and Asp549 of subunit *B* (Fig. 2*a*). The acidic face contains a
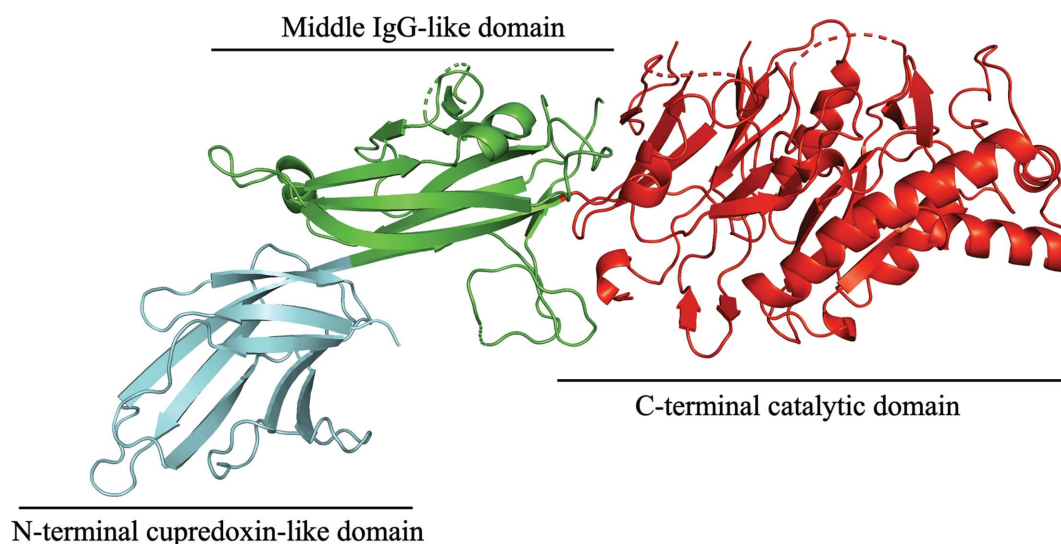

Figure 1
Cartoon representation of the PAD3 monomer structure. The N-terminal cupredoxin-like domain, the central IgG-like domain and the C-terminal catalytic domain are coloured cyan, green and red, respectively.

Table 6
Structural superpositions of PAD3 with other PAD isozymes.

	PDB code	Overall (3–664)		N-terminal cupredoxin-like domain (3–115)		Middle IgG-like domain (116–292)		C-terminal catalytic domain (293–664)	
		Sequence identity (%)	R.m.s.d. (Å)	Sequence identity (%)	R.m.s.d. (Å)	Sequence identity (%)	R.m.s.d. (Å)	Sequence identity (%)	R.m.s.d. (Å)
PAD1, Ca^{2+}	5hp5	58	2.98	40	2.32	48	2.69	69	1.91
PAD2	4n20	52	3.73	24	2.82	46	2.58	64	2.38
PAD2, Ca^{2+}	4n2b	52	3.73	24	2.82	46	3.73	64	2.12
PAD4	1wd8	56	2.65	35	2.72	49	2.92	69	2.00
PAD4, Ca^{2+}	1wd9	56	2.67	35	2.22	49	2.62	69	2.41

concave groove that forms the active-site cleft, and the equivalent region is subject to a conformational change upon Ca^{2+} binding in PAD4 (Arita *et al.*, 2004).

PAD3 and PAD4 share 56% sequence identity, and the overall r.m.s.d. between their two structures when overlaid is 2.65 Å (Table 6). Domain-based sequence alignments between PAD3 and the other PAD isozymes show that the IgG-like domains have the lowest sequence identity (25–50%), whereas the C-terminal domains have the highest sequence identity

(60–70%) (Table 6). In keeping with these statistics, the C-terminal domains of the PADs are structurally more conserved than the IgG-like domains, which show a greater structural variety.

The active site of PAD3 is formed by a conserved catalytic tetrad composed of Asp350, His470, Asp472 and Cys646. The loop connecting strands $\beta35$ and $\beta36$ is poorly ordered, and is in a conformation that positions Cys646 at a distance of ~10 Å from the inferred active-site centre. Thus, this structure shows
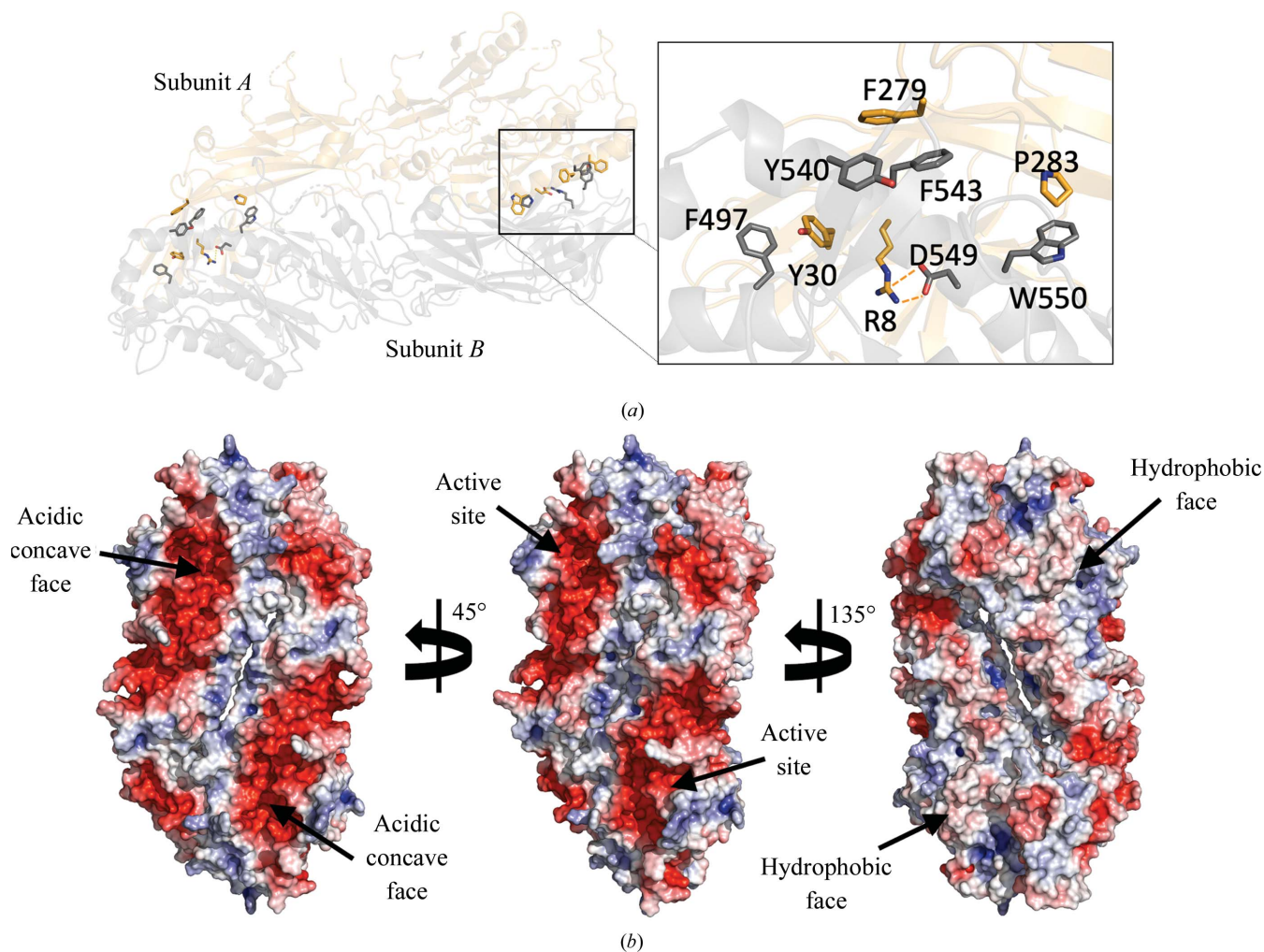

Figure 2
The PAD3 dimer. (*a*) Cartoon representation of two PAD3 molecules arranged in a head-to-tail manner. Residues involved in the head-to-tail dimer interface are represented as sticks. Residues Arg8/Asp549 form a salt bridge and Phe279/Tyr540/Phe543, Phe497/Tyr30 and Pro283/Trp550 form hydrophobic clusters. (*b*) The whole dimer is shown with a solvent-accessible surface coloured by electrostatic potential, showing the two faces of the molecule. Electrostatic potentials were calculated using *APBS* (Baker *et al.*, 2001; Jurrus *et al.*, 2018; red = −90 k_BT/e_c, blue = +90 k_BT/e_c).

a catalytically inactive state of the enzyme. There are known conformational differences in the C-terminal domains of Ca^{2+}-free and Ca^{2+}-bound forms of other PAD isoforms that are essential for enzyme function: in PAD2, the equivalent active-site cysteine residue (Cys645) is distant from the active site in the calcium-free form, but Ca^{2+} binding moves it by ~5 Å, thus facilitating nucleophilic attack by the cysteine on the C^{ζ} atom of the substrate (Arita *et al.*, 2004), and an equivalent conformational change is presumably also important in PAD3.

Previous PAD structures have shown up to six bound Ca^{2+} ions. In order to determine the putative Ca^{2+}-binding sites of PAD3, we performed structural and sequence alignments of PAD1, PAD2, PAD3 and PAD4 (Fig. 3). Examination of the alignments indicates that the amino acids coordinating Ca^{2+} ions at sites 1, 2, 3 and 4 are conserved among all PADs, whereas site 5 is not present in PAD1 and site 6 is not present in either PAD1 or PAD4. The structure of PAD1 contains two Ca^{2+} ions, even without added Ca^{2+} in the crystallization

conditions, but no such ions are evident in the structure of PAD3. Hence, in order to gain insight into the Ca^{2+}-mediated regulatory processes that control PAD3 activity, we sought to determine the structure of PAD3 in its Ca^{2+}-bound state. We attempted to do so by co-crystallization and also by soaking apo PAD3 crystals in crystallization solutions containing up to 50 mM $CaCl_2$. However, neither approach produced good-quality crystals.

The crystal structure of apo PAD3 described here shows that like other PAD isoforms, PAD3 has a conserved architecture and forms a head-to-tail dimer similar to other PAD isoforms (Arita *et al.*, 2004; Slade *et al.*, 2015; Saijo *et al.*, 2016). The Ca^{2+}-free form of the enzyme corresponds to the low-activity state of the enzyme (Knuckley *et al.*, 2010).

The biological function of PAD enzymes is only partially understood, with the molecular details of the specificity of PAD3 for its substrates (keratins, filaggrin and trichohyalin) remaining unclear, hampered by difficulties in obtaining

|	PAD1	PAD2	PAD3	PAD4
Ca^{2+} 1	Q351	Q350	Q349	Q349
	E355	E354	E353	E353
	F409	F408	F407	F407
	L412	L411	L410	L410
	D413	E412	E411	E411
Ca^{2+} 2	E353	E352	E351	E351
	D371	D370	D369	D369
	S372	S371	S370	S370
	N375	D374	N373	N373
Ca^{2+} 3	N153	N154	N153	N153
	D155	D156	D155	D155
	D157	D158	D155	D157
	D165	D166	D165	D165
	D176	D177	D176	D176
	D179	D180	D179	D179
Ca^{2+} 4	D155	D156	D155	D155
	D157	E158	D157	D157
	D179	D180	D179	D179
	D390	D389	D388	D388
Ca^{2+} 5	D165	D166	D165	D165
	S169	D169	D168	D168
	L171	K171	H170	E170
	D176	D177	D176	D176
	E252	E253	E252	E252
Ca^{2+} 6	D122	D123	D122	D120
	G124	D125	N124	T124
	T127	D127	E126	T126
	V129	V129	Q129	V127
	K130	E131	D130	K128

(*a*)

(*b*)

(*c*)
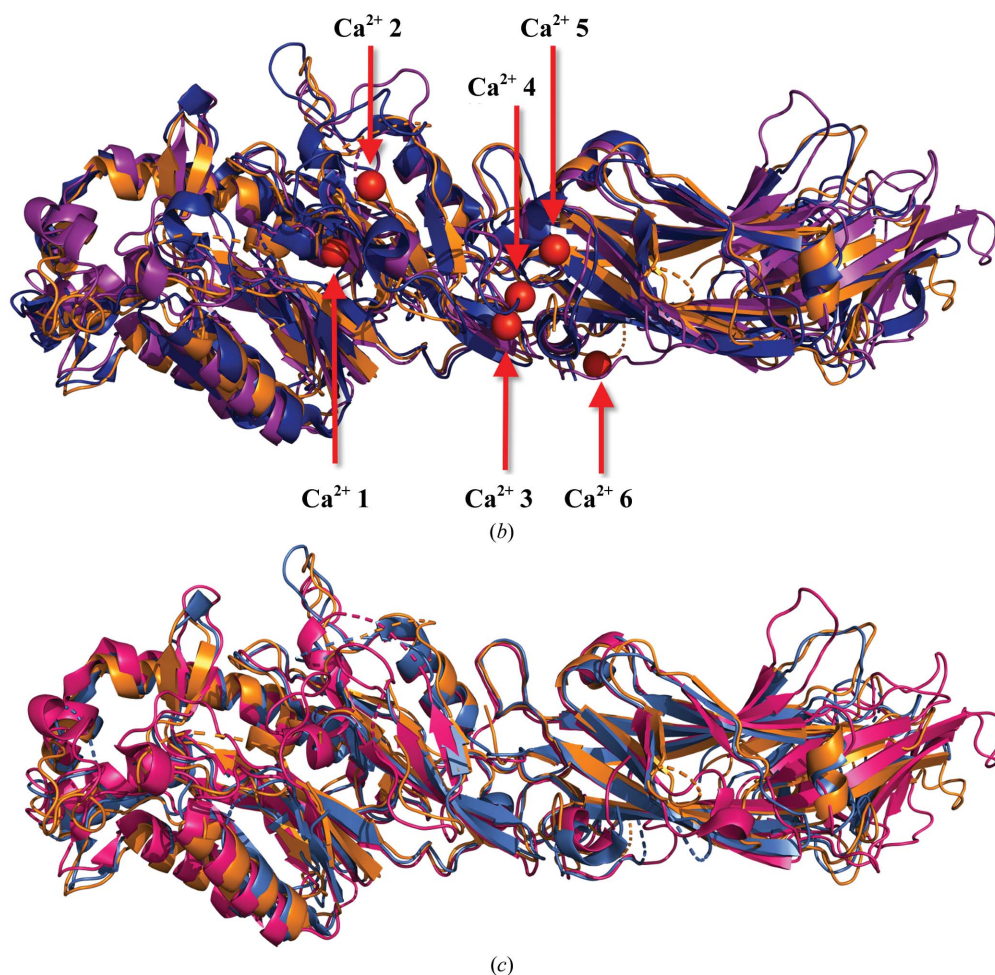
Figure 3
The Ca^{2+}-binding sites of PADs. (*a*) Ca^{2+}-binding sites and their respective coordinating residues determined from X-ray crystallographic structures of PAD1 (PDB entry 5hp5), PAD2 (PDB entries 4n2b and 4n2c) and PAD4 (PDB entries 1wd8 and 1wd9). A structural alignment of PADs (PAD1, PAD2, PAD3 and PAD4) was conducted to determine the residues in PAD3 that are likely to be involved in Ca^{2+} coordination. The Ca^{2+}-coordinating residues in PAD1, PAD2 and PAD4 that were visualized by X-ray crystallography (Arita *et al.*, 2004; Slade *et al.*, 2015; Saijo *et al.*, 2016) are listed in black. The putative Ca^{2+}-coordinating residues in PAD3 are listed in blue. The residues shown to not coordinate Ca^{2+} in PAD1 and PAD4 are listed in green. (*b*) Structure superposition of PAD3 (PDB entry 6ce1, light orange) with Ca^{2+}-bound PAD2 (PDB entry 4n2b, purple) and Ca^{2+}-bound PAD4 (PDB entry 1wd9, blue). All Ca^{2+} sites are shown in red for descriptive purposes. (*c*) Structure superposition of PAD3 (PDB entry 6ce1, light orange) with apo PAD2 (PDB entry 4n20, hot pink) and apo PAD4 (PDB entry 1wd8, light blue).

soluble substrate proteins. Structures of PAD3 in complex with Ca^{2+} and substrate analogues are required to more precisely define the molecular determinants of substrate recognition in this protein.

References

Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* D**68**, 352–367.

Arita, K., Hashimoto, H., Shimizu, T., Nakashima, K., Yamada, M. & Sato, M. (2004). *Nat. 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**, 5291–5296.6.

Assohou-Luty, C., Raijmakers, R., Benckhuijsen, W. E., Stammen-Vogelzangs, J., de Ru, A., van Veelen, P. A., Franken, K. L., Drijfhout, J. W. & Pruijn, G. J. (2014). *Biochim. Biophys. Acta*, **1844**, 829–836.

Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.

Basmanav, F. B. Ü., Cau, L., Tafazzoli, A., Méchin, M.-C., Wolf, S., Romano, M. T., Valentin, F., Wiegmann, H., Huchenq, A., Kandil, R., Garcia Bartels, N., Kilic, A., George, S., Ralser, D. J., Bergner, S., Ferguson, D. J. P., Oprisoreanu, A. M., Wehner, M., Thiele, H., Altmüller, J., Nürnberg, P., Swan, D., Houniet, D., Büchner, A., Weibel, L., Wagner, N., Grimalt, R., Bygum, A., Serre, G., Blume-Peytavi, U., Sprecher, E., Schoch, S., Oji, V., Hamm, H., Farrant, P., Simon, M. & Betz, R. C. (2016). *Am. J. Hum. Genet.* **99**, 1292–1304.

Bruschi, M., Petretto, A., Bertelli, R., Galetti, M., Bonanni, A., Pratesi, F., Migliorini, P., Candiano, G., Vaglio, A. & Ghiggeri, G. M. (2017). *Clin. Chim. Acta*, **464**, 12–16.

Chavanas, S., Adoue, V., Méchin, M.-C., Ying, S., Dong, S., Duplan, H., Charveron, M., Takahara, H., Serre, G. & Simon, M. (2008). *PLoS One*, **3**, e3408.

Chavanas, S., Méchin, M.-C., Nachat, R., Adoue, V., Coudane, F., Serre, G. & Simon, M. (2006). *J. Dermatol. Sci.* **44**, 63–72.

Das, K., Butler, G. H., Kwiatkowski, V., Clark, A. D. Jr, Yadav, P. & Arnold, E. (2004). *Structure*, **12**, 657–667.

Dong, S., Kanno, T., Yamaki, A., Kojima, T., Shiraiwa, M., Kawada, A., Méchin, M.-C., Chavanas, S., Serre, G., Simon, M. & Takahara, H. (2006). *Biochem. J.* **397**, 449–459.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.

Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.

Gorrec, F. (2009). *J. Appl. Cryst.* **42**, 1035–1042.

Ishida-Yamamoto, A., Takahashi, H., Iizuka, H., Senshu, T., Akiyama, K. & Nomura, K. (2000). *J. Investig. Dermatol.* **114**, 701–705.

Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., Brookes, D. H., Wilson, L., Chen, J., Liles, K., Chun, M., Li, P., Gohara, D. W., Dolinsky, T., Konecny, R., Koes, D. R., Nielsen, J. E., Head-Gordon, T., Geng, W., Krasny, R., Wei, G.-W., Holst, M. J., McCammon, J. A. & Baker, N. A. (2018). *Protein Sci.* **27**, 112–128.

Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.

Kizawa, K., Takahara, H., Troxler, H., Kleinert, P., Mochida, U. & Heizmann, C. W. (2008). *J. Biol. Chem.* **283**, 5004–5013.

Knuckley, B., Bhatia, M. & Thompson, P. R. (2007). *Biochemistry*, **46**, 6578–6587.

Knuckley, B., Causey, C. P., Jones, J. E., Bhatia, M., Dreyton, C. J., Osborne, T. C., Takahara, H. & Thompson, P. R. (2010). *Biochemistry*, **49**, 4852–4863.

Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.

Kubilus, J. & Baden, H. P. (1983). *Biochim. Biophys. Acta*, **745**, 285–291.

Kubilus, J., Waitkus, R. F. & Baden, H. P. (1980). *Biochim. Biophys. Acta*, **615**, 246–251.

Kubilus, J., Waitkus, R. W. & Baden, H. P. (1979). *Biochim. Biophys. Acta*, **581**, 114–121.

Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* D**75**, 861–877.

Liu, Y.-L., Chiang, Y.-H., Liu, G.-Y. & Hung, H.-C. (2011). *PLoS One*, **6**, e21314.

Malki, L., Sarig, O., Romano, M. T., Méchin, M. C., Peled, A., Pavlovsky, M., Warshauer, E., Samuelov, L., Uwakwe, L., Briskin, V., Mohamad, J., Gat, A., Isakov, O., Rabinowitz, T., Shomron, N., Adir, N., Simon, M., McMichael, A., Dlova, N. C., Betz, R. C. & Sprecher, E. (2019). *N. Engl. J. Med.* **380**, 833–841.

Méchin, M.-C., Enji, M., Nachat, R., Chavanas, S., Charveron, M., Ishida-Yamamoto, A., Serre, G., Takahara, H. & Simon, M. (2005). *Cell. Mol. Life Sci.* **62**, 1984–1995.

Mondal, S. & Thompson, P. R. (2019). *Acc. Chem. Res.* **52**, 818–832.

Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I. & Perrakis, A. (2005). *Acta Cryst.* D**61**, 1426–1431.

Painter, J. & Merritt, E. A. (2006a). *Acta Cryst.* D**62**, 439–450.

Painter, J. & Merritt, E. A. (2006b). *J. Appl. Cryst.* **39**, 109–111.

Rogers, G. E., Harding, H. W. & Llewellyn-Smith, I. J. (1977). *Biochim. Biophys. Acta*, **495**, 159–175.

Rothnagel, J. A. & Rogers, G. E. (1984). *Methods Enzymol.* **107**, 624–631.

Saijo, S., Nagai, A., Kinjo, S., Mashimo, R., Akimoto, M., Kizawa, K., Yabe-Wada, T., Shimizu, N., Takahara, H. & Unno, M. (2016). *J. Mol. Biol.* **428**, 3058–3073.

Senshu, T., Kan, S., Ogawa, H., Manabe, M. & Asaga, H. (1996). *Biochem. Biophys. Res. Commun.* **225**, 712–719.

Shirai, H., Blundell, T. L. & Mizuguchi, K. (2001). *Trends Biochem. Sci.* **26**, 465–468.

Slade, D. J., Fang, P., Dreyton, C. J., Zhang, Y., Fuhrmann, J., Rempel, D., Bax, B. D., Coonrod, S. A., Lewis, H. D., Guo, M., Gross, M. L. & Thompson, P. R. (2015). *ACS Chem. Biol.* **10**, 1043–1053.

Steinert, P. M., Parry, D. A. & Marekov, L. N. (2003). *J. Biol. Chem.* **278**, 41409–41419.

Tarcsa, E., Marekov, L. N., Mei, G., Melino, G., Lee, S. C. & Steinert, P. M. (1996). *J. Biol. Chem.* **271**, 30709–30716.

Unno, M., Kizawa, K., Ishihara, M. & Takahara, H. (2012). *Acta Cryst.* F**68**, 668–670.

Vossenaar, E. R., Zendman, A. J., van Venrooij, W. J. & Pruijn, G. J. (2003). *Bioessays*, **25**, 1106–1118.

Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. & Richardson, J. S. (2018). *Protein Sci.* **27**, 293–315.