

# SCIENTIFIC REPORTS



OPEN

## Structure and mechanism of a bacterial host-protein citrullinating virulence factor, *Porphyromonas gingivalis* peptidylarginine deiminase

Received: 16 April 2015

Accepted: 12 June 2015

Published: 01 July 2015

Theodoros Goulas<sup>1,\*</sup>, Danuta Mizgalska<sup>4,\*</sup>, Irene Garcia-Ferrer<sup>1</sup>, Tomasz Kantyka<sup>4,5</sup>, Tibisay Guevara<sup>1</sup>, Borys Szmigielski<sup>4</sup>, Aneta Sroka<sup>4</sup>, Claudia Millán<sup>3</sup>, Isabel Usón<sup>3,6</sup>, Florian Veillard<sup>7,†</sup>, Barbara Potempa<sup>7</sup>, Piotr Mydel<sup>8</sup>, Maria Solà<sup>2</sup>, Jan Potempa<sup>4,7</sup> & F. Xavier Gomis-Rüth<sup>1</sup>

Citrullination is a post-translational modification of higher organisms that deiminates arginines in proteins and peptides. It occurs in physiological processes but also pathologies such as multiple sclerosis, fibrosis, Alzheimer's disease and rheumatoid arthritis (RA). The reaction is catalyzed by peptidylarginine deiminases (PADs), which are found in vertebrates but not in lower organisms. RA has been epidemiologically associated with periodontal disease, whose main infective agent is *Porphyromonas gingivalis*. Uniquely among microbes, *P. gingivalis* secretes a PAD, termed PPAD (*Porphyromonas* peptidylarginine deiminase), which is genetically unrelated to eukaryotic PADs. Here, we studied function of PPAD and its substrate-free, substrate-complex, and substrate-mimic-complex structures. It comprises a flat cylindrical catalytic domain with five-fold  $\alpha/\beta$ -propeller architecture and a C-terminal immunoglobulin-like domain. The PPAD active site is a funnel located on one of the cylinder bases. It accommodates arginines from peptide substrates after major rearrangement of a "Michaelis loop" that closes the cleft. The guanidinium and carboxylate groups of substrates are tightly bound, which explains activity of PPAD against arginines at C-termini but not within peptides. Catalysis is based on a cysteine-histidine-asparagine triad, which is shared with human PAD1-PAD4 and other guanidino-group modifying enzymes. We provide a working mechanism hypothesis based on 18 structure-derived point mutants.

<sup>1</sup>Proteolysis Lab; Department of Structural Biology ("María de Maeztu" Unit of Excellence); Molecular Biology Institute of Barcelona, CSIC; Barcelona Science Park, Helix Building; c/Baldiri Reixac, 15–21; E-08028 Barcelona Spain.

<sup>2</sup>Structural MitoLab; Department of Structural Biology ("María de Maeztu" Unit of Excellence); Molecular Biology Institute of Barcelona, CSIC; Barcelona Science Park, Helix Building; c/Baldiri Reixac, 15–21; E-08028 Barcelona Spain.

<sup>3</sup>Crystallographic Methods; Department of Structural Biology ("María de Maeztu" Unit of Excellence); Molecular Biology Institute of Barcelona, CSIC; Barcelona Science Park, Helix Building; c/Baldiri Reixac, 15–21; E-08028 Barcelona Spain.

<sup>4</sup>Laboratory of Microbiology; Faculty of Biochemistry, Biophysics and Biotechnology; Jagiellonian University; Gonostajowa 7; PL-30–387 Kraków Poland.

<sup>5</sup>Małopolska Center of Biotechnology, Jagiellonian University; Gonostajowa 7; PL-30–387 Kraków Poland.

<sup>6</sup>Institució Catalana de Recerca i Estudis Avançats; Passeig Lluís Companys, 23; E-08003 Barcelona Spain.

<sup>7</sup>Department of Oral Immunology and Infectious Diseases; University of Louisville School of Dentistry; Louisville, KY 40202 USA.

<sup>8</sup>Broegelmann Research Laboratory; Department of Clinical Science; The Gade Institute; University of Bergen; N-5021 Bergen Norway.

<sup>†</sup>Current address: Institut de Biologie Moléculaire et Cellulaire, UPR9022 CNRS, F-37000 Strasbourg France.

\*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M. S. (email: msvcri@ibmb.csic.es) or J. P. (email: jan.potempa@louisville.edu) or F. X. G.-R. (email: xgrcri@ibmb.csic.es)

L-Citrulline ( $N^5$ -carbamoyl-L-ornithine) is a non-proteinogenic amino acid that is an intermediate in the Krebs-Henseleit urea cycle in animals<sup>1</sup>. It is produced by the enterocytes of the small bowel in humans and its accumulation in plasma can cause citrullinemia, an autosomal recessive disorder characterized by increased citrulline secretion in the urine and neuropsychiatric symptoms<sup>2</sup>. Citrulline is also a biological precursor for nitric oxide and its therapeutic administration has been proposed for the mitochondrial MELAS syndrome<sup>3</sup>. Citrulline further results from free arginine by citrullination, which entails replacement of the guanidino group with an ureido group through deimination. This removes the positive charge of the arginine side chain and liberates ammonia.

Of greater physiological relevance, however, is the citrullination of arginines in peptides and proteins through post-translational modification<sup>4,5</sup>. Given the limited number of genes in the genomes of higher organisms, such post-translational modifications increase the structural and functional diversity of the proteomes<sup>4</sup>. Citrullination may result in changes in fold, function and half-life of proteins and peptides, and the reaction is catalyzed in a calcium-dependent manner by peptidylarginine deiminases (PADs). These occur only in vertebrates, where five close paralogs (PAD1-PAD4 and PAD6) have been described<sup>6–8</sup>. Their activity is essential for skin keratinization, neuron insulation, and plasticity of the central nervous system as well as histone core-protein regulation<sup>5,7</sup>. Furthermore, through involvement of PADs in apoptosis, autophagy, and NETosis, citrullination plays a major role in the immune system.

However, citrullination also has an established role in pathology, which has lately catapulted interest in the reaction since increased levels of citrullinated proteins are found in several if not all inflammatory diseases<sup>9</sup> and have been directly implicated in Alzheimer's disease, prion diseases, psoriasis, multiple sclerosis and tumorigenesis<sup>5</sup>. In a specific genetic background, citrullinated proteins act as autoantigens to generate anti-citrullinated protein antibodies, which participate in an abnormal autoimmune response, a hallmark of rheumatoid arthritis (RA<sup>10</sup>). The latter is a common systemic disease affecting ~1% of the general population in the developed world that is characterized by chronic inflammation of the synovial joints, eventually leading to progressive joint destruction and, despite many years of intensive research, its mechanisms of disease progression are still poorly understood. As to etiology, genetic factors, environmental influences—such as smoking and oral contraceptives—, and concomitant microbial infections are risk factors for developing RA<sup>11</sup>.

Inflammation is also a hallmark of chronic periodontal disease (PD), which is among the most prevalent infectious diseases of mankind<sup>12</sup>. In its severe form, the disease affects the gums of 10–15% of adults, potentially leading to tissue destruction and tooth loss<sup>13</sup>. Its major causal agent is *Porphyromonas gingivalis*, a bacterium that is also implicated in cardiovascular diseases, respiratory diseases, diabetes, osteoporosis, and pre-term low birth-weight. More recently, epidemiological studies have further reported an increased prevalence of PD in RA<sup>14,15</sup>, which is consistent with the antique claim made by Hippocrates ~2,400 years ago that removal of bad teeth cures arthritis<sup>16</sup>.

Within the virulence-factor armamentarium of *P. gingivalis* are several secreted cysteine peptidases such as lysine (Kgp) and arginine gingipains A and B (RgpA and RgpB). These are cysteine endopeptidases cleaving after lysines and arginines, respectively, and they participate not only in nutrient acquisition but also in host-tissue destruction and defense inactivation<sup>17</sup>. Uniquely among microbes to date, *P. gingivalis* also produces a secreted PAD (called PPAD<sup>18</sup>), which protects *P. gingivalis* during acidic cleansing in the mouth through ammonia generated during host and endogenous protein citrullination<sup>19</sup>. PPAD does not require calcium for catalysis<sup>20</sup> and is genetically unrelated with animal PADs and, like the latter and cysteine peptidases, its main catalytic residue is a cysteine (C<sup>351</sup> in PPAD<sup>20</sup>).

Host PADs process arginines within polypeptide chains but not at their termini, i.e. they are efficient endodeiminases but poor exodeiminases<sup>4</sup>. In contrast, PPAD citrullinates C-terminal arginines like those generated by the prior action of Rgps<sup>17</sup>, which may be facilitated by the surface co-localization of Rgps and PPAD<sup>20</sup>. In this way, PPAD complements endogenous PADs and creates new exogenous epitopes for autoimmune response, which have been associated with RA disease progression<sup>15</sup>. Taken together, all these results suggest that the link between RA and *P. gingivalis*-induced PD may result from PPAD-mediated citrullination<sup>15</sup>.

To shed light on the molecular aspects of this key enzyme for pathogenicity, we analyzed the structure and function of PPAD in various functional states and proposed a working model for the enzyme based on mutational studies, which places PPAD in a wider context with PADs and functionally more distant enzymes.

## Results and Discussion

**Molecular structure of PPAD.** PPAD was recently reported to belong to a family of secreted *P. gingivalis* proteins, which includes Kgp and Rgps<sup>20</sup>. These proteins possess a ~75-residue C-terminal domain (CTD) for maturation and translocation through the outer membrane *via* the PorSS, PerioGate or Type-IX secretion system, which removes the CTD upon secretion<sup>21</sup>. Accordingly and similarly to Kgp and Rgps, full-length PPAD would span a pro-peptide, a catalytic domain (CD), an immunoglobulin-superfamily domain (IgSF), and a CTD<sup>18</sup>. We obtained a fragment of PPAD by homologous overexpression in *P. gingivalis* that was equivalent to the purified form from *P. gingivalis* supernatant<sup>18</sup> and, thus, lacked the pro-peptide and the CTD (residues A<sup>44</sup>-A<sup>475</sup>; see Table 1). We solved three distinct structures to high

Name	Sequence (5'–>3')	Name	Sequence (5'–>3')
pTCowPPADf	<u>tgca</u> gctagctccttaagggtgagatatacg	H236Fs	ggcaagtattggcaccgaa
pTCowPPADr	ctgac <u>gcat</u> gccaatcggtcgtaggtctcc	H236Rs	gatataatccgcttcggatct
W127AFt	tcttacgctacacgcgactataccggttggtcgcaa	H236AFt	aacgctgtgactgtggggcaagtattggcaccgaa
W127ARt	gtcgcgtgtagcgaagagtcagtttcccaatgatgaa	H236ARt	ccaacagtcacacggttgatataatccgcttcggatct
D130Fs	tataccggttggtcgcaa	H236NFt	aacaatgtgactgtggggcaagtattggcaccgaa
D130Rs	gctagtttcccaatgatgaa	H236NRt	ccaacagtcacattgttgatataatccgcttcggatct
D130AFt	tcttactggacacgcgcttataccggttggtcgcaa	D238AFt	aacctgtgcttggggcaagtattggcaccgaa
D130ARt	agcgcgtgtcagtaagagtcagtttcccaatgatgaa	D238ARt	ccaacaagccacatggttgatataatccgcttcggatct
D130NFt	tcttactggacacgcaactataccggttggtcgcaa	D238NFt	aacctgtgaattgtggggcaagtattggcaccgaa
D130NRt	gttgcgtgtccagtaagagtcagtttcccaatgatgaa	D238NRt	ccaacaattcacatggttgatataatccgcttcggatct
R152AFs	cctcgtcctaacgatgaa	D239AFt	aacctgtgacgctggggcaagtattggcaccgaa
R152ARs	cacgagcactcttctgtatc	D239ARt	ccaagctccacatggttgatataatccgcttcggatct
R152AFt	gactttataaacgcccctgcttaacgatgaa	D239Eft	aacctgtggacgaatggggcaagtattggcaccgaa
R152ARt	ggcgttataaataagtcacgagacctcttctgtatc	D239ERt	ccattcgtccacatggttgatataatccgcttcggatct
R154Fs	gaattccccaaatacgaagc	D239SFt	aacctgtgactctggggcaagtattggcaccgaa
R154Rs	gcggttataaataagtcacg	D239SRt	ccaagctccacatggttgatataatccgcttcggatct
R154AFt	cctgctcctaacgatgaaatccccaaatacgaagc	N297AFs	acaacagggtattgttctg
R154ARt	atcatcgttaggacagggcggttataaataagtcacg	N297ARs	gtacggttcttctgtg
R154EFt	cctgaacctaacgatgaaatccccaaatacgaagc	N297AFt	cggcttctctgattcgaacaacagggtattgttctg
R154ERt	atcatcgttaggtcagggcggttataaataagtcacg	N297ARt	tcagaaacagagaagcgtgacggttcttctgtg
T180Fs	acatgacggacggatgga	C351Fs	ggtagcggataaggctatctc
T180Rs	tgagcttaccgcaacatc	C351Rs	agggctctgttcttaaccaag
T180AFt	agcaggctgtggcaactacatgacggacggatgga	C351AFt	gcatgctgtactcagaggtgacggataaggctatctc
T180ARt	agttgccaccagctgctgagcttaccgcaacatc	C351ARt	tcgtgagtagcagcatgacggacatcttcttaaccaag
G182AFt	agcagaccggtgtaactacatgacggacggatgga	C351SFt	gcattctgactcagaggtgacggataaggctatctc
G182ARt	agttagcaccgctgctgagcttaccgcaacatc	C351SRt	tcgtgagtagcagaaatgacggacatcttcttaaccaag

**Table 1. Primers used for PPAD single-point mutagenesis.** Restriction enzyme recognition sequences are underlined.

resolution from different protein preparations, which crystallized in different space groups: substrate-free (to 1.5 Å resolution), a substrate-mimic (1.4 Å), and a substrate complex (1.8 Å; see Table 2).

The PPAD two-domain moiety (CD *plus* IgSF; Fig. 1) shows approximate maximal dimensions of 55 Å (height) × 57 Å (width) × 50 Å (depth) according to the orientation of Fig. 1a and lacks any bound calcium ion, thus explaining why it is not needed for activity. Overall, it resembles a tooth—with the 316-residue CD featuring the crown and IgSF the root—which is reminiscent of the gross overall shape of Kgp and RgpB despite completely different functions and CD architectures (see Fig. 2b in<sup>22</sup> and Fig. 2a in<sup>23</sup>). The neck is the interface between the two domains, and the active site is at the cusp, on the grinding surface (see below). The CD (A<sup>44</sup>-K<sup>359</sup>; see Fig. 1a–c) comprises eight helices and 20 β-strands and is a flat cylinder made up by a distorted five-fold α/β-propeller arranged around a central shaft. The PPAD CD cylinder has an upper entry base, which coincides with the tooth cusp, and an opposite lower exit base at the neck (Fig. 1a). Around the shaft, five propeller blades (I to V) spanning between 47 (blade III) and 76 (blade I) residues are sequentially arranged counterclockwise according to Fig. 1b,c. Each blade starts on the entry base with a loop connected to the previous blade and consists at least of a three-stranded twisted β-sheet with an inner, a middle and an outer strand, *plus* one helix. The inner strand runs across the cylinder to the exit base paralleling the central shaft. A short loop links the inner strand with the antiparallel middle strand, which runs in the opposite direction towards the entry base. This strand is connected through another loop with the helix, which lines the cylinder side wall. Finally, the helix is linked to the outer strand, which parallels the middle strand and likewise lines the cylinder side wall. Into this minimal architecture—found only in blade V (Fig. 1c)—, additional structural elements are inserted in each blade, thus accounting for overall blade asymmetry and chain lengths. In particular, a sodium ion is pinched by the inner strand and the consensus helix of blade II and is bound in an octahedral manner by six oxygens at distances of 2.30–2.63 Å: D<sup>148</sup>O, D<sup>158</sup>O, and two solvent molecules coplanar with the cation; and apically by D<sup>147</sup>Oδ1 and D<sup>158</sup>Oδ1.

Preceding the first blade, an N-terminal extension (A<sup>44</sup>-R<sup>63</sup>) is found attached to blade II on the cylinder side wall running from the entry base to the exit base (Fig. 1a–c). Here, the polypeptide undergoes

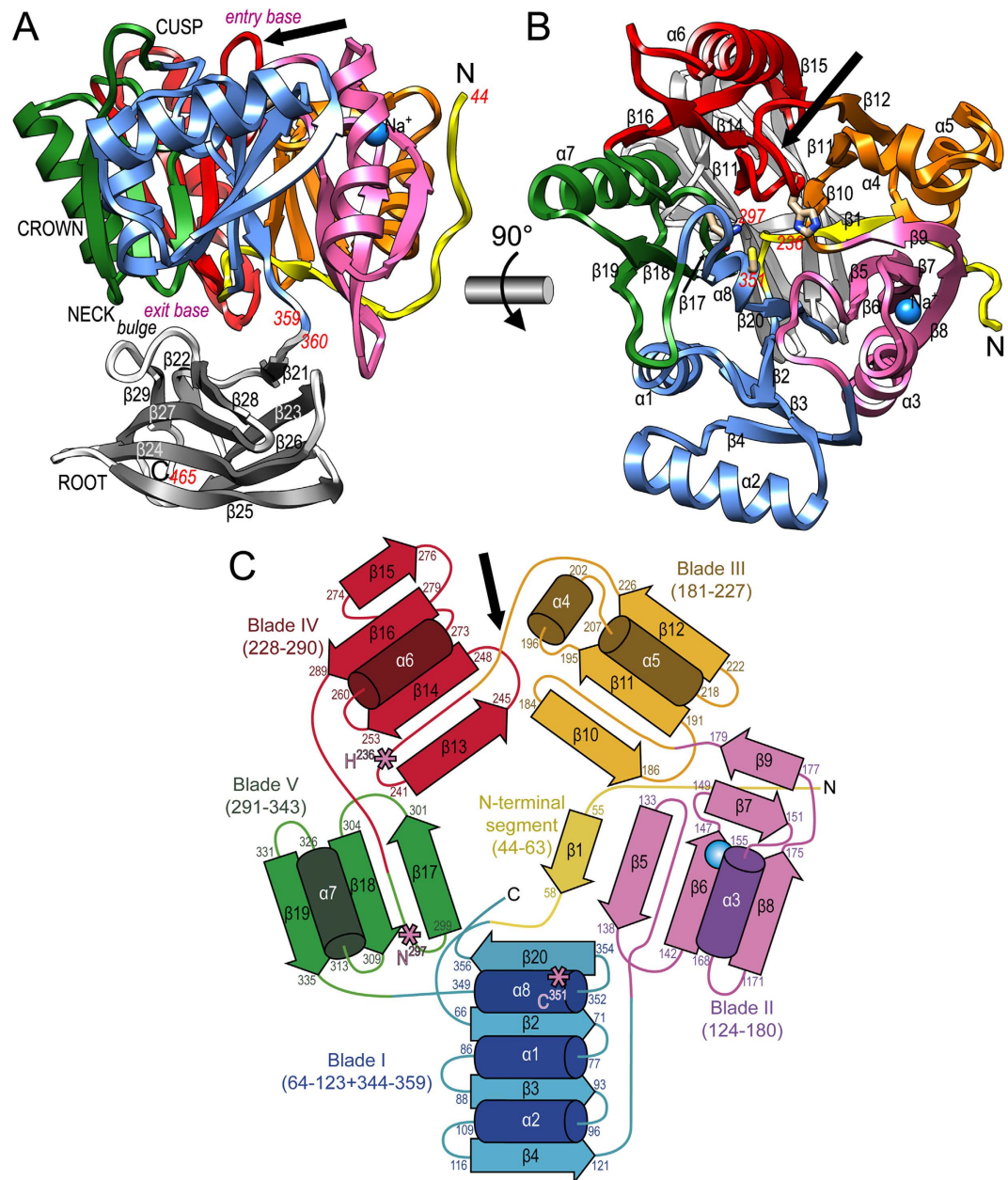
Dataset	PPAD-TP (substrate free)	PPAD(C <sup>351</sup> A)+M-R (substrate complex)	PPAD+D-Q (substrate-mimic complex)
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	C2	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell constants (a, b, c, in Å; β in °)	58.56, 60.30, 113.68, 90.00	105.36, 59.32, 84.61, 126.60	60.53, 71.31, 105.66, 90.00
Wavelength (Å)	0.9786	0.9795	0.9795
No. of measurements/unique reflections	288,438/64,233	254,674/38,882	1,133,182/90,283
Resolution range (Å) (outermost shell) <sup>a</sup>	42.0–1.50 (1.59–1.50)	67.9–1.80 (1.90–1.80)	46.1–1.40 (1.48–1.40)
Completeness (%)	98.5 (95.7)	99.3 (98.9)	99.6 (97.5)
R <sub>merge</sub> <sup>b</sup>	0.041 (0.190)	0.062 (0.529)	0.042 (0.305)
R <sub>crim</sub> [=R <sub>meas</sub> ] <sup>c</sup> / CC(1/2) <sup>c</sup>	0.047 (0.217)/0.998 (0.973)	0.067 (0.584)/0.999 (0.927)	0.043 (0.321)/1.000 (0.973)
Average intensity <sup>d</sup>	21.0 (7.4)	22.5 (3.7)	36.3 (8.5)
B-Factor (Wilson) (Å <sup>2</sup> )/Aver. multiplicity	25.0/4.5 (4.1)	29.2/6.5 (5.6)	19.3/12.6 (10.1)
Resolution range used for refinement (Å)	42.0–1.50	48.6–1.80	46.2–1.40
No. of reflections used (test set)	64,233 (772)	38,882 (699)	90,283 (903)
Crystallographic R <sub>factor</sub> (free R <sub>factor</sub> ) <sup>b</sup>	0.157 (0.177)	0.156 (0.182)	0.146 (0.149)
No. of protein atoms/solvent molecules/neutral (covalent) ligands/ionic ligands	3,321/460/7 glycerol, 3 p-thiopyridine, 1 S-oxo/1 Na <sup>+</sup>	3,296/426/5 glycerol, 1 Cys, 1 Met-Arg dipept./1 Na <sup>+</sup> , 2 N <sub>3</sub> <sup>-</sup> , 1 Cl <sup>-</sup>	3,300/689/3 glycerol, 1 imidazole, 1 Asp-Gln dipept./1 Na <sup>+</sup> , 1 N <sub>3</sub> <sup>-</sup> , 1 Cl <sup>-</sup> , 5 PO <sub>4</sub> <sup>3-</sup>
Rmsd from target values <sup>e</sup>			
bonds (Å)/angles (°)	0.010/1.03	0.010/0.99	0.010/1.05
Average B-factors (Å <sup>2</sup> )	24.4	28.6	17.4
All-atom contacts and geometry analysis <sup>f</sup>			
Residues			
in favored regions/outliers/all residues	409 (97.6%)/0/419	401 (97.1%)/0/413	411 (97.9%)/0/420
with poor rotamers/bad bonds/bad angles	2 (0.56%)/0/0	3 (0.85%)/0/1 (0.24%)	2 (0.56%)/0/0
with Cβ deviations >0.25 Å/ clashscore	0/3.62 (96 <sup>th</sup> percentile)	1/3.34 (98 <sup>th</sup> percentile)	0/3.47 (96 <sup>th</sup> percentile)
MolProbity score	1.23 (97 <sup>th</sup> percentile)	1.29 (98 <sup>th</sup> percentile)	1.17 (97 <sup>th</sup> percentile)

**Table 2. Crystallographic data.** <sup>a</sup>Data processing values in parenthesis refer to the outermost resolution shell. <sup>b</sup>For definitions, see Table 1 in<sup>55</sup>. <sup>c</sup>For definitions, see<sup>56,57</sup>. <sup>d</sup>According to the XDS program<sup>40</sup>. Average intensity is  $\langle I/\sigma(I) \rangle$  of unique reflections after merging. <sup>e</sup>According to Engh and Huber<sup>58</sup>. <sup>f</sup>According to MOLPROBITY<sup>53,59</sup>.

a kink and, paralleling the inner strand of blade II, runs along the exit base between blades II and III until the central shaft. There, it runs upward as the middle strand of blade I. The C-terminal segment after blade V enters blade I and provides an extra helix followed by the inner strand of the consensus topology, thus internally fastening the molecule like a Velcro strip. Thereafter, the polypeptide reaches the exit base of the CD and enters the C-terminal 106-residue IgSF domain.

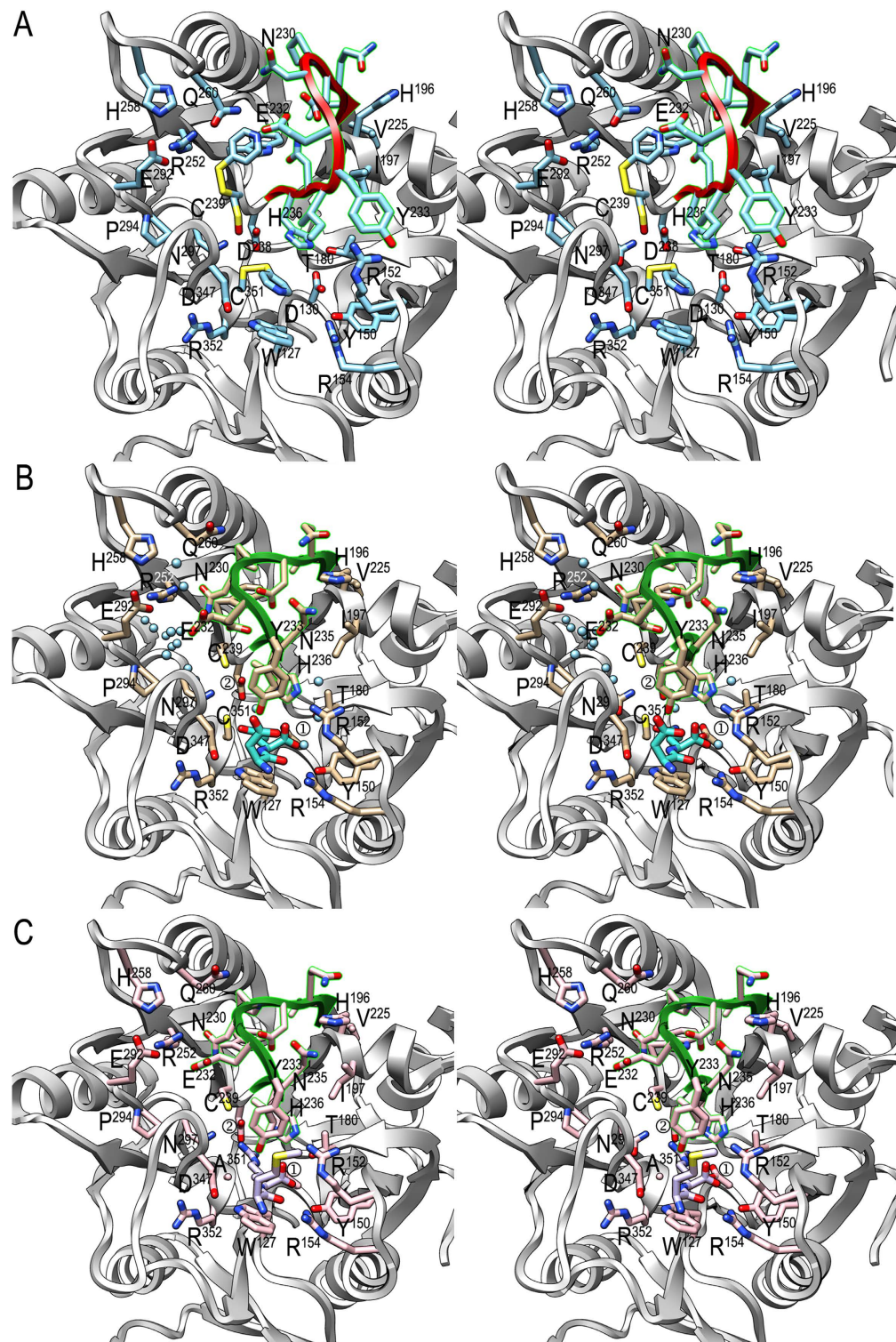
The IgSF domain (G<sup>360</sup>-E<sup>465</sup>) is a distorted 4 + 5-stranded β-sandwich (strands β21↓-β29) with an antiparallel back sheet (β21↓-β23↑-β26↓-β25↑) and a mixed front sheet (β22↓-β28 + β29↓-β27↑-β24↓-β25↑) whose planes are rotated away by ~25°. The right lateral flank of the domain is closed by strand β25, whose N-terminal and C-terminal halves participate, respectively, in the front and back sheets. The left lateral flank is much wider and open, and contains a bulge dividing the second strand from the left of the front sheet in two (β28 and β29). This bulge interacts with the exit base of the CD (Fig. 1a). Overall, the topology and strand-connectivity of PPAD IgSF is strongly reminiscent of that of Kgp and RgpB<sup>22-24</sup>, but while the width (~25 Å) and depth (~20 Å) of the domains are similar, the length—along the strands of the sheets—is much greater in PPAD than in gingipains (~50 Å vs. ~35 Å).

**Active site of PPAD.** The propeller shaft in PPAD is rather solid, with a shallow cavity on its entry base coinciding with the tooth cusp that contains the active site (Fig. 1a,b). The latter is mainly a narrow funnel-like hole, which accommodates an arginine side chain of a peptidic or protein substrate. It is



**Figure 1. Overall structure and topology of PPAD.** (A) Ribbon-type plot of PPAD in a lateral view revealing its tooth-like shape, which consists of regions assignable to cusp, crown, neck and root. The upper N-terminal cylindrical catalytic domain (CD; residues 44–359; top entry base and bottom exit base) is shown with the N-terminal segment in yellow and each of its constituting blades (I to V) in one color (blue, magenta, orange, red, and green). The C-terminal IgSF-like domain (residues 360–465) is shown in grey for its  $\beta$ -strands (labeled  $\beta$ 22– $\beta$ 29) and white for loops and coils. A sodium ion is shown as a blue sphere and a black arrow pinpoints the Michaelis-loop. (B) Top view onto the entry base of the CD cylinder after a horizontal 90°-rotation of (A). The helices ( $\alpha$ 1– $\alpha$ 8) and strands ( $\beta$ 1– $\beta$ 20) of the CD are labeled. Catalytic-triad-residue (C<sup>351</sup>, H<sup>236</sup> and N<sup>297</sup>) side chains are shown and labeled in red to highlight the active site in the center of the  $\alpha/\beta$ -propeller. A black arrow pinpoints the Michaelis-loop. (C) Topology scheme of the five-bladed PPAD CD with strands as arrows and helices as cylinders with their respective limiting residues; coloring as in panels (A) and (B). The three catalytic residues of (B) are shown as pink asterisks, and the Michaelis-loop is denoted by a black arrow.

framed by the main chain and side chains of the loops connecting blades I and II, II and III, III and IV, and V and I; segment  $\beta$ 7-loop  $\beta$ 7 $\alpha$ 3– $\alpha$ 3 of blade II; and helix  $\alpha$ 8 of blade I (Figs 1b and 2a–c). In the substrate-free structure, which was obtained with DTD-treated wild-type (wt) protein, catalytic C<sup>351</sup>, nearby C<sup>239</sup> and distal C<sup>462</sup> residues are covalently modified by what was conservatively interpreted as



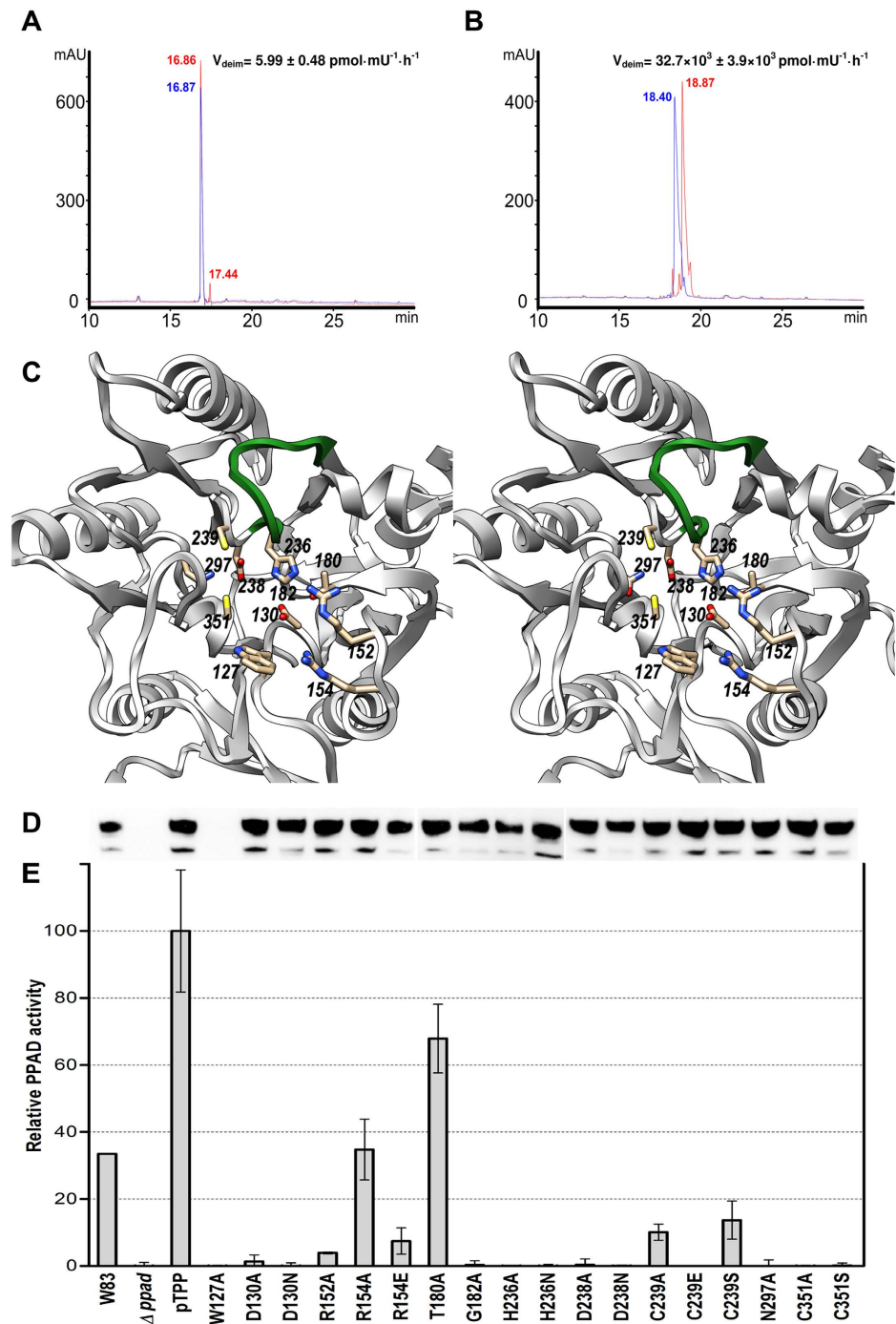
**Figure 2. Active-site architecture.** (A) Stereo image of substrate-free PPAD CD, which actually corresponds to a thiopyridine modified state, with the Michaelis-loop (V<sup>226</sup>-V<sup>237</sup>) shown in red. Selected side chains are displayed with their carbons in light blue and labeled. (B) Same view as in (A) of the substrate-mimic complex, with the Michaelis-loop in green, unmodified cysteines, side-chain carbons in tan and relevant solvent molecules to illustrate the NH<sub>3</sub>-exit/H<sub>2</sub>O-entry and hydroxide channels as spheres in light blue (see also Fig. 4a). The bound aspartate-glutamine dipeptide is further shown with its carbons in turquoise. ① labels D<sup>130</sup> and ② labels D<sup>238</sup>. (C) Same view as in (A) and (B) of the substrate complex, with the Michaelis-loop in green, unmodified cysteine C<sup>239</sup> (C<sup>351</sup> is replaced by alanine), and side-chain carbons in pink. The bound methionine-arginine dipeptide is further depicted with its carbons in purple. ① labels D<sup>130</sup> and ② labels D<sup>238</sup>.

a 4-thiopyridyl moiety. The C<sup>239</sup> side chain is even found in two alternate conformations, one bound to thiopyridine and the other with the sulfur as sulfoxide (Fig. 2a). This indicates overall flexibility of active site residues in PPAD due to the absence of a bound substrate and suggests that the covalent modifications of the S<sub>γ</sub> atoms do not distort the general unbound conformation of PPAD. In addition, segment V<sup>226</sup>-V<sup>237</sup> of the loop connecting blades III and IV, hereafter the “Michaelis loop”, is in an open conformation, thus consistent with a structure that can bind a substrate. In particular, Y<sup>233</sup> at the most exposed part of the loop points to bulk solvent (Fig. 2a). We further obtained a substrate-mimic complex of DTDP-untreated wt PPAD with dipeptide aspartate-glutamine, and a true substrate complex of DTDP-untreated PPAD-C<sup>351</sup>A with dipeptide methionine-arginine. The identification of the peptides was based on high-resolution Fourier maps and surrounding binding partners. The complexes were obtained serendipitously, and all attempts to obtain complexes with other substrates or products failed. We hypothesize that DTDP-treatment precludes substrate binding and, thus, protects the unbound conformation, while lack of such treatment causes the enzyme to trap substrates or mimics during biosynthesis or purification. The complex structures are equivalent, including the backbone of the bound dipeptides (Fig. 2b,c), except for some minimal displacement and the differing side chains (Fig. 2b,c), so the substrate complex is taken hereafter as reference except for issues dealing with C<sup>351</sup>S<sub>γ</sub>, for which the substrate-mimic complex will be referred to.

The complex structures allowed us to identify PPAD elements required for substrate binding and catalysis. Comparison with the substrate-free structure revealed overall coincidence of the complexes except for the rearrangement of the Michaelis loop (maximal displacement 7.5 Å at N<sup>230</sup>Cα), which adopts a closed conformation that traps the substrate arginine side chain (Fig. 2b). This causes H<sup>196</sup> to be rotated ~100° around its χ<sub>1</sub> angle toward bulk solvent and Y<sup>233</sup> to be displaced by 4.1 Å and slightly reoriented for its side chain to bind the substrate (see below). Michaelis-loop rearrangement further causes ~90° rotation of H<sup>236</sup> around its χ<sub>2</sub> angle, so that its Nδ1 atom is apical to the guanidinium plane (3.2 Å away from arginine C<sub>ζ</sub> atom; see below) and may play a role in catalysis (see below). Catalytic C<sup>351</sup>S<sub>γ</sub>, at the bottom of the cleft, occupies the opposite apical position and the atom is further in binding distance from N<sup>297</sup>Oδ1 (3.3 Å), which could potentially assist in catalysis (see below). N<sup>297</sup>Nδ2, in turn, is in binding distance of D<sup>238</sup>Oδ1 (3.3 Å). The guanidinium group is further tightly bound by D<sup>238</sup> through a double salt bridge with arginine Nη1 and Nη2 atoms (2.9 Å and 3.0 Å), by the main-chain carbonyl of T<sup>346</sup> (3.2 Å away from atom Nη1), and by D<sup>130</sup> through a second double salt bridge with arginine Nε and Nη2 atoms (2.8 Å and 2.9 Å). D<sup>130</sup> becomes rotated around its χ<sub>1</sub> angle by ~60° upon substrate binding, thereby exchanging its tight hydrogen bond with T<sup>180</sup>Oγ1 (2.6 Å) in the substrate-free structure with binding of the substrate guanidinium group. These five interactions of the guanidinium group occur roughly in the plane of the latter. The aliphatic part of the arginine is bound between the hydrophobic side chains of I<sup>234</sup> and W<sup>127</sup> (both 3.7 Å apart). The latter is held in place by a hydrogen bond between its Nε1 atom and D<sup>347</sup>Oδ1 (2.9 Å), which also confers to the tryptophan a potential role in overall structure maintenance due to its stabilizing function of the loop connecting blades V and I (see below). Interestingly, two small solvent-accessible channels are found roughly on either side of the guanidinium plane, on the right and the left in Fig. 2b. The left channel, hereafter “NH<sub>3</sub>-exit/H<sub>2</sub>O-entry channel,” is framed by segments T<sup>290</sup>-N<sup>297</sup>, N<sup>230</sup>-E<sup>232</sup>, G<sup>345</sup>-T<sup>346</sup>, R<sup>252</sup>, H<sup>258</sup>, and, in particular, C<sup>239</sup>, which is closest to the substrate guanidinium and thus acts as a gatekeeper of the channel. The right channel, in turn, is shallower and does not reach the substrate but rather H<sup>236</sup>Nε2, which is bound to two solvent molecules (see below). This “hydroxide-entry channel” is framed by Y<sup>233</sup>-N<sup>235</sup>, N<sup>151</sup>-R<sup>152</sup>, I<sup>197</sup>, and E<sup>201</sup>.

On the outer border of the active-site cavity, the main chain of the substrate is tightly bound through six interactions. The C-terminal carboxylate is linked by a double salt bridge with R<sup>152</sup>Nη2 (3.0 Å) and Nε (2.8 Å). In addition, one of the carboxylate oxygens is further bound by R<sup>154</sup>Nη1 (2.9 Å) and the other by Y<sup>233</sup>Oη (2.8 Å). The latter atom also binds the main-chain amido nitrogen (3.4 Å), and the preceding peptide carbonyl is hydrogen-bonded by R<sup>154</sup>Nη2 (2.7 Å). This interaction seems to be the main factor responsible for the selectivity of PPAD for peptidylarginines over free arginine<sup>18</sup>. In addition, these interactions draw an intricate network to fix the substrate in the cleft, which makes it difficult to imagine how a substrate with C-terminal extension to the arginine, i.e. an endodeiminase substrate, would be bound by PPAD, as a C-terminally extended peptide would collide with Y<sup>233</sup> and R<sup>152</sup> side chains (Fig. 2b,c). Finally, lack of specific interactions with atoms upstream of the last peptide bond of the substrate accounts for PPAD's capacity to non-specifically turn over both peptides and proteins with C-terminal arginines, i.e. as long as the C-terminus is freely accessible.

**Peptidylarginine deiminase activity and mutant studies *in vitro*.** PPAD is an efficient deiminase of peptides including bradykinin and benzoylglucylarginine<sup>18</sup>, EGF and anaphylatoxin C5a<sup>20</sup>, and Rgp-derived fibrinogen peptides, as well as a large set of bacterial cell-envelope proteins truncated by Rgps. To provide additional data on the endo- and exodeiminase activities of PPAD *in vitro*, we tested two octapeptides of equivalent charge derived from the physiologically-relevant bradykinin precursor sequence, respectively with an arginine at position six (G-F-S-P-F-R-S-S; Fig. 3a) and at the C-terminus (P-P-G-F-S-P-F-R; Fig. 3b). We found that peptidylarginine exodeiminase activity of PPAD was nearly 5,500 times higher than endodeiminase activity. This supports the structural findings above. In addition, detailed inspection of the final refined Fourier maps and thermal displacement parameters of atoms Nη1, C<sub>ζ</sub>, and Nη2 of all twelve internal arginines of the substrate-mimic complex of PPAD, which was refined



**Figure 3. PPAD activity assays.** (A) Endo- and (B) exo-deiminase activity assays *in vitro* of *P. gingivalis* W83 wt PPAD against peptides of sequence G-F-S-P-F-R-S-S and P-P-G-F-S-P-F-R, respectively. Peptides are shown before (blue HPLC chromatograms) and after reaction with PPAD (red HPLC chromatograms). Citrullination caused a shift in the retention time of the peptides when compared with the original ones and was confirmed by mass spectrometry. Based on peak integration, the velocity of reaction was calculated for both peptides, which indicated that peptidylarginine exodeiminase activity of PPAD was nearly 5,500 times higher than endodeiminase activity based on reaction velocity ( $32,700$  vs.  $6 \text{ pmol} \cdot \text{mU}^{-1} \cdot \text{h}^{-1}$ ). (C) Stereo image depicting the 11 positions subjected to point mutagenesis and activity measurements (see (C) and (D)). The Michaelis-loop is shown in green for reference. (D) PPAD expression monitoring through Western-blot analysis of whole bacterial cultures resolved on SDS-PAGE and probed with an anti-PPAD antibody. The samples correspond to those of the abscissa of panel (E). (E) Relative deiminase activity in front of *N*-acetylarginine of wt W83 strain supernatant (W83), of a PPAD-deletion mutant strain ( $\Delta ppad$ ), of the latter containing plasmid pTPP for wt PPAD overexpression (pTPP; reference 100%), and a cohort of single point mutants around the active site encoded by pTPP variants.



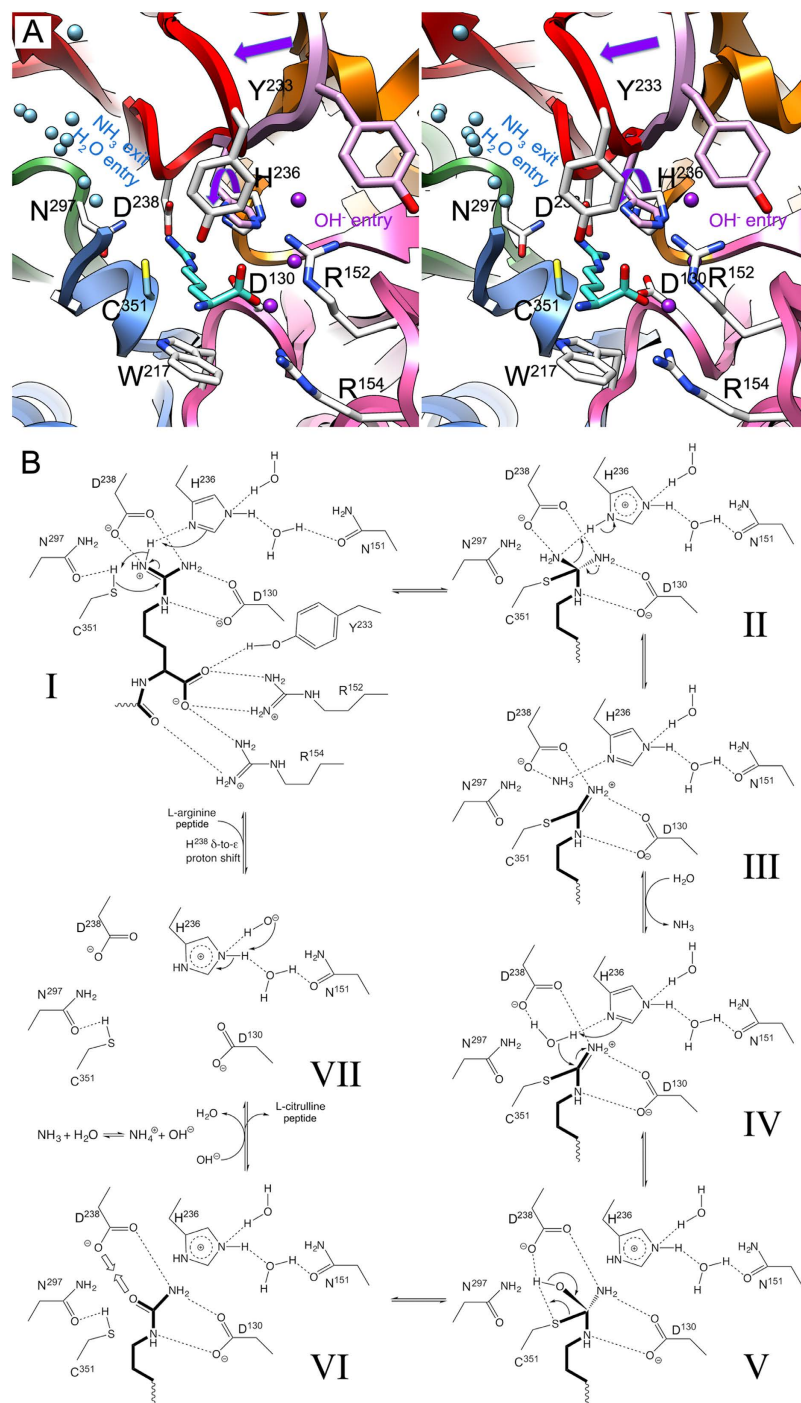
with data to very high resolution (1.4 Å; see Table 2), revealed no significant evidence for citrullination, strongly suggesting that PPAD produced by homologous overexpression in *P. gingivalis* is not endocitrullinated. Taken together, all these findings strongly support that PPAD is an exodeiminase, as already suggested in the initial report in 1999<sup>18</sup>, and that N-terminal arginines of peptides, endosubstrates and standalone arginines are only modified at a much lower rate, if at all<sup>18</sup>.

In order to discern the functional role of the distinct residues identified in the structures above, we constructed a cohort of 18 single-point mutants of positions 127, 130, 152, 154, 180, 182, 236, 238, 239, 297 and 351 (Table 1 and Fig. 3c) and assessed the deiminase activity of the respective cell cultures relative to the wt. Difficulties in the production of wt and mutant PPADs, which were obtained from *P. gingivalis* cultures, precluded more extensive enzymatic analyses with purified protein. Mutant expression levels were equivalent to those of the wt as monitored by Western-blot analysis, thus pointing to properly folded proteins. The sole exception was W<sup>127</sup>A, which in accordance with a structural role in addition to a substrate-binding role (see above), was not produced in detectable amounts (Fig. 3d). As expected, activity was completely abolished when mutating catalytic C<sup>351</sup>—to either alanine or serine—, but also when replacing D<sup>238</sup> or H<sup>236</sup>—to either alanine or asparagine—, which participate in substrate guanidinium C $\zeta$  atom pinching (Fig. 3e). N<sup>297</sup>, in binding distance of C<sup>351</sup>S $\gamma$ , likewise yielded an inactive enzyme when replaced with alanine. D<sup>130</sup>, which strongly binds the guanidinium, is also indispensable, and C<sup>239</sup>, the gatekeeper of one of the two solvent channels, is also relevant as its alanine and serine mutants were just ~8% active and its glutamate mutant was completely inactive. G<sup>182</sup>, in turn, is required to be side-chain depleted as it shuts the bottom of the pocket and is close to H<sup>236</sup> and D<sup>130</sup>. Its replacement with alanine yielded a complete loss of activity. In contrast, T<sup>180</sup>, which interacts with the two latter residues, is unessential and its alanine mutant still had ~66% activity. Interestingly, R<sup>152</sup>, which establishes a double salt bridge with the substrate carboxylate, is absolutely indispensable for activity, while the second carboxylate-binding arginine, R<sup>154</sup>, is less relevant, with its alanine mutant still showing ~30% activity. Its glutamate mutant, however, which introduces a negative charge next to the also negatively-charged substrate carboxylate, thus causing repulsion, was less than ~10% active.

**Mechanism of peptide citrullination by PPAD.** We propose the following chemical mechanism of function of PPAD, which includes a catalytic triad (C<sup>351</sup>-H<sup>236</sup>-N<sup>297</sup>) and seven steps proceeding over two tetrahedral and one planar-thiouonium covalent reaction intermediates (Fig. 4a,b).

In the substrate-free state, the Michaelis-loop containing Y<sup>233</sup> is in an open conformation, which enables peptides with a C-terminal arginine to be accommodated at the active site. The arginine becomes firmly anchored through electrostatic interactions of the guanidinium group with the side chains of D<sup>238</sup> and D<sup>130</sup>, being positioned in an extended conformation and appropriately oriented for catalysis. In addition, R<sup>152</sup> and R<sup>154</sup> bind the C-terminal carboxylate of the arginine and the carbonyl of the preceding peptide bond. Moreover, formation of this Michaelis complex (Fig. 4b, I) entails major rearrangement of the Michaelis loop, which occludes the active site and causes Y<sup>233</sup> to further bind the C-terminal carboxylate of the substrate. Rearrangement further entails that the side chain of H<sup>236</sup> is rotated, as a result of which the plane of the guanidinium group becomes pinched between H<sup>236</sup>N $\delta$ 1 and C<sup>351</sup>S $\gamma$ , and H<sup>236</sup>N $\epsilon$ 2 is solvent-bound in the hydroxide-entry channel (Fig. 4a,b). This geometry was determinant for the identification of H<sup>236</sup> as the general base/acid of the mechanism and of the guanidinium N $\eta$ 1 atom as the nitrogen atom of the leaving ammonia product. In addition, C<sup>351</sup>S $\gamma$  is hydrogen-bonded to N<sup>297</sup>O $\delta$ 1, which probably enhances the nucleophilicity of the catalytic sulfur. In the first step of the reaction, C<sup>351</sup>S $\gamma$  performs a nucleophilic attack on the sp<sup>2</sup>-like planar C $\zeta$  atom of the substrate guanidinium (Fig. 4b, I), giving rise to the first neutral tetrahedral reaction intermediate and yielding an sp<sup>3</sup>-like C $\zeta$  atom. Concomitantly, H<sup>236</sup>, which acts first as a general base, abstracts a proton from N $\eta$ 1, and the latter captures the proton from the catalytic thiol group. The histidine is now in a diprotonated state (Fig. 4b, II). The tetrahedral intermediate collapses to a positively-charged planar thiouonium covalent intermediate and ammonia, which receives a proton from H<sup>236</sup>N $\delta$ 1, now acting as a general acid (Fig. 4b, II and III). Ammonia leaves the active site through the NH<sub>3</sub>-exit/H<sub>2</sub>O-entry channel (Fig. 4a) and reaches the surface of the enzyme. In the next step, a solvent molecule—probably a water—occupies the former position of ammonia and becomes polarized by the side chain of D<sup>238</sup> and H<sup>236</sup>N $\delta$ 1. The latter again acts as a base and abstracts a proton from the water molecule, which performs a nucleophilic attack on the central carbon of the thiouonium (Fig. 4b, IV). This yields the second neutral intermediate centered on sp<sup>3</sup>-like tetrahedral C $\zeta$  and diprotonated H<sup>236</sup> (Fig. 4b, V). The intermediate itself collapses to a citrullinated product and the intact catalytic cysteine mercapto group, which becomes hydrogen-bonded to N<sup>297</sup>O $\delta$ 1. The repulsion between D<sup>238</sup> and the carbonyl oxygen of the neutral reaction product may provide the driving force for clearance of the latter from the active-site cleft (Fig. 4b, VI). Finally, a hydroxide resulting from the reaction of ammonia with water may enter the active site through the hydroxide-entry channel and replace one of the two solvent molecules bound to H<sup>236</sup>N $\epsilon$ 1. The latter histidine transfers a proton to the hydroxide and a proton shift from N $\delta$ 1 to N $\epsilon$ 2 restores the functional monoprotonated state of H<sup>236</sup>, thus leaving the active site posed for a new round of reaction (Fig. 4b, VII).

**Structural similarity of PPAD catalytic domain.** PPAD CD conforms to the structural requirements of the guanidino-group modifying enzyme superfamily (GME; see Fig. 5a–c), which adopts



**Figure 4. Proposed peptide citrullinating mechanism of PPAD.** (A) Composite picture in stereo of the active site of PPAD (see also Fig. 2) based on the substrate-mimic complex ribbon plot colored as in Fig. 2b. Only elements engaged in substrate binding and catalysis are depicted. Residue side chains taken from the substrate-mimic complex are shown with carbons in light blue (C<sup>351</sup>), those from the substrate complex in white (Y<sup>233</sup>, H<sup>236</sup>, D<sup>238</sup>, N<sup>297</sup>, R<sup>152</sup>, R<sup>154</sup>, and W<sup>217</sup>), and those from the unbound structure in pink (Y<sup>233</sup> and H<sup>236</sup>). The Michaelis loop is shown in the open conformation of the unbound structure in pink and in the occluded conformation of the substrate(-mimic) complexes in red, a purple straight arrow highlights the rearrangement upon substrate binding. The substrate arginine depicted belongs to the substrate complex (carbons in turquoise). Solvent molecules from the substrate-mimic complex in light blue highlight the NH<sub>3</sub>-exit/H<sub>2</sub>O-entry channel on the left and those in purple the hydroxide-entry channel on the right. The rotation of the H<sup>236</sup> side chain from the substrate-unbound to the bound conformation is pinpointed by a curved purple arrow. (B) Proposed biochemical mechanism of action of an enzymatic activity cycle in seven steps (I to VII). The substrate arginine and product citrulline are shown with bonds in bold, hydrogen bonds are shown as dashed lines.

similar five-fold  $\alpha/\beta$ -propeller folds and catalyzes chemical processing of (methylated) guanidine groups as found in the citrullinating GME members: PADs, PPAD, agmatine deiminases (AgDIs), and arginine deiminases (ADIs), which are all dimers or tetramers with the exception of PPAD<sup>25</sup>. AgDIs deiminate isolated agmatine (1-[4-aminobutyl]-guanidine) to *N*-carbamoylputrescine and ammonia as part of mechanisms by which energy is harnessed for growth<sup>26</sup>, and they are missing in higher eukaryotes<sup>25</sup>. ADIs, in turn, citrullinate standalone arginine and protect cells from acidic environments. They are found in plants and microorganisms but are likewise absent from animals<sup>27</sup>. Both families do not have extra domains further to the catalytic  $\alpha/\beta$ -propeller.

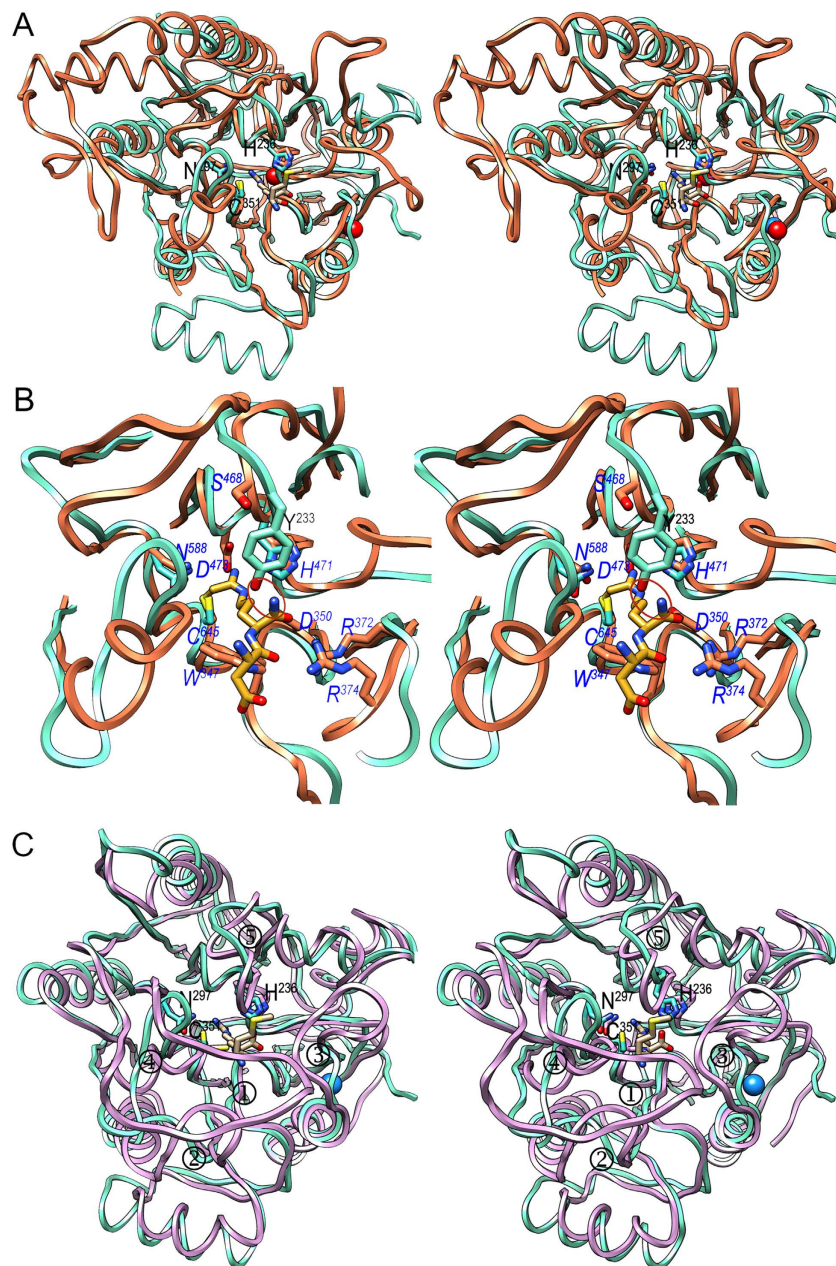
To date, only the structures of human PAD2 and PAD4 have been determined among PADs<sup>8,28</sup>, and they comprise a ~375-residue calcium-dependent  $\alpha/\beta$ -propeller domain preceded by two IgG domains (see Fig. 1 in<sup>8</sup> and<sup>28</sup>), which are unrelated to PPAD IgSF further to being all- $\beta$  protein domains. Among AgDIs, structures have been reported from *Enterococcus faecalis* (*Ef*; Protein Data Bank (PDB) access code 2JER<sup>26</sup>) and *Helicobacter pylori* (*Hp*; PDB 3HVM<sup>29</sup>), and other potential relatives have been deposited with the PDB but not functionally analyzed or published (PDB 2EWO, 1XKN, 1ZBR, 2CMU, 3H7C and 1VKP). Finally, ADI structures have been reported from *Streptococcus pyogenes* (*Sp*; PDB 4BOF<sup>30</sup>), *Pseudomonas aeruginosa* (*Pa*; PDB 1RXX<sup>31</sup>), and *Mycoplasma arginini* (*Ma*; PDB 1S9R<sup>32</sup>). Among all these, closest structural similarity of PPAD is found with AgDIs (Z-score of 35 according to program DALI<sup>33</sup>; see Fig. 5c), followed by PADs (Z = 18–21; Fig. 5a,b) and ADIs (Z = 18–19).

Superposition of the PPAD  $\alpha/\beta$ -propeller on that of human PAD4 (Fig. 5a,b), *Ef*AgDI (Fig. 5c), and *Pa*ADI, *Sp*ADI and *Ma*ADI (data not shown) reveals good overall conservation of the five-blade architectures, although several decorations in the distinct blades of each family account for large differences, especially in the loops surrounding the active-site cleft. In particular, PADs evince a large partially helical insertion between  $\beta$ 14 and  $\alpha$ 6 of PPAD blade IV and lack  $\alpha$ 2 of blade I (Fig. 5a). ADIs, in turn, evince a large helical sub-domain replacing  $\alpha$ 2 and  $\beta$ 4 of PPAD blade I. In common, all propellers are closed by the blade V-blade I Velcro mechanism (see above and<sup>25</sup>) and the catalytic cysteines and histidines are conserved, as well as the two aspartates anchoring the guanidine group to the bottom of the active site. In addition, PPAD shares with ADIs and PADs the two arginines binding the main chain of the substrate. While these firmly bind the substrate C-terminus in PPAD and ADIs, in PADs they are slightly reoriented and only bind what would be one of the two carboxylate oxygens in addition to the upstream peptide carbonyl (Fig. 5b). This, together with the replacement of PPAD Y<sup>233</sup> of the Michaelis loop by serine (S<sup>468</sup>; PAD4; residue numbering of proteins distinct from PPAD in italics) or threonine (T<sup>468</sup>; PAD2), provides enough space in PADs to allow for a C-terminal extension of the substrate. Furthermore, calcium-dependence of PADs is characterized by several calcium-binding sites<sup>8,28</sup>, two of which occur within the propeller domain: one close to the active site with evident implications for function and the other at the domain periphery (Fig. 5a). Interestingly, the latter coincides with the sodium site of PPAD, so a predominantly structural role for both is suggested. In contrast to PADs and PPAD, AgDIs and ADIs, which only process standalone residues, possess completely closed active sites (Fig. 5c).

Most notably, superposition also revealed that all these families possess an equivalent of PPAD asparagine N<sup>297</sup>, i.e. with a potential role in catalysis (PAD2, N<sup>590</sup>; PAD4, N<sup>588</sup>; *Ef*AgDI, N<sup>306</sup>; *Hp*AgDI, N<sup>274</sup>; *Pa*ADI, N<sup>360</sup>; *Ma*ADI, N<sup>352</sup>; and *Sp*ADI, N<sup>355</sup>). To our knowledge, this was previously unnoticed since this residue, which is strictly conserved across citrullinating GMEs, was merely recognized as an important residue for proper active-site conformation conserved in the consensus helix of blade V of all families (see Fig. 4 in<sup>25</sup>). In PADs, this asparagine is also conserved in distant orthologs from zebrafish and chicken within a shared motif (M/L-V-N-M<sup>34</sup>), which complements the consensus motif encompassing the catalytic cysteine residue (G-E-I/V-H-C-G-T/S). The only notable exception is human PAD6, which lacks both motifs and the calcium sites that are essential for activity in the other paralogs and orthologs<sup>34</sup>. This absence, together with the lack of direct evidence for activity *in vitro* with the assays routinely employed for the other PADs, poses the question as to whether PAD6 is an active peptidylarginine deiminase or whether it may require further factors or interacting partners for activity<sup>35</sup>. In any case, it is likely to follow a different catalytic mechanism.

In all the above structures, the asparagine is at suitable distances and in appropriate orientations to polarize the catalytic cysteine, as found in papain-like cysteine peptidases—in particular, Kgp and RgpB have N<sup>510</sup>—so we suggest that citrullinating GMEs all have a cysteine-histidine-asparagine catalytic triad as shown for PPAD (see above). However, in contrast to cysteine peptidases, the three residues do not establish a charge-relay system for proton transfer, but rather cysteine-asparagine and histidine act separately on opposite faces of the plane of the guanidinium (Fig. 4a,b).

**Concluding remarks.** Structural considerations identified PPAD as a closer relative of AgDIs, which are found across bacteria, than of PADs, which are found only in vertebrates. This, in turn, enables us to hypothesize that PPAD was acquired through horizontal gene transfer of a bacterial single-domain agmatine-citrullinating enzyme. The latter would then have evolved in a different bacterial environment under fusion to two new C-terminal domains like those found in cognate RgpB, to be secreted through a distinct system. This evolution further yielded a unique function among citrullinating enzymes: deimination of peptides with a C-terminal arginine. This activity, which complements that of R-type gingipain



**Figure 5. Structural similarities.** (A) Superposed ribbon-plots in stereo of PPAD in its substrate-mimic complex (cyan) and human PAD4 (coral; PDB 4DKT<sup>54</sup>) as found in its covalent thiouronium reaction intermediate mimic complex. The side chains of the respective catalytic triads (labeled for PPAD only), as well as the two calcium ions of PAD4 (red spheres) and the sodium ion of PPAD (blue sphere) are shown, as is the methionine-arginine dipeptide from the PPAD substrate complex (carbons in tan). Most loops connecting the blades and the consensus secondary elements within each blade differ in length and conformation. (B) Close-up of (A). The side chains of the catalytic triad (not labeled) and Y<sup>233</sup> (labeled in black) of PPAD are depicted (carbons in cyan), as are several representative residues from human PAD4 (carbons in coral; labeled in blue italics) and the covalently bound intermediate (carbons in goldenrod). The mechanistically-relevant equivalent positions (see Fig. 4a,b) in PPAD/human PAD4 (in italics) are C<sup>351</sup>/C<sup>645</sup>, H<sup>236</sup>/H<sup>471</sup>, N<sup>297</sup>/N<sup>588</sup>, D<sup>238</sup>/D<sup>473</sup>, D<sup>130</sup>/D<sup>350</sup>, W<sup>217</sup>/W<sup>347</sup>, Y<sup>233</sup>/S<sup>468</sup>, R<sup>152</sup>/R<sup>372</sup>, and R<sup>154</sup>/R<sup>374</sup>. A red ellipse highlights the clash an endodeiminase substrate would have with PPAD Y<sup>233</sup>. The latter is equivalent to S<sup>468</sup> in human PAD4, which allows for free space for C-terminally elongated substrates. (C) Same as (A) showing PPAD (cyan) and AgDI from *Enterococcus faecalis* (purple; PDB 2JER<sup>26</sup>) as found in a covalent adduct with an agmatine-derived amidine reaction intermediate. The respective catalytic triads are depicted and that of PPAD is also labeled. AgDI main-chain segments diverging from PPAD and mainly accounting for a closed active site are pinpointed (① to ⑤). The mechanistically-relevant equivalent positions (see Fig. 4a,b) in PPAD/AgDI (in italics) are C<sup>351</sup>/C<sup>357</sup>, H<sup>236</sup>/H<sup>218</sup>, N<sup>297</sup>/N<sup>306</sup>, D<sup>238</sup>/D<sup>220</sup>, D<sup>130</sup>/D<sup>96</sup>, and W<sup>217</sup>/W<sup>93</sup>.

virulence factors (gingipain-null mutants are devoid of endogenous citrullination), has been demonstrated for several substrates.

Pathogenic bacteria have evolved sophisticated mechanisms in response to the changing environment and host antimicrobial defense systems. Post-translational modifications are hailed as one of the main factors of pathogens to breach immune tolerance. Among these modifications, citrullination of endogenous proteins seems to be a key process in the initiation of autoimmune reactions. To date, *P. gingivalis* is the only prokaryote that is able to citrullinate proteins and peptides. It has been proposed as a mechanistic link between PD and RA through its potential capacity of generating citrullinated epitopes distinct from endogenous PADs, thus contributing to aggravation of RA. This activity is induced by the sole bacterial peptidylarginine deiminase reported to date, PPAD, which also has a role in the interaction with host cells, so it may be considered as a double target for PD and RA. In contrast, other abundant odontopathogens responsible for PD such as *Prevotella intermedia* and *Fusobacterium nucleatum*, which both lack a PPAD ortholog, do not evince a link with RA.

## Methods

**Protein production, purification, and characterization.** *P. gingivalis* PPAD (UniProt database [UP] access code Q9RQJ2 or GenBank entry WP\_005873463.1 for NCBI gene tag PG\_1424) was obtained through small-scale homologous overexpression as a secreted protein from plasmid-transformed *P. gingivalis* W83 PPAD-deletion mutant strain  $\Delta ppad$ . Briefly, plasmid pT-COW, which confers resistance against tetracycline<sup>36</sup>, was used as expression vector, and plasmid derivatives encoding the wild type (wt) and a total of 18 PPAD point mutants (W<sup>127</sup>A, D<sup>130</sup>A, D<sup>130</sup>N, R<sup>152</sup>A, R<sup>154</sup>A, R<sup>154</sup>E, T<sup>180</sup>A, G<sup>182</sup>A, H<sup>236</sup>A, H<sup>236</sup>N, D<sup>238</sup>A, D<sup>238</sup>N, C<sup>239</sup>A, C<sup>239</sup>E, C<sup>239</sup>S, N<sup>297</sup>A, C<sup>351</sup>S, and C<sup>351</sup>A; see Table 1) were generated. For this, the wt gene sequence plus 1081 upstream base pairs and 267 downstream base pairs was amplified from *P. gingivalis* W83 genomic DNA with primers pTCowPPADf and pTCowPPADr (see Table 1), which contained recognition sequences for restriction endonucleases *NheI* and *SphI*, respectively. The PCR fragment obtained was ligated into pT-COW, previously digested with *NheI* and *SphI*, to yield plasmid pTPP. Point mutations were thereafter introduced into pTPP by the SLIM method<sup>37</sup> using primers listed under Table 1 and confirmed by DNA sequencing. Plasmid pTPP or its PPAD-mutating variants were introduced into *P. gingivalis* W83  $\Delta ppad$  by conjugation, and bacteria were grown under anaerobic conditions (85% N<sub>2</sub>, 5% H<sub>2</sub> and 10% CO<sub>2</sub>) in liquid Schaedler broth supplemented with hemin (5 mg/ml), menadione (0.5 mg/ml), L-cysteine (50 mg/ml), 1  $\mu$ g/ml tetracycline, and in the presence or absence of 4,4'-dithiodipyridine (DTDP). Expression levels were monitored by Western-blot analysis. For this, 30  $\mu$ l of *P. gingivalis* liquid cultures at OD<sub>600</sub> = 1.0 were separated on SDS-PAGE and transferred to PVDF membranes. Primary PPAD antibodies (kindly provided by Patrick Venables, Oxford) were used in 1:1,000 dilution, secondary HRP-conjugated goat anti-rabbit (Amersham) antibodies were used at 1:10,000 dilution. Cell cultures obtained in the absence of DTDP were used for functional tests (see below). In addition, preparations at a somewhat larger scale—limited by the intrinsic difficulties of cultivating *P. gingivalis*—for structural studies were performed for wt PPAD (DTDP-treated and -untreated) and DTDP-untreated PPAD mutant C<sup>351</sup>A (PPAD-C<sup>351</sup>A) and purified according to<sup>20</sup>.

Protein identity and purity were assessed by 15% Tricine-SDS-PAGE stained with Coomassie blue, peptide-mass fingerprinting of tryptic protein digests (PMF), N-terminal sequencing through Edman degradation, and mass spectrometry (MS). Ultrafiltration steps were performed with Vivaspinn 15 and Vivaspinn 500 filter devices of 10 kDa cut-off (Sartorius Stedim Biotech). Protein concentrations were estimated applying the respective theoretical extinction coefficients by measuring A<sub>280</sub> in a spectrophotometer (NanoDrop). Concentrations were also measured by the BCA Protein Assay Kit (Thermo Scientific) with bovine serum albumin as a standard.

**Activity assays.** PPAD endo- and exodeiminase activities were determined against kininogen-derived peptides of sequence G-F-S-P-F-R-S-S and P-P-G-F-S-P-F-R, respectively. Briefly, peptides (30  $\mu$ g) were incubated for 2 h at 37 °C in 100 mM Tris-HCl, pH 7.5 supplemented with 10 mM L-cysteine in the presence of *P. gingivalis* PPAD (0.12, 1.2, 12 and 120 mU) in 30  $\mu$ l-reaction volumes (final peptide concentration 1 mg/ml). Respective controls were prepared with the same amount of peptide incubated in the reaction buffer alone. Reactions were stopped by addition of 80  $\mu$ l 0.5% trifluoroacetic acid (TFA) in HPLC-quality water and the samples were further analyzed by HPLC using an ÄKTA Micro chromatography system (GE Healthcare) coupled with an Aeris Peptide XB-C18 4.6/150 column (Phenomenex). Peptides were resolved in 19 column volumes using a 2–80% gradient of phase A (0.1% TFA) and phase B (80% acetonitrile, 0.08% TFA) at 1.5 ml/min flow rate. Eluted peaks were fractionated and citrullination was assessed by MS using a HCT Ultra ETD II ESI Iontrap mass spectrometer (Bruker). To determine the velocity of deimination, peptides were incubated with 0.12 mU (1 h) and 120 mU (2 h) PPAD, respectively, in triplicates. Peak integration data were used to determine the amount of modified peptide in each peak (~11% and ~4.5% for P-P-G-F-S-P-F-Cit and G-F-S-P-F-Cit-S-S, respectively) and estimate the reaction velocity (in pmol·mU<sup>-1</sup>·h<sup>-1</sup> ± SD). We found that when P-P-G-F-S-P-F-R became completely citrullinated after overnight incubation (0.12–12 mU), G-F-S-P-F-R-S-S was not modified. Only at ten-fold higher PPAD concentration (120 mU) was certain time-dependent citrullination of the endosubstrate observed, with 5% of peptide being modified after 2 h. Comparatively, 11% of the exosubstrate was citrullinated after 1 h at a thousand-fold lower PPAD concentration (0.12 mU; see Fig. 3a,b).

Competence of wt and mutant PPADs was assessed by the amount of citrulline produced according to a sensitive colorimetric assay<sup>38</sup>. Results obtained from three independent assays were adjusted to  $OD_{600} = 1.0$  and presented as % of the activity of pTPP-transformed  $\Delta ppad$  producing wt PPAD.

**Crystallization and diffraction data collection.** Prior to crystallization, DTDP-treated and -untreated wt PPAD and DTDP-untreated PPAD-C<sup>351</sup>A were dialyzed overnight against buffer A (20 mM Tris-HCl, 20 mM sodium chloride, pH 7.5) and further purified by ionic-exchange chromatography on a TSKgel DEAE-2SW column (TOSOH Bioscience) equilibrated with buffer A. A gradient of 4–60% buffer B (20 mM Tris-HCl, 500 mM sodium chloride, pH 7.5) was applied over 30 ml and samples were collected and pooled. Finally, each pool was concentrated by ultrafiltration and subjected to size-exclusion chromatography on a Superdex 75, 10/300 column (GE Healthcare Life Sciences) equilibrated with buffer C (20 mM Tris-HCl, 150 mM sodium chloride, pH 7.5).

Crystallization assays were performed by the sitting-drop vapor diffusion method. Reservoir solutions were prepared by a Tecan robot and 100 nL crystallization drops were dispensed on  $96 \times 2$ -well MRC plates (Innovadyne) by a Phoenix nanodrop robot (Art Robbins) or a Cartesian Microsys 4000 XL (Genomic Solutions) robot at the joint IBM/IRB Automated Crystallography Platform at Barcelona Science Park. Plates were stored in Bruker steady-temperature crystal farms at 4 °C and 20 °C. Successful conditions were scaled up to the microliter range in 24-well Cryschem crystallization dishes (Hampton Research).

The best crystals of wt PPAD with 4-thiopyridine but without substrate (PPAD-TP; *substrate free*) resulting from DTDP treatment during production (see above) were obtained at 20 °C from 1  $\mu$ l:1  $\mu$ l drops with protein solution at 20–25 mg/ml concentration in 20 mM Tris-HCl pH 7.4, 100 mM sodium chloride and 100 mM sodium acetate (pH 4.5), 25% [w/v] polyethylene glycol 3,350 as reservoir solution. PPAD mutant C<sup>351</sup>A in complex with the dipeptide methionine-arginine (PPAD-C<sup>351</sup>A+M-R; *substrate complex*) was crystallized similarly but with 100 mM tri-sodium citrate, 20% [w/v] polyethylene glycol 3,000, pH 5.5–6.5 as reservoir solution instead. Finally, wt DTDP-untreated PPAD in complex with the dipeptide aspartate-glutamine (PPAD+D-Q; *substrate-mimic complex*) was crystallized with 100 mM tri-sodium citrate, 2 M ammonium sulfate, pH 5.5–6.5 as reservoir solution. All crystals contained protein spanning A<sup>44</sup>-A<sup>475</sup> as determined by Edman degradation and MS analysis. Crystals were cryo-protected by rapid passage through drops containing increasing concentrations of glycerol (up to 15% [v/v]). Complete diffraction datasets were collected at 100 K from liquid-N<sub>2</sub> flash cryo-cooled crystals (Oxford Cryosystems 700 series cryostream) on a Pilatus 6M pixel detector (from Dectris) at beam line XALOC of ALBA synchrotron (Barcelona, Spain<sup>39</sup>). Further data were collected on the same detector type at beam line ID23-1 of ESRF synchrotron (Grenoble, France) within the Block Allocation Group “BAG Barcelona.” Diffraction data were integrated, scaled, merged, and reduced with program XDS<sup>40</sup>. PPAD-TP, PPAD-C<sup>351</sup>A+M-R, and PPAD+D-Q crystals all contained one protein molecule per asymmetric unit (solvent content, respectively, 41%, 44% and 48%), had the symmetry of the space groups P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, C2, and P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, respectively, and had different cell constants (see Table 2 for data processing statistics).

**Structure solution and refinement.** A similarity search with programs PSI-BLAST and HHPRED identified only low homology models (PDB 3HVM, 1ZBR, 1XKN, 2JER, 3H7C, and 2EWO), which failed to render a solution by conventional molecular replacement and Patterson-search methods. At this point, wt PPAD-TP crystal diffraction data were used for structure solution with ARCIMBOLDO<sup>41–43</sup>. Therefore, 16 datasets with resolutions ranging from 3.0 Å to 1.5 Å from different native protein crystals or heavy-ion soaks with similar cell dimensions were merged with program XPREP. A collection of structure fragments was generated from the six aforementioned distant structural relatives, and ARCIMBOLDO runs were set up in parallel with these fragments and libraries<sup>41,42</sup>. These calculations eventually enabled structure solution (see<sup>44,45</sup> for details), and the resulting phase set was subjected to density modification and autotracing with SHELXE<sup>46</sup>, which yielded an improved set of phases and a partial model. These phases and the resulting Fourier map enabled subsequent manual model building with the COOT program<sup>47</sup>, which alternated with crystallographic refinement with PHENIX<sup>48</sup> and BUSTER/TNT<sup>49</sup> under inclusion of TLS refinement, until the final refined model of PPAD-TP was obtained. This consisted of residues A<sup>44</sup>-N<sup>464</sup>, one structural sodium ion, seven glycerols, 460 solvent molecules, and 4-thiopyridine moieties respectively attached to the S $\gamma$  atoms of C<sup>351</sup>, C<sup>462</sup>, and C<sup>239</sup>. The final Fourier map indicated that the side chain of the latter residue was present in two alternate conformations, one bound to thiopyridine and the other with the sulfur as sulfoxide. See Table 2 for final refinement and model quality statistics.

The structure of PPAD-C<sup>351</sup>A+M-R was solved with PHASER within the PHENIX<sup>50</sup> package using the refined coordinates of PPAD-TP. The adequately rotated and translated molecule yielded accurate phases, which enabled calculation of an initial Fourier map. Subsequent model completion and refinement proceeded as above. The final model of PPAD-C<sup>351</sup>A+M-R contained residues A<sup>44</sup>-M<sup>463</sup>, one structural sodium cation, a dipeptide of tentative sequence methionine-arginine, five glycerols, one chloride, two azides, 426 solvent molecules, and a free cysteine disulfide-bonded to C<sup>462</sup>. See Table 2 for final refinement and model quality statistics.

The structure of PPAD+D-Q was solved similarly. Model completion and refinement proceeded as above. The final model comprised residues A<sup>44</sup>-E<sup>465</sup>, one sodium cation, a dipeptide of tentative sequence

aspartate-glutamine (the distinction between aspartate/asparagine and glutamate/glutamine was performed based on surrounding interacting partners), three glycerols, five phosphates, one chloride, one azide, and 689 solvent molecules. See Table 2 for final refinement and model quality statistics.

**Miscellaneous.** Ideal coordinates and parameters for crystallographic refinement of non-standard ligands were obtained from the PRODRG server<sup>51</sup>. Structural similarity searches were performed with DALI<sup>33</sup>, and structure figures were prepared with programs COOT and CHIMERA<sup>52</sup>. Experimental structures were validated with MOLPROBITY<sup>53</sup>. The final coordinates of *P. gingivalis* PPAD-TP (substrate free), PPAD-C<sup>351</sup>A+M-R (substrate complex), and PPAD+D-Q (substrate-mimic complex) are deposited with the PDB at www.pdb.org (access codes 4YT9, 4YTG, and 4YTB).

## References

- Krebs, H. A. & Henseleit, K. Untersuchungen über die Harnstoffbildung im Tierkörper. *Hoppe-Seylers Z. Physiol. Chem.* **210**, 33–66 (1932).
- Woo, H. I., Park, H. D. & Lee, Y. W. Molecular genetics of citrullinemia types I and II. *Clin. Chim. Acta* **431**, 1–8 (2014).
- El-Hattab, A. W., Emrick, L. T., Chanprasert, S., Craigen, W. J. & Scaglia, F. Mitochondria: role of citrulline and arginine supplementation in MELAS syndrome. *Int. J. Biochem. Cell Biol.* **48**, 85–91 (2014).
- György, B., Toth, E., Tarcsa, E., Falus, A. & Buzas, E. I. Citrullination: a posttranslational modification in health and disease. *Int. J. Biochem. Cell Biol.* **38**, 1662–1677 (2006).
- Baka, Z. *et al.* Citrullination under physiological and pathological conditions. *Joint Bone Spine* **79**, 431–436 (2012).
- Muller, S. & Radic, M. Citrullinated autoantigens: from diagnostic markers to pathogenetic mechanisms. *Clin. Rev. Allergy Immunol.* **48**, in press (2015).
- Nijenhuis, S., Zendman, A. J. W., Vossenaar, E. R., Pruijn, G. J. M. & van Venrooij, W. J. Autoantibodies to citrullinated proteins in rheumatoid arthritis: clinical performance and biochemical aspects of an RA-specific marker. *Clin. Chim. Acta* **350**, 17–34 (2004).
- Slade, D. J. *et al.* Protein arginine deiminase 2 binds calcium in an ordered fashion: implications for inhibitor design. *ACS Chem. Biol.* **10**, 1043–1053 (2015).
- Gudmann, N. S., Hansen, N. U., Jensen, A. C., Karsdal, M. A. & Siebuhr, A. S. Biological relevance of citrullinations: diagnostic, prognostic and therapeutic options. *Autoimmunity* **48**, 73–79 (2014).
- Mallya, R. K. *et al.* Anti-keratin antibodies in rheumatoid arthritis: frequency and correlation with other features of the disease. *Clin. Exp. Immunol.* **51**, 17–20 (1983).
- MacGregor, A. J. *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum.* **43**, 30–37 (2000).
- Seymour, G. J., Ford, P. J., Cullinan, M. P., Leishman, S. & Yamazaki, K. Relationship between periodontal infections and systemic disease. *Clin. Microbiol. Infect.* **13 Suppl 4**, 3–10 (2007).
- Preshaw, P. M. *et al.* Periodontitis and diabetes: a two-way relationship. *Diabetologia* **55**, 21–31 (2012).
- de Pablo, P., Chapple, I. L., Buckley, C. D. & Dietrich, T. Periodontitis in systemic rheumatic diseases. *Nat. Rev. Rheumatol.* **5**, 218–224 (2009).
- Maresz, K. J. *et al.* *Porphyromonas gingivalis* facilitates the development and progression of destructive arthritis through its unique bacterial peptidylarginine deiminase (PAD). *PLoS Pathog.* **9**, e1003627 (2013).
- Mayo, C. H. Focal infection of dental origin. *Dent. Cosmos* **64**, 1206–1208 (1922).
- Potempa, J., Sroka, A., Imamura, T. & Travis, J. Gingipains, the major cysteine proteinases and virulence factors of *Porphyromonas gingivalis*: structure, function and assembly of multidomain protein complexes. *Curr. Prot. Pept. Sci.* **4**, 397–407 (2003).
- McGraw, W. T., Potempa, J., Farley, D. & Travis, J. Purification, characterization, and sequence analysis of a potential virulence factor from *Porphyromonas gingivalis*, peptidylarginine deiminase. *Infect. Immun.* **67**, 3248–3256 (1999).
- Casiano-Colon, A. & Marquis, R. E. Role of the arginine deiminase system in protecting oral bacteria and an enzymatic basis for acid tolerance. *Appl. Environ. Microbiol.* **54**, 1318–1324 (1988).
- Bielecka, E. *et al.* Peptidyl arginine deiminase from *Porphyromonas gingivalis* abolishes anaphylatoxin C5a activity. *J. Biol. Chem.* **289**, 32481–32487 (2014).
- Sato, K. *et al.* Identification of *Porphyromonas gingivalis* proteins secreted by the Por secretion system. *FEMS Microbiol. Lett.* **338**, 68–76 (2013).
- de Diego, I. *et al.* Structure and mechanism of cysteine peptidase gingipain K (Kgp), a major virulence factor of *Porphyromonas gingivalis* in periodontitis. *J. Biol. Chem.* **289**, 32291–32302 (2014).
- de Diego, I. *et al.* *Porphyromonas gingivalis* virulence factor gingipain RgpB shows a unique zymogenic mechanism for cysteine peptidases. *J. Biol. Chem.* **288**, 14287–14296 (2013).
- Eichinger, A. *et al.* Crystal structure of gingipain R: an Arg-specific bacterial cysteine proteinase with a caspase-like fold. *EMBO J.* **18**, 5453–5462 (1999).
- Shirai, H., Mokrab, Y. & Mizuguchi, K. The guanidino-group modifying enzymes: structural basis for their diversity and commonality. *Proteins* **64**, 1010–1023 (2006).
- Llácer, J. L. *et al.* The gene cluster for agmatine catabolism of *Enterococcus faecalis*: study of recombinant putrescine transcarbamylase and agmatine deiminase and a snapshot of agmatine deiminase catalyzing its reaction. *J. Bacteriol.* **189**, 1254–1265 (2007).
- Curran, T. M., Lieou, J. & Marquis, R. E. Arginine deiminase system and acid adaptation of oral streptococci. *Appl. Environ. Microbiol.* **61**, 4494–4496 (1995).
- Arita, K. *et al.* Structural basis for Ca<sup>2+</sup>-induced activation of human PAD4. *Nat. Struct. Mol. Biol.* **11**, 777–783 (2004).
- Jones, J. E. *et al.* Characterization and inactivation of an agmatine deiminase from *Helicobacter pylori*. *Bioorg. Chem.* **38**, 62–73 (2010).
- Henningham, A. *et al.* Structure-informed design of an enzymatically inactive vaccine component for group A Streptococcus. *MBio* **4**, e00509–e00513 (2013).
- Galkin, A. *et al.* Structural insight into arginine degradation by arginine deiminase, an antibacterial and parasite drug target. *J. Biol. Chem.* **279**, 14001–14008 (2004).
- Das, K. *et al.* Crystal structures of arginine deiminase with covalent reaction intermediates; implications for catalytic mechanism. *Structure* **12**, 657–667 (2004).
- Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
- Vossenaar, E. R., Zendman, A. J., van Venrooij, W. J. & Pruijn, G. J. PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. *Bioessays* **25**, 1106–1118 (2003).

35. Rose, R., Rose, M. & Ottmann, C. Identification and structural characterization of two 14-3-3 binding sites in the human peptidylarginine deiminase type VI. *J. Struct. Biol.* **180**, 65–72 (2012).
36. Bélanger, M., Rodrigues, P. & Progulske-Fox, A. Genetic manipulation of *Porphyromonas gingivalis*. *Curr. Protoc. Microbiol.* Chapter **13**, Unit13C 12 (2007).
37. Chiu, J., March, P. E., Lee, R. & Tillett, D. Site-directed, ligase-independent mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4h. *Nucleic Acids Res.* **32**, e174 (2004).
38. Boyde, T. R. & Rahmatullah, M. Optimization of conditions for the colorimetric determination of citrulline, using diacetyl monoxime. *Anal. Biochem.* **107**, 424–431 (1980).
39. Juanhuix, J. *et al.* Developments in optics and performance at BL13-XALOC, the macromolecular crystallography beamline at the ALBA synchrotron. *J. Synchrotron Radiat.* **21**, 679–689 (2014).
40. Kabsch, W. XDS. *Acta Crystallogr. sect. D* **66**, 125–132 (2010).
41. Rodriguez, D. D. *et al.* Crystallographic *ab initio* protein structure solution below atomic resolution. *Nat. Methods* **6**, 651–653 (2009).
42. Sammito, M. *et al.* Exploiting tertiary structure through local folds for crystallographic phasing. *Nat. Methods* **10**, 1099–1101 (2013).
43. Millán, C. *et al.* Combining phase information in reciprocal space for ARCIMBOLDO. *Acta Crystallogr. sect. D* **71**, in press (2015).
44. Sammito, M. *et al.* Structure solution with ARCIMBOLDO using fragments derived from distant homology models. *FEBS J.* **281**, 4029–4045 (2014).
45. Millán, C., Sammito, M. & Usón, I. Macromolecular *ab initio* phasing enforcing secondary and tertiary structure. *IUCrJ* **2**, 95–105 (2015).
46. Thorn, A. & Sheldrick, G. M. Extending molecular-replacement solutions with SHELXE. *Acta Crystallogr. sect. D* **69**, 2251–2256 (2013).
47. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. sect. D* **66**, 486–501 (2010).
48. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. sect. D* **68**, 352–367 (2012).
49. Smart, O. S. *et al.* Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. sect. D* **68**, 368–380 (2012).
50. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
51. Schüttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. sect. D* **60**, 1355–1363 (2004).
52. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
53. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. sect. D* **66**, 12–21 (2010).
54. Jones, J. E. *et al.* Synthesis and screening of a haloacetamide containing library to identify PAD4 selective inhibitors. *ACS Chem. Biol.* **7**, 160–165 (2012).
55. García-Castellanos, R. *et al.* Three-dimensional structure of MecI: Molecular basis for transcriptional regulation of staphylococcal methicillin resistance. *J. Biol. Chem.* **278**, 39897–39905 (2003).
56. Weiss, M. S. Global indicators of X-ray quality. *J. Appl. Cryst.* **34**, 130–135 (2001).
57. Karplus, P. A. & Diederichs, K. Linking crystallographic model and data quality. *Science* **336**, 1030–1033 (2012).
58. Engh, R. & Huber, R. Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallogr. sect. A* **47**, 392–400 (1991).
59. Davis, I. W. *et al.* MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383 (2007).

## Acknowledgments

We are grateful to Joan Pous and Xandra Kreplin from the joint IBMB/IRB Automated Crystallography Platform for assistance during crystallization experiments and to Robin Rycroft for very helpful contributions to the manuscript. We further acknowledge the help provided by local contacts at the ESRF and ALBA synchrotrons. We are also grateful to Adam Lesner from the University of Gdansk and Oliwia Bochenska from the Małopolska Center of Biotechnology for peptide synthesis and mass spectrometry analysis. This study was financially supported in part by grants from European, US American, Polish, Spanish, and Catalan agencies (UMO-2012/04/A/NZ1/00051, UMO-2012/05/B/NZ6/00581, UMO-2013/08/W/NZ1/00696, 2137/7.PR-EU/2011/2, 2975/7.PR/13/2014/2, NIH NIDCR DE09761; FP7-PEOPLE-2011-ITN-290246 “RAPID”; FP7-HEALTH-2012-306029-2 “TRIGGER”; FP7-HEALTH-2010-261460 “Gums&Joints”; BFU2012-32862; BFU2012-33516; BFU2012-35367; BIO2013-49320-EXP; MDM-2014-0435; 2014SGR9 and 2014SGR997). IGF acknowledges an FPU-fellowship (AP2010-3799) from the former Spanish Ministry for Education, Culture and Sport. TG acknowledges a “Juan de la Cierva” research contract (JCI-2012-13573) from the Spanish Ministry for Economy and Competitiveness. The Department of Structural Biology of IBMB is a “Maria de Maeztu” Unit of Excellence from the Ministry of Economy and Competitiveness. Funding for data collection was provided in part by ESRF. The Faculty of Biochemistry, Biophysics and Biotechnology of the Jagiellonian University at Kraków (Poland) is a partner of the Leading National Research Center (KNOW) supported by the Polish Ministry of Science and Higher Education.

## Author Contributions

F.X.G.-R., M.S. and J.P. designed the research. T.G., D.M., T.K., I.G.-F., T.G., B.S., A.S., C.M., I.U., F.V., B.P., P.M., M.S. and F.X.G.-R. performed the experiments and/or data analysis. F.X.G.-R., M.S. and J.P. wrote the paper with input from all authors.



### Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Goulas, T. *et al.* Structure and mechanism of a bacterial host-protein citrullinating virulence factor, *Porphyromonas gingivalis* peptidylarginine deiminase. *Sci. Rep.* **5**, 11969; doi: 10.1038/srep11969 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>