Structural Basis for Histone N-terminal Peptides Recognition by Human Peptidylarginine Deiminase 4

The eukaryotic genome is organized within the nucleosome core particle, which is composed of core histones and their associated DNA. A diverse array of core particle modifications play a fundamental role in specific nucleosome regulation [1]. In these modifications, the Ca^{2+}-dependent citrullination (or deimination) of arginine residues of histone by peptidylarginine deiminase 4 (PAD4) has come into focus as a novel post-translational modification linked to transcriptional regulation in eukaryotes. To date, five types of human PAD, namely, PAD1-PAD4 and PAD6, have been characterized by cDNA cloning [2]. Some of these PADs are also associated with human diseases such as rheumatoid arthritis (RA), in which autoantibodies against proteins citrullinated by PAD are found at early phases of the manifestation of the disease in the majority of RA patients. Furthermore, a significant association was reported to exist between RA and functional variants of the gene encoding PAD4 in the Japanese population [3]. PAD4 is expressed in bloodstream granulocytes, and is the only type of PAD that has a nuclear localization signal and is thus localized in the cell nucleus. In view of this, we previously determined the crystal structures of Ca^{2+}-free PAD4 and a Ca^{2+}-bound inactive mutant with and without a substrate, benzoyl-L-arginine amide, and showed a novel mechanism of the Ca^{2+}-induced activation of the enzyme [4]. Here, we define the structural basis for histone N-terminal peptides recognition by determining the crystal structures of the inactive Ca^{2+}-bound C645A mutant in complexes with three histone N-terminal peptides named peptides H3-1, H3-2, and H4, each consisting of 10 residues that include one target arginine residue for the enzyme [5]. Diffraction data for the Ca^{2+}-bound PAD4 (C645A) in complex with H3-1 and H4 were collected on **BL38B1** at SPring-8 using an ADSC Quantum 4R CCD detector, and those for the complex with peptide H3-2 were collected on NW12 at PF-AR using an ADSC Quantum 210 CCD detector.

The structures of Ca^{2+}-bound PAD4 (C645A) in complexes with the three histones N-terminal peptides are similar to each other. PAD4 has five non-EF-hand Ca^{2+}-binding sites and adopts an elongated fold with two domains (Fig. 1). The N-terminal domain (Met1 to Pro300) is divided into two immunoglobulin-like subdomains. Subdomain 1 has nine β-strands and a nuclear localization signal on the molecular surface in a loop region. The amino acid alterations in functional variants of the enzyme in the Japanese RA population are positioned in subdomain 1, far from the active site in the C-terminal domain, and are unlikely to affect catalytic function directly. Subdomain 2 has ten β-strands, four short α-helices, and three Ca^{2+} ions. The C-terminal domain (Asn301 to Pro663) has a structure of five β β α β modules called an α/β propeller. The active site cleft, in which each histone N-terminal peptide and two Ca^{2+} ions are bound, is found in the α/β propeller structure in the C-terminal domain.

In each histone N-terminal peptide, five successive residues at (N – 2), (N – 1), N, (N + 1), and (N + 2), where N– and N+ define the positions immediately preceding and following the target arginine residue, were unambiguously assigned in $|F_o| - |F_c|$ maps (Fig. 2). The remaining five residues are highly disordered in the three structures. The side chain of the target arginine residue is recognized within the active site cleft in the same manner as that seen for the

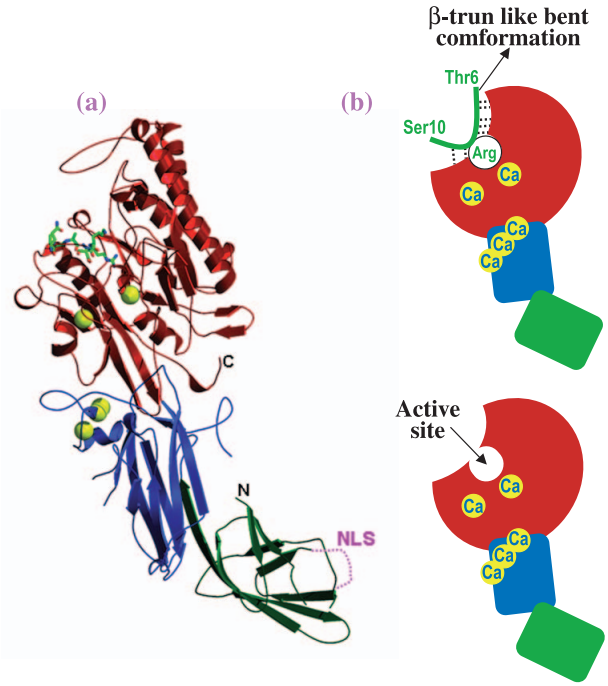

Fig. 1. Structure of Ca^{2+}-bound PAD4 (C645A) in complex with peptide H3-1. **(a)** Ribbon representation of the structure. Ca^{2+} ions and the histone peptide are shown as yellow balls and as a green stick model, respectively. N-terminal subdomains 1 (residues 1-118) and 2 (residues 119-300) and the C-terminal domain (residues 301-663) are colored green, blue, and red, respectively. The nuclear localization signal (NLS) region in subdomain 1 is shown as a dotted line. **(b)** Schematic representations of the structure shown in **(a)** (upper panel). The colors of the N-terminal (subdomains 1 and 2) and C-terminal domains are the same as those of **(a)**. Dotted lines show hydrogen bonds that form a consensus recognition motif at the molecular surface near the active site. For reference, the structure of the Ca^{2+}-bound PAD4 (C645A) is shown in the bottom panel.

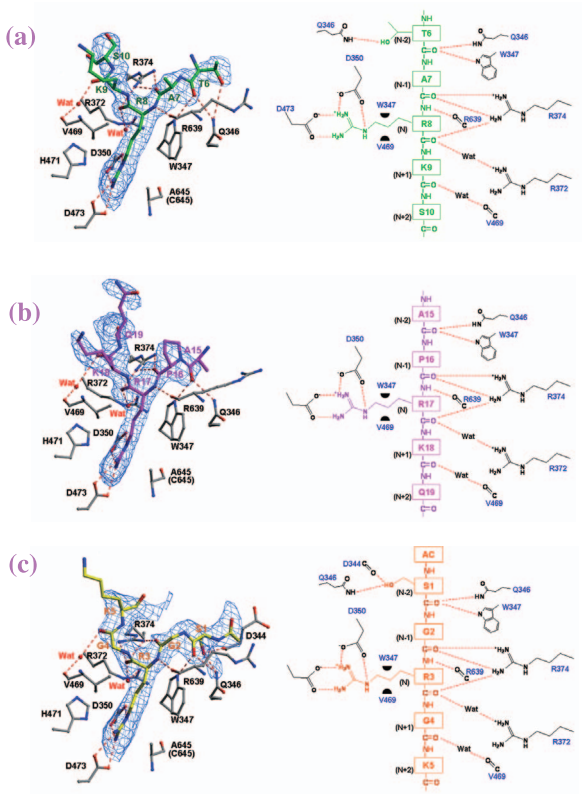

Fig. 2. Structures around active sites of Ca^{2+}-bound PAD4 (C645A) in complex with peptides H3-1, H3-2, and H4. Left, ball-and-stick representation of the structures. The protein moiety is colored grey, and the peptides H3-1 **(a)**, H3-2 **(b)**, and H4 **(c)** are colored green, magenta, yellow, and cyan, respectively. The $|F_o| - |F_c|$ electron densities of the peptides are superimposed, contoured at 2σ. Right, schematic diagrams of the structures on the left. Dotted lines and green half-circles show hydrogen bonds and hydrophobic interactions, respectively.

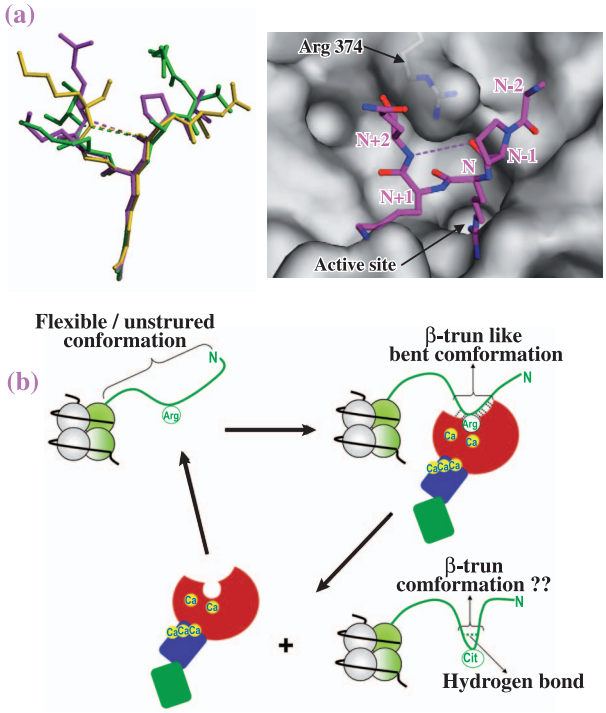

Fig. 3. Histone N-terminal peptide structures. **(a)** Left, structural comparison of PAD4-bound forms. Three peptides, namely H3-1, H3-2, and H4, are shown as ball-and-stick representations colored green, magenta, and yellow, respectively, as in Fig. 2 **(a)**, **(b)**, and **(c)**. Right, top view of the peptide H3-2 structure shown on the left of this figure, together with a molecular surface representation near the active site cleft. The weak intrapeptide interactions between the backbone oxygen at the (N − 1) position and the backbone nitrogen at the (N + 2) position are shown as dotted lines. **(b)** Possible conformational change of N-terminal histone tail in histone citrullination. The N-terminal histone inal tail protruding from the nucleosome core particle is shown as green ribbon.

previously determined BA complex [4]. Peptide recognition at the molecular surface near the active site cleft occurs commonly through backbone atoms of the peptides at the (N – 2), (N – 1), N, and (N + 1) positions (Fig. 2 and Fig. 3). This indicates that PAD4 can target multiple arginine sites in N-terminal histone tails without sequence specificity.

N-terminal histone tails are flexible and protrude from the nucleosome core particle [5]. However, the enzyme induces β-turn-like bent conformations in the histone peptides at the molecular surface near the active site cleft (Fig. 3(a)). It is, therefore, important for target arginine recognition by PAD4 that the local peptide around the target arginine residue takes a highly disordered conformation, because the enzyme recognizes a flexible/unstructured peptide at the molecular surface near the active site cleft and induces a β-turn-like bent conformation (Fig. 3(b)). Most histone-modifying enzymes interact extensively with their peptide ligands and recognize a specific amino acid in a sequence-specific manner. But, this is not the case for PAD4. The enzyme can recognize multiple arginine sites in flexible/unstructured peptides in a non-sequence-specific manner.

Kyouhei Arita, Toshiyuki Shimizu, Hiroshi Hashimoto and Mamoru Sato*

Field of Supramolecular Biology,
Yokohama City University

*E-mail: msato@tsurumi.yokohama-cu.ac.jp

References

[1] S. Khorasanizadeh: Cell **116** (2004) 259.
[2] E.R. Vossenaar *et al.*: BioEssays **25** (2003) 1106.
[3] A. Suzuki *et al.*: Nature Genet. **34** (2003) 395.
[4] K. Arita *et al.*: Nature Struct. Mol. Biol. **11** (2004) 777.
[5] *K. Arita, T. Shimizu, H. Hashimoto, Y. Hidaka, M. Yamada and M. Sato: Proc. Natl. Acad. Sci. USA **103** (2006) 5291.*