

# Structure and mechanism of cysteine peptidase Kgp, a major virulence factor of *Porphyromonas gingivalis* in periodontitis \*\*

Iñaki de Diego<sup>1</sup>, Florian Veillard<sup>2,5</sup>, Maryta N. Sztukowska<sup>2</sup>, Tibisay Guevara<sup>1</sup>, Barbara Potempa<sup>2</sup>, Anja Pomowski<sup>4</sup>, James A. Huntington<sup>4</sup>, Jan Potempa<sup>2,3,\*</sup> and F.Xavier Gomis-Rüth<sup>1,\*</sup>

**Running title:** Crystal structure of gingipain K catalytic and immunoglobulin-type domains

<sup>1</sup> Proteolysis Lab; Molecular Biology Institute of Barcelona; Spanish Research Council CSIC; Barcelona Science Park; Helix Building; c/Baldiri Reixac,15-21; 08028 Barcelona, Catalonia (Spain).

<sup>2</sup> Oral Immunology and Infectious Disease; University of Louisville School of Dentistry; Louisville, KY 40202 (US).

<sup>3</sup> Department of Microbiology; Faculty of Biochemistry, Biophysics and Biotechnology; Jagiellonian University; Ul. Gronostajowa 7; 30-387 Kraków (Poland).

<sup>4</sup> Department of Haematology; University of Cambridge; Cambridge Institute for Medical Research; Wellcome Trust/MRC Building; Hills Road; Cambridge CB2 0XY(UK).

<sup>5</sup> Present address: Institut de Biologie Moléculaire et Cellulaire; URP9022 CNRS; Strasbourg (France).

\* Corresponding authors. E-mail: fxgr@ibmb.csic.es; phone: (+34) 934 020 186; fax : (+34) 934 034 979 or e-mail : jspote01@louisville.edu; phone (+1) 502 852 5572 ; Fax (+1) 502 852 5572.

**Keywords:** cysteine protease; X-ray crystal structure; microbial infection; oral pathogen

**Background:** The cysteine peptidase gingipain K is a major proteolytic virulence factor of *Porphyromonas gingivalis*.

**Results:** The structure of the catalytic and immunoglobulin-type domains has been solved in complex with a covalent inhibitor.

**Conclusion:** A distinct S<sub>1</sub> pocket explains its high specificity for lysines.

**Significance:** The structural details reveal the working mechanism and may lead to the design of drugs to selectively treat periodontitis.

## SUMMARY

Cysteine peptidases are key proteolytic virulence factors of the periodontopathogen *Porphyromonas gingivalis*, which causes chronic periodontitis, the most prevalent dysbiosis-driven disease in humans. Two peptidases, gingipain K (Kgp) and R (RgpA and RgpB), which differ in their selectivity after lysines and arginines, respectively, collectively account for 85% of the extracellular proteolytic activity of *P. gingivalis* at the site of infection. Therefore, they are promising targets for the design of specific inhibitors. While the structure of the catalytic domain of RgpB is known, little is known about Kgp, which shares only 27% sequence identity. We report the high-resolution crystal structure of a competent fragment of Kgp encompassing the catalytic cysteine peptidase domain and a downstream immunoglobulin superfamily-like domain, which is required for folding and secretion of Kgp *in vivo*. The structure, which strikingly resembles a tooth, was serendipitously trapped with a fragment of a covalent inhibitor targeting the catalytic cysteine. This

provided accurate insight into the active site and suggested that catalysis may require a catalytic triad, C<sup>477</sup>-H<sup>444</sup>-D<sup>388</sup>, rather than the cysteine-histidine dyad normally found in cysteine peptidases. In addition, a 20Å-long solvent-filled interior channel traverses the molecule and links the bottom of the specificity pocket with the molecular surface opposite the active-site cleft. This channel, absent in RgpB, may enhance the plasticity of the enzyme, which would explain the much lower activity *in vitro* towards comparable specific synthetic substrates. Overall, the present results report the architecture and molecular determinants of the working mechanism of Kgp, including interaction with its substrates.

Bacteria are normally part of the commensal flora that is generally beneficial for human health (1). However, in a susceptible host, some are pathogenic and invade cells and tissues causing infection and disease. Moreover, the emergence of resistant strains, which are currently responsible for half of all infections, has exacerbated the danger (2). In the US alone, resistant pathogens infect at least 2 million people every year, which makes such infections more common than cancer, and they cause 23,000 deaths (2). The only way to keep pace with these extremely adaptive pathogens is *via* continuous effort in the development of new antimicrobials. Inexplicably, however, the pharmaceutical industry has neglected the development of new antibiotics in recent decades: only four new drug applications were approved by the US FDA in 2005-2012 (3). Responsibly, academic research must fill the gap.

Among the most prevalent human bacterial commensals turned into a pathogen is *Porphyromonas gingivalis*, a Gram-negative oral anaerobe that causes periodontitis, an inflammatory disease that afflicts half the adult population in the US, destroys the gums, and leads to tooth loss (4). It was even detected in the 5,300-year old mummy of the Tyrolean Iceman “Ötzi” in what may well

be the earliest report of gingival infection in *Homo sapiens* (5). *P. gingivalis* invades periodontal tissues by colonizing the gingival sulcus and proliferating in the subgingival plaque. It evades the host defense mechanisms through a panel of virulence factors that deregulate innate immune and inflammatory responses. In addition, bacteria and their products can enter the circulation contributing to development and severity of systemic diseases at distal sites, such as cardiovascular diseases (6), rheumatoid arthritis (7), and diabetes (8), and also preterm delivery (9). Currently, specific treatment of severe periodontitis consists only of curettage of the affected area, which is time consuming, painful and needs frequent repetition (10), and the adjunct doxycycline hyclate (periostat), which targets matrix metalloproteinases and was approved by the FDA in 1988 (11). Consequently, there is an urgent need for the development of novel therapeutic approaches.

Peptidases are a substantial part of the infective armamentarium of *P. gingivalis* (12-14). Most are cysteine peptidases, and the best characterized are the gingipains K (*alias* Kgp) and R (RgpA and RgpB) (4,15), which are major virulence factors of the pathogen (16). Gingipains are cell-surface-anchored or soluble, and responsible for up to 85% of the total extracellular proteolytic activity of *P. gingivalis* (17). This activity yields nutrient acquisition, cleavage of host-cell surface receptors, signaling *via* protease-activated receptors, and inactivation of cytokines and components of the complement system. The pathogen thus keeps host bactericidal activity in check and maintains chronic inflammation (4). In particular, Kgp cleaves many constituents of human connective tissue and plasma, including immunoglobulins, fibronectin, plasma kallikrein, fibrinogen, iron-, heme- and hemoglobin-transporting proteins, and peptidase inhibitors, thus contributing to bleeding and vascular permeability, as well as to heme and iron uptake by the bacterium (18,19). Further pathophysiologically relevant substrates of Kgp include cadherins at the cell-adherence junction, membrane TNF $\alpha$ , interleukin-8, the interleukin-6 receptor, thrombomodulin, complement regulatory protein CD46, and osteoprotegerin (18). Kgp thus contributes far more to the pathogenicity of *P. gingivalis* than any other peptidase (20) and so it is essential for bacterial survival and the pathological outcome of periodontitis (21). This was further confirmed by the reduction of bacterial virulence observed in a mouse model of infection upon specific inhibition of Kgp (22). Accordingly, Kgp—as RgpA and RgpB—is a promising target for the development of therapeutic inhibitors to treat periodontitis (18,23).

Functionally, RgpA and RgpB specifically cleave bonds after arginines, whereas Kgp cleaves after lysines (21,24). Structurally, these enzymes are translated as multi-domain proteins made up of at least: a signal peptide, a pro-domain, a catalytic domain (CD), an immunoglobulin-superfamily domain (IgSF), and a C-terminal domain. RgpB shows just this minimal configuration (21). RgpA has four additional hemagglutinin/adhesion domains (termed RgpA<sub>A1</sub>-RgpA<sub>A4</sub>) inserted between the IgSF and the C-terminal domain. Kgp, in turn, has between three and five such domains (termed KgpA<sub>A1</sub>-KgpA<sub>A5</sub>), depending on the bacterial strain, thus spanning up to 1,723-1,732 residues (21). Both Kgp and RgpA are subjected to extensive post-translational proteolytic processing and are secreted as non-covalent but very tight complexes of the catalytic and

hemagglutinin/adhesion domains, which are held together through oligomerization motifs (25).

Detailed structural and functional knowledge of target virulence factors at the molecular level can lead to the development of new drugs following *rational drug design* strategies (26). Atomic structural data are available for the catalytic and IgSF domains of RgpB, for both a zymogen complex and the active form (27,28), as well as for the ancillary hemagglutinin/adhesion domains KgpA<sub>A1</sub>, KgpA<sub>A2</sub> and KgpA<sub>A3</sub> of Kgp (29,30). The latter, however, do not provide insight into the proteolytic function and mechanism of Kgp. Given the importance of the distinct but complementary cleavage specificities of RgpB and Kgp, which may be related to the differences between their respective CD+IgSF moieties (27% identity; see Fig. 1), we analyzed the 3D structure of a catalytically competent 455-residue fragment of Kgp from *P. gingivalis* strain W83 [hereafter Kgp(CD+IgSF)] and assessed its molecular determinants of action and specificity.

## EXPERIMENTAL PROCEDURES

*Protein production* — Kgp(CD+IgSF) of *P. gingivalis* strain W83 (sequence D<sup>229</sup>-P<sup>683</sup>; see UniProt database access code Q51817) *plus* a C-terminal hexahistidine tag was purified by affinity chromatography on Nickel-Sepharose beads from culture medium of *P. gingivalis* mutant strain ABM1 expressing recombinant Kgp with one oligomerization motif disrupted by hexahistidine insertion (25,31). In contrast to wild-type strain W83, which secretes only heterooligomeric complexes of Kgp and RgpA, strain ABM1 can release soluble and functional Kgp fragments into the medium (25). This facilitates purification of a protein variant that is compatible with crystallization. To avoid autoproteolysis, the sample was incubated with *N* $\alpha$ -tosyl-L-lysinechloromethane (TLCK; Sigma) prior to elution from the beads.

*Crystallization and diffraction data collection* — 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 96x2-well MRC plates (Innovadyne) by a Phoenix nanodrop robot (Art Robbins) or a Cartesian Microsys 4000 XL (Genomic Solutions) robot at the Automated Crystallography Platform (PAC) 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 were obtained at 20°C with protein solution (at 5.7mg/mL in 5mM Tris-HCl pH 7.4; 0.02% sodium azide) and 22% polyethylene glycol 8000; 0.1M sodium cacodylate pH6.5; 0.2M calcium acetate as reservoir solution from 2:1  $\mu$ L drops. Crystals were cryo-protected by immersion in harvesting solution (18% polyethylene glycol 8000; 0.08M sodium cacodylate pH6.5; 0.16M calcium acetate; 20%[v/v] glycerol). A complete diffraction dataset was collected from a liquid-N<sub>2</sub> flash-cryo-cooled crystal at 100K (cooled by an Oxford Cryosystems 700 series cryostream) on an ADSC Q315R CCD detector at beam line ID14-4 of the European Synchrotron Radiation Facility

(ESRF, Grenoble, France) within the Block Allocation Group “BAG Barcelona.” This crystal was orthorhombic and contained one Kgp(CD+IgSF) moiety per asymmetric unit ( $V_M=2.2\text{\AA}^3/\text{Da}$ ; 44% solvent content; (32)), and was tightly packed (for comparison, see e.g. (33)). Diffraction data were integrated, scaled, merged, and reduced with programs XDS and XSCALE (34) (see Table 1).

**Structure solution and refinement** — The structure of Kgp(CD+IgSF) was solved by likelihood-scoring molecular replacement with program PHASER (35) using the protein part of the structure of RgpB(CD+IgSF) of *P. gingivalis* strain HG66 (GenBank AAB41892; Protein Data Bank [PDB] access code 1CVR; (28)) as a searching model. The side chains were trimmed from the model with program CHAINSAW within the CCP4 suite (36) based on sequence alignment performed with MULTALIN (37). A two-body search was performed with the CD and the IgSF separately to obtain suitable phases. These calculations rendered two unambiguous solutions at final Eulerian angles ( $\alpha$ ,  $\beta$ ,  $\gamma$ , in  $^\circ$ ) of 225.7, 87.3, 30.7 and 226.2, 92.0, 34.2; and fractional cell coordinates ( $x$ ,  $y$ ,  $z$ ) of 0.019, 0.265, 0.096 and 0.959, 0.259, 0.090, respectively. The initial values for the rotation/translation function Z-scores were 7.9/10.1 and 5.4/7.2, respectively, which confirmed  $P2_12_12_1$  as the correct space group. A Fourier map calculated with the appropriately rotated and translated model was then subjected to density modification and model extension with ARP/wARP (38). The model obtained was completed through successive rounds of manual model building using programs TURBO-FRODO (39) and COOT (40), and crystallographic refinement with program BUSTER/TNT (41), which included TLS refinement (one TLS group for each domain), until completion of the model. The final model contained Kgp residues D<sup>229</sup>-V<sup>680</sup> (the last three residues and the hexahistidine tag were not visible in the final map) plus four glycerol, one standalone histidine, two acetate, and three azide molecules, in addition to 533 solvent molecules and six (tentatively assigned) cations: two calciums, three sodiums, and one nickel (see Table 1). Of these, the two calcium ions (numbered Ca<sup>999</sup> and Ca<sup>998</sup>) and two of the sodium ions (numbered Na<sup>997</sup> and Na<sup>996</sup>) were intrinsic parts of the structure and are described under “Results and Discussion.” The nickel ion, in turn, was observed at the interface between two symmetric molecules, tetrahedrally coordinated by the carboxylate oxygens of E<sup>379</sup>, a symmetric E<sup>355</sup>, and an acetate from the mother liquor, as well as by the Ne2-atom of an isolated histidine, possibly resulting from digestion of the C-terminal hexahistidine tag during purification. The third sodium ion was bound 10Å away from the nickel by the main-chain carbonyl of T<sup>415</sup>, four solvent molecules, and the carboxylate side chain of a symmetric E<sup>498</sup> residue. We attribute these two metal sites to purification and crystallization artifacts.

Although the main chain and most of the side chains of the entire molecule were fully defined in the final Fourier map owing to the high resolution and quality of the diffraction data, the CD moiety was more rigid and better defined than the downstream IgSF, as indicated by the lower average thermal displacement parameter (13.6 vs. 34.5Å<sup>2</sup>, see Table 1). This is reminiscent of the structure of RgpB(CD+IgSF) in complex with its pro-domain (27). The side chain of the catalytic cysteine residue (C<sup>477</sup>; residue type CKC), showed additional density (see “Results and

Discussion”). Moreover, M<sup>594</sup> showed alternate occupancy for its side chain. All other sulfur-containing side chains (three cysteines and ten methionines) were apparently unaltered according to the final Fourier map. The only Ramachandran outlier of the structure was A<sup>443</sup> (Table 1), which, however, was unambiguously defined by the final Fourier map and had similar main-chain angles in the RgpB structures (27,28). Three proline residues were found in *cis* conformation (P<sup>241</sup>, P<sup>424</sup>, and P<sup>453</sup>).

**Miscellaneous** — The structure-based sequence alignment in Fig. 1 was performed with program EXPRESSO within T-COFFEE v. 10.0 (42) and represented with program ESPript 3.0 (43). Ideal coordinates and parameters for crystallographic refinement of non-standard ligands were obtained from the PRODRG server (44). Structural similarity searches were performed with DALI (45), and structure figures were prepared with programs COOT and CHIMERA (46). The model was validated with MOLPROBITY (47). The final coordinates of *P. gingivalis* Kgp(CD+IgSF) are deposited with the PDB at www.pdb.org (access code 4RBM).

## RESULTS AND DISCUSSION

**Overall structure of Kgp catalytic domain** — The structure of Kgp(CD+IgSF) is elongated, with approximate maximal dimensions 75x50x45Å (Fig. 2a). Curiously, it resembles a tooth, with CD featuring the crown and IgSF the root (Fig. 2b). The neck is the interface between the two domains, and the active site is at the cusp, on the grinding surface (see below).

The globular CD (D<sup>229</sup>-P<sup>600</sup>; see Figs. 1 and 2a) is a competent cysteine peptidase domain and conforms to the  $\alpha/\beta$ -hydrolase or PLEES fold (48,49). It contains four cation-binding sites (two sodium and two calcium ions; Fig. 3a-c), which generally contribute to tertiary structure integrity (50). It is subdivided into a smaller N-terminal (or A-) sub-domain (NSD; D<sup>229</sup>-K<sup>375</sup>) and a larger C-terminal (or B-) sub-domain (CSD; S<sup>376</sup>-P<sup>600</sup>). The NSD starts on the left of the molecule (orientation hereafter according to Fig. 2a) with a small helical segment ( $\alpha 1$ ; for regular secondary structure elements, see Figs. 1 and 2a,d), and the polypeptide chain follows an extended trace downwards along the surface. At P<sup>241</sup>, the chain makes a sharp turn upward and enters a four-stranded parallel pleated  $\beta$ -sheet (sheet I) through the second strand from the right ( $\beta 1$ ). This sheet (from left to right:  $\beta 6$ - $\beta 3$ - $\beta 1$ - $\beta 2$ ) has connectivity +1x,-2x,-1x according to (51) and is twisted by  $\sim 40^\circ$  but not arched or curved. The NSD is a three-layer ( $\alpha\beta\alpha$ )-sandwich, thus sheet I is flanked by two almost parallel helices on its right ( $\alpha 2$  and  $\alpha 5$ ) and two more ( $\alpha 3$  and  $\alpha 4$ ) on its left. While in general the regular secondary structure elements are connected by tight loops, the one connecting  $\beta 3$  with  $\beta 6$  (L $\beta 3\beta 6$ ) exceptionally spans 30 residues and contains a  $\beta$ -ribbon ( $\beta 4\beta 5$ ) and a calcium-binding site (Ca<sup>998</sup>) in addition to helix  $\alpha 3$ . This cation is oxygen-coordinated in an octahedral manner—as is usual for calcium (52)—by D<sup>330</sup> O $\delta 1$ , two solvent molecules, and, bidentately, the carboxylate oxygens of E<sup>343</sup> in a plane with the ion. D<sup>337</sup> O $\delta 1$  and F<sup>341</sup> O are found at the apical positions (Fig. 3b). Therefore, one in-plane position of the octahedron is split into two ligands, and the binding distances range between

2.31 and 2.55Å, typical values for calcium (2.36-2.39Å; (53)).

After  $\alpha 5$ , at the front neck surface, the polypeptide enters the CSD with helix  $\alpha 6$ , which in turn leads to a central six-stranded twisted ( $\sim 40^\circ$ )—but not arched or curved—pleated  $\beta$ -sheet (sheet II;  $\beta 8$ - $\beta 7$ - $\beta 9$ - $\beta 13$ - $\beta 14$ - $\beta 15$  from left to right). The chain enters the sheet with the second strand from the left and has connectivity -1x,+2x,+1x,+1x,+1, i.e. all strands are parallel and run upward except for the rightmost one,  $\beta 15$ , which runs downward. The latter creates the junction with the NSD and runs parallel to the leftmost strand of sheet I but is horizontally rotated  $\sim 60^\circ$  away, giving rise to a pseudo-continuous ten-stranded supersheet. Like NSD, CSD is a three-layer ( $\alpha\beta\alpha$ )-sandwich, which contains the active-site cleft (see below). Five helices ( $\alpha 7$ ,  $\alpha 11$ ,  $\alpha 12$ ,  $\alpha 13$ , and  $\alpha 14$ ) are found on the left side of the sheet and three more ( $\alpha 8$ ,  $\alpha 9$ , and  $\alpha 10$ ) on the right flank. Interestingly, helices  $\alpha 11$  and  $\alpha 12$  are aligned with respect to their axes and almost in phase. Such interrupted helices are exceedingly rare in protein structures (51), and only segment G<sup>522</sup>-G<sup>539</sup> (like prolines, flanking glycines are observed in hinges (54)) prevents these helices from being a single continuous unit. This intercalated segment gives rise to an extended loop that protrudes from the molecular surface and folds back to cover helix  $\alpha 7$  like a cape, and it contains a sodium-binding site (Na<sup>997</sup>). This cation is octahedrally coordinated by six oxygen ligands—the most common coordination number for sodium (55)—through T<sup>536</sup> O $\gamma$ , V<sup>521</sup> O, D<sup>542</sup> O $\delta 2$ , and a solvent molecule coplanar with the metal and through S<sup>537</sup> O and Y<sup>402</sup> O $\eta$  at the apical positions. Coordinating distances span 2.31-2.59Å (Fig. 3c), which is consistent with most common distances for sodium being 2.38-2.41Å (53). Nearby, L $\alpha 12\alpha 13$  contains a second octahedral oxygen-liganded sodium site (Na<sup>996</sup>), 13.4Å from the former. This ion is bound at distances of 2.36-2.69Å by Y<sup>550</sup> O, N<sup>551</sup> O $\delta 1$ , A<sup>543</sup> O, and L<sup>546</sup> O in the plane, and apically by T<sup>544</sup> O and S<sup>549</sup> O (Fig. 3c). As found for the segment connecting strands  $\beta 3$  and  $\beta 6$  in NSD sheet I, the segment connecting  $\beta 9$  and  $\beta 13$  in the CSD is elaborate and includes a small three-stranded antiparallel  $\beta$ -sheet (sheet III;  $\beta 10$ ,  $\beta 11$ , and  $\beta 12$ ), which is almost perpendicular to sheet II. The downstream loop, L $\beta 13\alpha 10$  contributes, together with L $\beta 3\beta 4$  of the NSD, to a second calcium site (Ca<sup>999</sup>), which may thus have a role in maintaining the structural integrity of the NSD-CSD interface. This calcium is bound by F<sup>482</sup> O, two solvent molecules and, bidentately, the carboxylate oxygens of E<sup>491</sup> in a plane with the ion. A further solvent molecule and, bidentately, the carboxylate oxygens of D<sup>313</sup> are in the respective apical positions (Fig. 3a). Therefore, two positions of the octahedron are split into two ligands, and the binding distances range from 2.33 to 2.55Å (Fig. 3a). After strand  $\beta 15$ , at segment D<sup>590</sup>-S<sup>592</sup>, the polypeptide abruptly changes direction and runs horizontally outward from the interface with the NSD to the left molecular surface. At A<sup>598</sup>, the chain turns abruptly downward and leads to the interface between CD and IgSF at P<sup>600</sup>-K<sup>601</sup>.

Finally, an internal channel is found within the CSD, vertically traversing the molecule over 20Å from the bottom of the specificity pocket in the active site (see below) to the lower outer surface of the sub-domain (Fig. 3d), where it emerges through a crater surrounded by P<sup>595</sup>,

N<sup>551</sup>, and R<sup>597</sup>. It is filled with 13 solvent molecules, which are well defined in the final Fourier map (average thermal displacement parameter 17.0Å<sup>2</sup>; for comparison, the overall value for the CD is 13.6Å<sup>2</sup> and that of all solvent molecules is 29.8Å<sup>2</sup>). The outermost solvent molecule of the channel at the domain surface is bound by S<sup>669</sup> O $\gamma$  from the downstream domain IgSF. This channel is embraced by sheet II and helices  $\alpha 7$ ,  $\alpha 11$ + $\alpha 12$  and  $\alpha 13$ . Such extended solvent channels traversing the inner core of proteins are rare and here its role, if any, is unknown. The channel is not wide enough to evacuate reaction products, as e.g. in catalase (56) or the ribosome (57), and is too far away from the active-site cysteine to serve as supplier of solvent for the deacylation step in catalysis. In addition, in the structurally related RgpB, this channel is replaced by a compact hydrophobic core (see PDB 1CVR; (28)).

Solvent molecules buried in internal cavities, which are integral structural components of proteins, interchange with the external bulk solvent (58), thus conferring a “breathing” motion to a protein. By serving as mobile hydrogen bonding donors or acceptors, internal waters may facilitate transition and structural rearrangement between different functional states (59), and they cluster at internal cavities of functional importance such as hinge regions or along channels (60). However, the generation of cavities inside a protein at places where a compact hydrophobic core is found in close structural relatives usually reduces stability (61), although water-filled cavities destabilize less than empty cavities: the water molecules may still interact favorably with neighboring protein residues (62). On the other hand, the hydrogen bonding potential of a water molecule inside a protein structure is less exploited than in the aqueous phase, and moving a solvent molecule from bulk solvent to the interior of a protein entails entropic costs (63) and energy costs for hydration of the cavity (64), which, in turn, destabilize protein structures. In Kgp(CD+IgSF), the extended solvent channel could contribute more to the overall plasticity and flexibility of the enzyme than in RgpB its compact hydrophobic core. While certain flexibility, at least around the active site, is a prerequisite for efficient catalysis (65), destabilization of the overall enzyme moiety contravenes the axiom that proteins must adopt a stable tertiary folds to be wholly functional (66). This would be consistent with much lower activity of Kgp *in vitro* than RgpB against comparable synthetic substrates mimicking their respective specificities (24,67). This, in turn, would apparently contract the superior role of Kgp as proteolytic virulence factor (20). It must be kept in mind, however, that native Kgp occurs as a complex of the catalytic and hemagglutinin-adhesion domains, which work as a homing device to deliver Kgp to its targets and exert essential functions for *P. gingivalis* such as agglutination of red blood cells, acquisition of heme and binding to the extracellular matrix (4,18,19).

*Overall structure and similarity of Kgp immunoglobulin superfamily domain* — With K<sup>601</sup>, the polypeptide chain enters the IgSF, which is essential for folding of Kgp: no properly folded CD is detected by specific monoclonal antibodies if IgSF is ablated, despite the truncated *kgp* gene being transcribed (68). In addition, only residual Kgp-specific activity is detected in such deletion mutants (68). Structurally, IgSF consists of a six-

stranded antiparallel open  $\beta$ -barrel adopting a Greek-key topology for its first four strands ( $\beta$ 16- $\beta$ 19- $\beta$ 18- $\beta$ 17) followed by a final  $\beta$ -ribbon structure ( $\beta$ 20 $\beta$ 21). The initial segment of IgSF (K<sup>601</sup>-P<sup>608</sup>) runs in rather extended conformation and partially closes the open side of the barrel but, due to a bulge at P<sup>608</sup>-P<sup>612</sup>, it only interacts through one hydrogen bond with both neighboring strands  $\beta$ 16 (T<sup>606</sup> O-Q<sup>621</sup> O; 2.99Å) and  $\beta$ 21 (K<sup>601</sup> O-L<sup>672</sup> N; 2.86Å), so, strictly speaking, it cannot be considered a proper  $\beta$ -strand (Fig. 2a). Overall, the IgSF fold corresponds to classic immunoglobulin-like domains, which usually function as cell-adhesion molecules (69). In particular, it best fits into the C2 set represented by the second domain of vascular cell-adhesion molecule-1 (PDB 1VCA). Consistently, structural similarity searches identified, in addition to the homologous domain from RgpB (PDB 1CVR and 4IEF; DALI Z-score=14.4; *rmsd*=1.2Å; length of alignment=79; seq. identity=23%; see also Fig. 1), the N-terminal immunoglobulin-like domains of complement components C3 (PDB 2WII; 8.1; 2.4Å; 75; 15%) and C5 (PDB 3PRX; 7.4; 2.3Å; 73; 12%) as related. Interestingly, these proteins themselves are degraded by Kgp (70). In addition, the macroglobulin-like MG domains are also similar (71).

The IgSF contacts the bottom of the CD through an interface that generates the neck of the tooth and involves L $\alpha$ 1 $\beta$ 1,  $\alpha$ 2, L $\alpha$ 2 $\beta$ 2 and the end of  $\alpha$ 5 of CD, which fit into the concave outer surface of the IgSF barrel. In turn, L $\beta$ 20 $\beta$ 21 of IgSF inserts like a wedge between CD C-terminal segment S<sup>592</sup>-P<sup>600</sup> and the Na<sup>996</sup>-stabilized loop L $\alpha$ 12 $\alpha$ 13.

*Catalytic site and active-site cleft* — Catalysis in Kgp occurs at the cusp of the tooth through binding of peptide substrates to the active-site cleft (Fig. 2a,d,e). As occurs in  $\alpha/\beta$ -hydrolase- or PLEES-fold enzymes, active-site residues are provided by loops connecting strands at the C-terminal edge of the central  $\beta$ -sheet, here sheet II of the CSD. We serendipitously trapped the structure of Kgp(CD+IgSF) in a covalent reaction intermediate mimic, which was interpreted—based on the excellent quality of the Fourier map (Fig. 2c)—as an L-lysylmethyl (LM) moiety attached to the S $\gamma$  atom of the catalytic cysteine, C<sup>477</sup> (provided by L $\beta$ 13 $\alpha$ 10). The latter was identified as such by active-site labeling and confirmed by mutagenesis (72). LM introduces an extra methylene group between C<sup>477</sup> S $\gamma$  and the carbonyl mimicking the scissile carbonyl (Fig. 2c), so it does not yield a thioester and cannot be hydrolyzed. This covalent modification resulted from the irreversible cysteine-peptidase chloromethylketone inhibitor, TLCK, used during purification (see “Experimental procedures”). Inhibition of cysteine peptidases by TLCK was first reported by Cohen (73) and such chloromethylketones are routinely used during protein purification to prevent degradation (74). Noteworthy, while these compounds target the catalytic cysteine in cysteine peptidases (75,76), they also inhibit serine peptidases, but by covalent attachment to the aromatic N $\epsilon$ 2 nitrogen of the respective catalytic histidines instead (77). In a second step, the nearby catalytic serine of serine peptidases may (78) or may not (77) attack the carbonyl of the N $\epsilon$ 2-attached carboxymethyl moiety and yield a tetrahedral reaction intermediate-like product covalently bound to both serine O $\gamma$  and histidine N $\epsilon$ 2.

The LM moiety enabled us to identify the active-site cleft (Figs. 2d,e and 4). When viewed in the standard orientation of cysteine peptidases (28,79), i.e. with a substrate binding horizontally from left (non-primed side) to right (primed side), the cleft is rather flat and delimited by H<sup>444</sup>-E<sup>447</sup> from  $\beta$ 10, D<sup>388</sup>-V<sup>395</sup> from L $\beta$ 7 $\alpha$ 7, and F<sup>527</sup> at its bottom; C<sup>476</sup>-I<sup>478</sup> from L $\beta$ 13 $\alpha$ 10 at its back; and P<sup>509</sup>-W<sup>513</sup> from L $\beta$ 14 $\alpha$ 11 and I<sup>573</sup>-H<sup>575</sup> from L $\alpha$ 13 $\alpha$ 14 at its top. L $\beta$ 9 $\beta$ 10 contains H<sup>444</sup>, which strongly hydrogen-bonds through its N $\delta$ 1 atom the carbonyl group oxygen of the LM moiety (2.61Å apart; Figs. 2d,e and 4). This oxygen is close to where the scissile carbonyl oxygen is expected to be in a true substrate after the acylation step of catalysis. This is consistent with H<sup>444</sup> playing a major role in catalysis, potentially as part of a charge relay system together with C<sup>477</sup>, as previously suggested (18).

While most cysteine peptidases comprise a cysteine-histidine dyad as catalytic residues (80,81), the position and proximity of one of the carboxylate oxygens of D<sup>388</sup> from L $\beta$ 7 $\alpha$ 7 to H<sup>444</sup> N $\epsilon$ 2 in the present structure (2.68Å apart; Fig. 2d,e) strongly suggests a role for this aspartate in catalysis in Kgp, as already described in the foot-and-mouth-disease-virus leader cysteine peptidase (82) and as postulated for RgpB. In the latter, however, glutamate E<sup>381</sup> (RgpB of *P. gingivalis* strain HG66 numbering in *italics* according to GenBank entry AAB41892, see Fig. 1; subtract 229 for the protein numbering employed in PDB 1CVR and (28)) is found instead of an aspartate (27,28). This hypothesis would entail that Kgp had a catalytic triad spanning C<sup>477</sup>-H<sup>444</sup>-D<sup>388</sup> and that the cleavage mechanism would include a thiolate-imidazolium ion pair making an oxyanion hole-assisted nucleophilic attack on the scissile peptide carbonyl in the acylation step (81). Alternatively, the imidazole may also function as a general base and abstract a proton from the cysteine thiol group (81). In either case, the histidine imidazolium would thereafter transfer a proton to the leaving  $\alpha$ -amino group of the downstream cleavage product, and the upstream part of the substrate would remain covalently bound as a thioester to the catalytic cysteine. We hypothesize that the aspartate would also have a role in protonation and, thus, side-chain orientation of the histidine imidazole/ium during catalysis, as in serine peptidases. Unfortunately, site-specific mutagenesis failed to demonstrate the catalytic efficacy of third residues in other cysteine peptidases such as papain (81), so this hypothesis remains to be verified by other methods.

The complex of Kgp(CD+IgSF) with LM also revealed that the specificity pocket in S<sub>1</sub> can accommodate a lysine side chain, whose  $\epsilon$ -amino group is tetrahedrally bound by one of the carboxylate atoms of D<sup>516</sup> (2.76Å apart), T<sup>442</sup> O $\gamma$  (2.99Å), N<sup>475</sup> O (3.06Å), and one of three solvent molecules forming a buried solvent cluster at the pocket bottom (2.69Å; Figs. 2e and 4). This solvent molecule is further bound to the second water of the cluster (2.89Å), W<sup>513</sup> O (2.91Å), D<sup>516</sup> C $\beta$  (3.16Å), and V<sup>395</sup> C $\gamma$ 1 (2.75Å) in a distorted square-based pyramidal manner. The second solvent, in turn, is further bound to T<sup>399</sup> O $\gamma$ 1 (2.89Å), T<sup>442</sup> O $\gamma$ 1 (2.89Å) and the third solvent molecule (2.61Å) in a tetrahedral manner. Finally, the latter solvent is further bound to D<sup>516</sup> O (2.79Å) and S<sup>520</sup> O $\gamma$  (2.73Å). The latter oxygen bridges the three-solvent cluster at the bottom of the pocket with the internal

solvent channel (see above and Figs. 3d and 4). The aliphatic part of the lysine side chain of LM, in turn, is sandwiched between the side chain of  $W^{513}$ , the main chain at  $N^{475}$ - $C^{477}$ , and  $S^{511}$  O $\gamma$ . Replacement of lysine in P<sub>1</sub> with an arginine, which would match the specificity of RgpB, would entail rupture of the salt bridge with  $D^{516}$  and the hydrogen bonds with  $N^{475}$  O and  $T^{442}$  O $\gamma$ , and, possibly, clashes with the latter atom, thus explaining why Kgp is specific for lysines and not arginines. This lysine specificity resembles trypsin and, like several trypsin-like serine peptidases, Kgp has been shown to cleave proteins involved in the blood coagulation/fibrinolysis cascade (24). In contrast to these serine peptidases, however, Kgp needs an anaerobic environment as found in the periodontal pockets of infected patients. Finally, upstream of S<sub>1</sub>, S<sub>2</sub> is shallow and small, framed by  $Y^{512}$ ,  $H^{575}$ ,  $W^{513}$ , and  $W^{391}$  (Figs. 2d,e and 4). This explains why Kgp substrates do not have arginines or lysines at position P<sub>2</sub> (18).

*Structural kinship of the CD*—Despite low overall sequence identity, the closest structural similarity of Kgp CD was found with the corresponding domain of the two essentially identical RgpB strain variants studied to date, (strains HG66 and W83; PDB 1CVR and 4IEF; DALI Z-score=44.6 and 43.2; *rmsd*=2.1 and 2.0Å; length of alignment=336 and 328; seq. identity=24% and 26%; see also Fig. 1). While both Kgp and RgpB CDs generally fit well on top of each other, in particular at the regular secondary structure elements (Fig. 5a,b), large differences are found in a number of loops, as well as due to the long solvent-filled internal channel of Kgp, which is absent in RgpB. Notable insertions or deletions, or substantially different chain traces, are found at Kgp elements L $\beta$ 2 $\alpha$ 3, L $\beta$ 4 $\beta$ 5, La5 $\alpha$ 6 and the interface between NSD and CSD, L $\beta$ 7 $\alpha$ 7, La7 $\beta$ 8, L $\beta$ 8 $\alpha$ 8, La10 $\beta$ 14, La11 $\alpha$ 12, La12 $\alpha$ 13, and La13 $\alpha$ 14. In particular, L $\beta$ 7 $\alpha$ 7 contains a long flap,  $G^{382}$ - $S^{391}$ , in RgpB—absent in Kgp—, which partially replaces the elongated cap  $G^{522}$ - $G^{539}$  of Kgp—in turn, missing in RgpB. Consistently, the two Kgp helices  $\alpha$ 11 and  $\alpha$ 12 are a continuous single helix in RgpB. Overall, these differences also entail that, while both structures share the two calcium sites of Kgp—though none of its sodium sites—an extra calcium site is found in RgpB at the region flanked by  $\alpha$ 7 and  $\alpha$ 11-La11 $\alpha$ 12- $\alpha$ 12 in the CSD of the CD. Noteworthy, while catalytic activity of RgpB is ablated by the calcium-chelator EDTA (83), Kgp remains unaffected (24). This supports that it is actually the extra calcium site of RgpB that is targeted by the chelator while the two common ones likely remain bound to the respective protein moieties. These differences also have a direct consequence for L $\beta$ 7 $\alpha$ 7, which provides the third putative catalytic acidic residue,  $D^{388}$  (see above). This loop protrudes slightly more in Kgp, thus explaining why an aspartate suffices in the latter to approach the catalytic histidine—while a glutamate is required in RgpB (see Fig. 5b,c)—as the positions of the catalytic cysteines and histidines nicely coincide in both structures (Fig. 5c).

Substantial differences are also found in the specificity pockets. Interestingly, although in both structures an aspartate salt-bridges the tip of the specific lysine or arginine ( $D^{516}$  in Kgp,  $D^{392}$  in RgpB,

respectively), none of them is placed at the bottom of the pocket but rather on the side, though on opposite walls of the pocket ( $D^{392}$  is close to  $G^{396}$  and  $D^{516}$  to  $P^{516}$ , see Fig. 5c), so the C $\gamma$  atoms of these aspartates are ~8Å apart. This entails that, while the lysine is bound in extended conformation in Kgp, in RgpB the arginine side chain is rotated clockwise around C $\delta$ -N $\epsilon$  by ~50° to meet  $D^{392}$  by means of a double salt bridge through its N $\eta$ 1 and N $\eta$ 2 atoms. This is enabled by the presence of  $V^{471}$  and  $M^{517}$  and by a rearrangement of RgpB region  $S^{507}$ - $Q^{520}$ - $S^{507}$ - $S^{520}$  in Kgp—, which relocates  $W^{513}$  and thus avoids a clash with the arginine (Fig. 5c). Summarizing, although the CD structures of Kgp and RgpB are quite similar, structural peculiarities account for their varying specificities, and help to explain distinct catalytic efficiencies against comparable substrates (68).

Mechanistically, the CDs of gingipains belong to MEROPS database family C25 (<http://merops.sanger.ac.uk>; (80,84)). They have been grouped together with families C11 (clostripain), C13 (legumain), C14 (caspases), C50 (separases), C80 (RTX self-cleaving toxin from *Vibrio cholerae*), and C84 (PrH peptidase from *Tannerella forsythia*) into clan CD. All these families share the following properties (80): broad distribution across all kingdoms of life; the catalytic histidine is found in a histidine-glycine motif preceded by a block of four hydrophobic residues (motif II according to (85)); and the catalytic cysteine is found in an alanine-cysteine motif (exceptionally cysteine-cysteine in Kgp) preceded by a second block of four hydrophobic residues (motif III according to (85)). In addition, they are specific for residues in position P<sub>1</sub> of substrates (arginine, lysine, asparagine or aspartate for the distinct families), and resistant to the broad-spectrum cysteine peptidase inhibitor E-64 but susceptible to chloromethylketone inhibitors (80). Structurally, despite being cysteine peptidases, Kgp and RgpB CDs are unlike any other clans and proteins structurally characterized to date. A structure-based search identified legumain as the closest structural relative of Kgp and RgpB (Fig. 5d) but it only matches the CSD and shows negligible sequence identity (PDB 4AW9, 4AWA, 4AWB; DALI Z-score=13.5; *rmsd*=2.8Å; length of alignment=173; seq. identity=14%). A certain similarity of the entire CD of Kgp and RgpB is also found with 2+2 heterotetramers of caspases (see Fig. 6 in (28)).

*Conclusion* — The present detailed structural analysis of an essential virulence factor of a major human periodontopathogen has revealed the molecular determinants of its mode of action and specificity. This may lay the foundations for the rational design of specific inhibitors—complementary to but distinct from those against RgpB—that may curtail the survival of the pathogen and palliate the effects of periodontal disease and its associated systemic disorders. This approach is complementary to the one aimed at developing inhibitors simultaneously targeting both types of gingipains (23).

## REFERENCES

1. Le Chatelier, E., Nielsen, T., Qin, J., Prifti, E., Hildebrand, F., Falony, G., Almeida, M., Arumugam, M., Batto, J. M., Kennedy, S., Leonard, P., Li, J., Burgdorf, K., Grarup, N., Jorgensen, T., Brandslund, I., Nielsen, H. B., Juncker, A. S., Bertalan, M., Levenez, F., Pons, N., Rasmussen, S., Sunagawa, S., Tap, J., Tims, S., Zoetendal, E. G., Brunak, S., Clement, K., Dore, J., Kleerebezem, M., Kristiansen, K., Renault, P., Sicheritz-Ponten, T., de Vos, W. M., Zucker, J. D., Raes, J., Hansen, T., Meta, H. I. T. c., Bork, P., Wang, J., Ehrlich, S. D., and Pedersen, O. (2013) Richness of human gut microbiome correlates with metabolic markers. *Nature* **500**, 541-546
2. May, M. (2014) Drug development: time for teamwork. *Nature* **509**, S4-S5
3. Hede, K. (2014) Antibiotic resistance: An infectious arms race. *Nature* **509**, S2-S3
4. Bostanci, N., and Belibasakis, G. N. (2012) *Porphyromonas gingivalis*: an invasive and evasive opportunistic oral pathogen. *FEMS Microbiol. Lett.* **333**, 1-9
5. Maixner, F., Thomma, A., Cipollini, G., Widder, S., Rattei, T., and Zink, A. (2014) Metagenomic analysis reveals presence of *Treponema denticola* in a tissue biopsy of the Iceman. *PLoS one* **9**, e99994
6. Kurita-Ochiai, T., and Yamamoto, M. (2014) Periodontal pathogens and atherosclerosis: implications of inflammation and oxidative modification of LDL. *Biomed. Res. Int.* **2014**, 595981
7. Maresz, K. J., Hellvard, A., Sroka, A., Adamowicz, K., Bielecka, E., Koziel, J., Gawron, K., Mizgalska, D., Marcinska, K. A., Benedyk, M., Pyrc, K., Quirke, A. M., Jonsson, R., Alzabin, S., Venables, P. J., Nguyen, K. A., Mydel, P., and Potempa, J. (2013) *Porphyromonas gingivalis* facilitates the development and progression of destructive arthritis through its unique bacterial peptidylarginine deiminase (PAD). *PLoS Pathog.* **9**, e1003627
8. Tilakaratne, A., and Soory, M. (2014) Anti-inflammatory actions of adjunctive tetracyclines and other agents in periodontitis and associated comorbidities. *Open Dent. J.* **8**, 109-124
9. Page, R. C. (1998) The pathobiology of periodontal diseases may affect systemic diseases: inversion of a paradigm. *Ann. Periodontol.* **3**, 108-120
10. Haffajee, A. D., Socransky, S. S., and Gunsolley, J. C. (2003) Systemic anti-infective periodontal therapy. A systematic review. *Ann. Periodontol.* **8**, 115-181
11. Matesanz-Pérez, P., García-Gargallo, M., Figuero, E., Bascones-Martínez, A., Sanz, M., and Herrera, D. (2013) A systematic review on the effects of local antimicrobials as adjuncts to subgingival debridement, compared with subgingival debridement alone, in the treatment of chronic periodontitis. *J. Clin. Periodontol.* **40**, 227-241
12. Potempa, J., and Pike, R. N. (2005) Bacterial peptidases. *Contrib. Microbiol.* **12**, 132-180
13. Mallorquí-Fernández, N., Manandhar, S. P., Mallorquí-Fernández, G., Usón, I., Wawrzonek, K., Kantyka, T., Solà, M., Thogersen, I. B., Enghild, J. J., Potempa, J., and Gomis-Rüth, F. X. (2008) A new autocatalytic activation mechanism for cysteine proteases revealed by *Prevotella intermedia* interpain A. *J. Biol. Chem.* **283**, 2871-2882
14. Dubin, G., Koziel, J., Pyrc, K., Wladyka, B., and Potempa, J. (2013) Bacterial proteases in disease - role in intracellular survival, evasion of coagulation/ fibrinolysis innate defenses, toxicoses and viral infections. *Current pharmaceutical design* **19**, 1090-1113
15. Lamont, R. J., and Jenkinson, H. F. (1998) Life below the gum line: pathogenic mechanisms of *Porphyromonas gingivalis*. *Microbiol. Mol. Biol. Rev.* **62**, 1244-1263
16. Potempa, J., and Travis, J. (1996) *Porphyromonas gingivalis* proteinases in periodontitis, a review. *Acta Biochim. Pol.* **43**, 455-465
17. Potempa, J., Pike, R., and Travis, J. (1997) Titration and mapping of the active site of cysteine proteinases from *Porphyromonas gingivalis* (gingipains) using peptidyl chloromethanes. *Biol. Chem.* **378**, 223-230
18. Pike, R. N., and Potempa, J. (2013) Chapter 523 - Gingipain K. in *Handbook of Proteolytic Enzymes* (Rawlings, N. D., and Salvesen, G. eds.), Academic Press, Oxford. pp 2337-2344
19. Brochu, V., Grenier, D., Nakayama, K., and Mayrand, D. (2001) Acquisition of iron from human transferrin by *Porphyromonas gingivalis*: a role for Arg- and Lys-gingipain activities. *Oral Microbiol. Immunol.* **16**, 79-87
20. Pathirana, R. D., O'Brien-Simpson, N. M., Brammar, G. C., Slakeski, N., and Reynolds, E. C. (2007) Kgp and RgpB, but not RgpA, are important for *Porphyromonas gingivalis* virulence in the murine periodontitis model. *Infect. Immun.* **75**, 1436-1442
21. Yongqing, T., Potempa, J., Pike, R. N., and Wijeyewickrema, L. C. (2011) The lysine-specific gingipain of *Porphyromonas gingivalis*: importance to pathogenicity and potential strategies for inhibition. *Adv. Exp. Med. Biol.* **712**, 15-29
22. Curtis, M. A., Aduse Opoku, J., Rangarajan, M., Gallagher, A., Sterne, J. A. C., Reid, C. R., Evans, H. E. A., and Samuelsson, B. (2002) Attenuation of the virulence of *Porphyromonas gingivalis* by using a specific synthetic Kgp protease inhibitor. *Infect. Immun.* **70**, 6968-6975
23. Kataoka, S., Baba, A., Suda, Y., Takii, R., Hashimoto, M., Kawakubo, T., Asao, T., Kadowaki, T., and Yamamoto, K. (2014) A novel, potent dual inhibitor of Arg-gingipains and Lys-gingipain as a promising agent for periodontal disease therapy. *FASEB J.* **28**, 3564-3578
24. Pike, R., McGraw, W., Potempa, J., and Travis, J. (1994) Lysine- and arginine-specific proteinases from *Porphyromonas gingivalis*. Isolation, characterization, and evidence for the existence of complexes with hemagglutinins. *J. Biol. Chem.* **269**, 406-411
25. Sztukowska, M., Veillard, F., Potempa, B., Bogyo, M., Enghild, J. J., Thogersen, I. B., Nguyen, K. A., and Potempa, J. (2012) Disruption of gingipain oligomerization into non-covalent cell-surface attached complexes. *Biol. Chem.* **393**, 971-977
26. Mittl, P. R., and Grütter, M. G. (2006) Opportunities for structure-based design of protease-directed drugs. *Curr. Opin. Struct. Biol.* **16**, 769-775
27. de Diego, I., Veillard, F. T., Guevara, T., Potempa, B., Sztukowska, M., Potempa, J., and Gomis-Rüth, F. X. (2013) *Porphyromonas gingivalis* virulence factor gingipain RgpB shows a unique zymogenic mechanism for cysteine peptidases. *J. Biol. Chem.* **288**, 14287-14296
28. Eichinger, A., Beisel, H. G., Jacob, U., Huber, R., Medrano, F. J., Banbula, A., Potempa, J., Travis, J., and Bode, W. (1999) Crystal structure of gingipain R: an Arg-specific bacterial cysteine proteinase with a caspase-like fold. *EMBO J.* **18**, 5453-5462

29. Li, N., Yun, P., Jeffries, C. M., Langley, D., Gamsjaeger, R., Church, W. B., Hunter, N., and Collyer, C. A. (2011) The modular structure of haemagglutinin/adhesin regions in gingipains of *Porphyromonas gingivalis*. *Molecular microbiology* **81**, 1358-1373
30. Li, N., Yun, P., Nadkarni, M. A., Ghadikolaei, N. B., Nguyen, K. A., Lee, M., Hunter, N., and Collyer, C. A. (2010) Structure determination and analysis of a haemolytic gingipain adhesin domain from *Porphyromonas gingivalis*. *Molecular microbiology* **76**, 861-873
31. Potempa, J., and Nguyen, K. A. (2007) Purification and characterization of gingipains. *Curr. Protoc. Prot. Sci.* **Chapter 21**, Unit 21 20
32. Matthews, B. W. (1968) Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497
33. Trillo-Muyo, S., Jasilionis, A., Domagalski, M. J., Chruszcz, M., Minor, W., Kuisiene, N., Arolas, J. L., Solà, M., and Gomis-Ruth, F. X. (2013) Ultratight crystal packing of a 10-kDa protein. *Acta Crystallogr. sect. D* **69**, 464-470
34. Kabsch, W. (2010) XDS. *Acta Crystallogr. sect. D* **66**, 125-132
35. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658-674
36. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. sect. D* **67**, 235-242
37. Corpet, F. (1988) Multiple sequence alignment with hierarchical clustering. *Nucl. Acids Res.* **16**, 10881-10890
38. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171-1179
39. Carranza, C., Inisan, A.-G., Mouthuy-Knoops, E., Cambillau, C., and Roussel, A. (1999) Turbo-Frodo. in *AFMB Activity Report 1996-1999*, CNRS-UPR 9039, Marseille. pp 89-90
40. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. sect. D* **66**, 486-501
41. Blanc, E., Roversi, P., Vonnrhein, C., Flensburg, C., Lea, S. M., and Bricogne, G. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. *Acta Crystallogr. sect. D* **60**, 2210-2221
42. Di Tommaso, P., Moretti, S., Xenarios, I., Orobitch, M., Montanyola, A., Chang, J. M., Taly, J. F., and Notredame, C. (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucl. Acids Res.* **39**, W13-W17
43. Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucl. Acids Res.* **42**, W320-W324
44. Schüttelkopf, A. W., and van Aalten, D. M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. sect. D* **60**, 1355-1363
45. Holm, L., and Rosenström, P. (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545-W549
46. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605-1612
47. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. sect. D* **66**, 12-21
48. Ollis, D. L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S. M., Harel, M., Remington, S. J., Silman, I., Schrag, J., Sussman, J. L., Verschueren, K. H. G., and Goldman, A. (1992) The  $\alpha/\beta$  hydrolase fold. *Prot. Eng.* **5**, 197-211
49. Puente, X. S., and López-Otín, C. (1997) The PLEES-proteins - A family of structurally related enzymes widely distributed from bacteria to humans. *Biochem. J.* **332**, 947-949
50. Fraústo da Silva, J. J. R., and Williams, R. J. P. (2001) *The biological chemistry of the elements: the inorganic chemistry of life.*, Second Edition ed., Oxford University Press Inc., New York
51. Richardson, J. S. (1981) The anatomy and taxonomy of protein structure. *Adv. Prot. Chem.* **34**, 167-339
52. Harding, M. M. (2000) The geometry of metal-ligand interactions relevant to proteins. II. Angles at the metal atom, additional weak metal-donor interactions. *Acta Crystallogr. sect. D* **56**, 857-867
53. Harding, M. M. (2006) Small revisions to predicted distances around metal sites in proteins. *Acta Crystallogr. sect. D* **62**, 678-682
54. Richardson, J. S., and Richardson, D. C. (1989) Principles and patterns of protein conformation. in *Prediction of protein structure and the principles of protein conformation.* (Fasman, G. D. ed.), Plenum Press, New York. pp 1-98
55. Harding, M. M. (2002) Metal-ligand geometry relevant to proteins and in proteins: sodium and potassium. *Acta Crystallogr. sect. D* **58**, 872-874
56. Vidossich, P., Loewen, P. C., Carpena, X., Fiorin, G., Fita, I., and Rovira, C. (2014) Binding of the antitubercular pro-drug isoniazid in the heme access channel of catalase-peroxidase (KatG). A combined structural and metadynamics investigation. *J. Phys. Chem. B* **118**, 2924-2931
57. Nakatogawa, H., and Ito, K. (2002) The ribosomal exit tunnel functions as a discriminating gate. *Cell* **108**, 629-636
58. Williams, M. A., Goodfellow, J. M., and Thornton, J. M. (1994) Buried waters and internal cavities in monomeric proteins. *Prot. Sci.* **3**, 1224-1235
59. Rose, A., Theune, D., Goede, A., and Hildebrand, P. W. (2014) MP:PD--a data base of internal packing densities, internal packing defects and internal waters of helical membrane proteins. *Nucl. Acids Res.* **42**, D347-D351
60. Hildebrand, P. W., Gunther, S., Goede, A., Forrest, L., Frommel, C., and Preissner, R. (2008) Hydrogen-bonding and packing features of membrane proteins: functional implications. *Biophys. J.* **94**, 1945-1953
61. Lee, C., Park, S. H., Lee, M. Y., and Yu, M. H. (2000) Regulation of protein function by native metastability. *Proc. Natl. Acad. Sci. USA* **97**, 7727-7731
62. Takano, K., Yamagata, Y., and Yutani, K. (2003) Buried water molecules contribute to the conformational stability of a protein. *Prot. Eng.* **16**, 5-9
63. Dunitz, J. D. (1994) The entropic cost of bound water in crystals and biomolecules. *Science* **264**, 670-670
64. Zhang, L., and Hermans, J. (1996) Hydrophilicity of cavities in proteins. *Proteins* **24**, 433-438

65. Kokkinidis, M., Glykos, N. M., and Fadouloglou, V. E. (2012) Protein flexibility and enzymatic catalysis. *Adv. Protein. Chem. Struct. Biol.* **87**, 181-218
66. Schulenburg, C., and Hilvert, D. (2013) Protein conformational disorder and enzyme catalysis. *Top. Curr. Chem.* **337**, 41-67
67. Potempa, J., Pike, R., and Travis, J. (1995) The multiple forms of trypsin-like activity present in various strains of *Porphyromonas gingivalis* are due to the presence of either Arg-gingipain or Lys-gingipain. *Infect. Immun.* **63**, 1176-1182
68. Sztukowska, M., Sroka, A., Bugno, M., Banbula, A., Takahashi, Y., Pike, R. N., Genco, C. A., Travis, J., and Potempa, J. (2004) The C-terminal domains of the gingipain K polyprotein are necessary for assembly of the active enzyme and expression of associated activities. *Mol. Microbiol.* **54**, 1393-1408
69. Chothia, C., and Jones, E. Y. (1997) The molecular structure of cell adhesion molecules. *Annu. Rev. Biochem.* **66**, 823-862
70. Discipio, R. G., Daffern, P. J., Kawahara, M., Pike, R., Travis, J., Hugli, T. E., and Potempa, J. (1996) Cleavage of human complement component C5 by cysteine proteinases from *Porphyromonas (Bacteroides) gingivalis*. Prior oxidation of C5 augments proteinase digestion of C5. *Immunology* **87**, 660-667
71. Marrero, A., Duquerroy, S., Trapani, S., Goulas, T., Guevara, T., Andersen, G. R., Navaza, J., Sottrup-Jensen, L., and Gomis-Rüth, F. X. (2012) The crystal structure of human  $\alpha$ 2-macroglobulin reveals a unique molecular cage. *Angew. Chem. Int. Ed.* **51**, 3340-3344
72. Ishida, Y., Hu, J., Sakai, E., Kadowaki, T., Yamamoto, K., Tsukuba, T., Kato, Y., Nakayama, K., and Okamoto, K. (2008) Determination of active site of lysine-specific cysteine proteinase (Lys-gingipain) by use of a *Porphyromonas gingivalis* plasmid system. *Arch. Oral Biol.* **53**, 538-544
73. Shaw, E., Mares-Guia, M., and Cohen, W. (1965) Evidence for an active-center histidine in trypsin through use of a specific reagent, 1-chloro-3-tosylamido-7-amino-2-heptanone, the chloromethyl ketone derived from *N* $\alpha$ -tosyl-L-lysine. *Biochemistry* **4**, 2219-2224
74. Frydrych, I., and Mlejnek, P. (2008) Serine protease inhibitors N-alpha-tosyl-L-lysiny-chloromethylketone (TLCK) and N-tosyl-L-phenylalaninyl-chloromethylketone (TPCK) are potent inhibitors of activated caspase proteases. *J. Cell. Biochem.* **103**, 1646-1656
75. Drenth, J., Kalk, K. H., and Swen, H. M. (1976) Binding of chloromethyl ketone substrate analogues to crystalline papain. *Biochemistry* **15**, 3731-3738
76. Jia, Z., Hasnain, S., Hirama, T., Lee, X., Mort, J. S., To, R., and Huber, C. P. (1995) Crystal structures of recombinant rat cathepsin B and a cathepsin B-inhibitor complex. Implications for structure-based inhibitor design. *J. Biol. Chem.* **270**, 5527-5533
77. Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J., Powers, J. C., and Wilcox, P. E. (1972) An x-ray crystallographic study of the binding of peptide chloromethyl ketone inhibitors to subtilisin BPN'. *Biochemistry* **11**, 2439-2449
78. Bode, W., Mayr, I., Baumann, U., Huber, R., Stone, S. R., and Hofsteenge, J. (1989) The refined 1.9 Å crystal structure of human  $\alpha$ -thrombin: interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J.* **8**, 3467-3475
79. Schechter, I., and Berger, A. (1967) On the size of active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **27**, 157-162
80. Chen, J.-M., Rawlings, N. D., Stevens, R. A., and Barrett, A. J. (1998) Identification of the active site of legumain links it to caspases, clostripain and gingipains in a new clan of cysteine endopeptidases. *FEBS Lett.* **441**, 361-365
81. Polgár, L. (2013) Chapter 405 - Catalytic mechanisms of cysteine peptidases. in *Handbook of Proteolytic Enzymes* (Rawlings, N. D., and Salvesen, G. S. eds.), 3rd Ed., Academic Press, Oxford. pp 1773-1784
82. Guarné, A., Hampoelz, B., Glaser, W., Carpena, X., Tormo, J., Fita, I., and Skern, T. (2000) Structural and biochemical features distinguish the foot-and-mouth disease virus leader proteinase from other papain-like enzymes. *J. Mol. Biol.* **302**, 1227-1240
83. Chen, Z., Potempa, J., Polanowski, A., Wikström, M., and Travis, J. (1992) Purification and characterization of a 50-kDa cysteine proteinase (gingipain) from *Porphyromonas gingivalis*. *J. Biol. Chem.* **267**, 18896-18901
84. Rawlings, N. D., Waller, M., Barrett, A. J., and Bateman, A. (2014) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucl. Acids Res.* **42**, D503-D509
85. Aravind, L., and Koonin, E. V. (2002) Classification of the caspase-hemoglobinase fold: detection of new families and implications for the origin of the eukaryotic separins. *Proteins* **46**, 355-367
86. Dall, E., and Brandstetter, H. (2013) Mechanistic and structural studies on legumain explain its zymogenicity, distinct activation pathways, and regulation. *Proc. Natl. Acad. Sci. USA* **110**, 10940-10945
87. Weiss, M. S. (2001) Global indicators of X-ray quality. *J. Appl. Cryst.* **34**, 130-135
88. Karplus, P. A., and Diederichs, K. (2012) Linking crystallographic model and data quality. *Science* **336**, 1030-1033

## ACKNOWLEDGMENTS

We are grateful to the joint IBMB/IRB Automated Crystallography Platform for assistance during crystallization experiments and to Robin Rycroft for very helpful contributions to the manuscript. We acknowledge the help provided by local contacts at the ESRF synchrotron.

## FOOTNOTES

\*\*This study was financially supported in part by grants from European, US American, Polish, Spanish, and Catalan agencies (UMO-2012/04/A/NZ1/00051, 2137/7.PR-EU/2011/2, DE09761; FP7-HEALTH-2010-261460 “Gums&Joints”; FP7-PEOPLE-2011-ITN-290246 “RAPID”; FP7-HEALTH-2012-306029-2 “TRIGGER”; BFU2012-32862; BIO2013-49320-EXP; CSD2006-00015; and 2014SGR9). Funding for data collection was provided in part by ESRF. The Faculty of Biochemistry, Biophysics and Biotechnology of the Jagiellonian University is a beneficiary of structural funds from the European Union (POIG.02.01.00-12-064/08).

**Abbreviations:** CD, catalytic domain; CSD, C-terminal sub-domain; IgSF, immunoglobulin superfamily-like domain; Kgp, lysine-specific gingipain; KgpA<sub>A1</sub>- KgpA<sub>A5</sub>, Kgp hemagglutinin/adhesion domains A1 to A5; LM, L-lysylmethyl; NSD, N-terminal sub-domain; PAC, Automated Crystallography Platform at IBMB/IRB; PDB, Protein data Bank; RgpA, RgpB, arginine-specific gingipains A and B; RgpA<sub>A1</sub>- RgpA<sub>A4</sub>, RgpA hemagglutinin/adhesion domains A1 to A4; and TLCK, *N*α-tosyl-L-lysylchloromethane.

## FIGURE LEGENDS

**Figure 1 – Gingipain sequences.** Structure-guided sequence alignment of the CD and IgSF moieties (separately framed) of Kgp from *P. gingivalis* strain W83 (UniProt Q51817; top rows) and RgpB from *P. gingivalis* strain HG66 (GenBank AAB41892; PDB 1CVR, (28); bottom rows). The sequence of the latter differs from that of the ortholog from strain W83 at 12 positions (UniProt 95493; PDB 4IEF; (27)). The amino-acid numbering and the regular secondary structure elements (strands as black arrows labeled β1-β21 and helices as loops labeled α1-α14) above the alignment correspond to the Kgp(CD+IgSF) structure (this study), those below the alignment correspond to RgpB(CD+IgSF)(taken from Fig. 2c of (28)). Identical residues are in bold white over black background, similar residues are in bold black over gray, and the overall sequence identity is 27%. (Potential) catalytic residues of Kgp CD are pinpointed by an open rhombus; residues framing the S<sub>1</sub> pocket are indicated by an arrow.

**Figure 2 – General architecture of KgpB.** (A) Ribbon-type plot of Kgp(CD+IgSF) showing the regular secondary-structure elements (CD, α-helices as orange ribbons, β-strand as yellow arrows and coils in tan; IgSF, β-strand as blue arrows labeled β16-β21 and coils in lilac). The N- and C-termini and the four structural cations (calciums Ca<sup>999</sup> and Ca<sup>998</sup> in red, sodiums Na<sup>997</sup> and Na<sup>996</sup> in blue) are depicted. The (putative) catalytic triad—C<sup>477</sup> (covalently modified at Sy with an L-lysylmethyl group, carbons in lilac), H<sup>444</sup> and D<sup>388</sup>—is further shown as sticks. Black arrows point to the solvent channel (see Fig. 3d). (B) Picture of a tooth with its parts labeled. (C) (2mF<sub>obs</sub>-DF<sub>calc</sub>)-type Fourier map of the region around the catalytic cysteine, C<sup>477</sup>, obtained with diffraction data to 1.75Å resolution and contoured at 1σ above threshold. (D) Structure of Kgp(CD+IgSF) in cross-eye stereo in standard orientation (28,79), which corresponds to a horizontal 90°-rotation of the view in (A), i.e viewing the CD cusp region. Regular secondary-structure elements of the CD (helices α1-α14 and strands β1-β15) are labeled. The NSD is on top, the CSD is at the bottom, see also Fig. 5d. (E) Close-up of (D) in stereo centered on the non-primed side of the active site. Residues framing the specificity pocket S<sub>1</sub> and pocket S<sub>2</sub> are labeled. Small green spheres represent solvent molecules.

**Figure 3 – Ion sites and solvent channel.** (A) Detail of the Kgp(CD+IgSF) structure around Ca<sup>999</sup> shown as a red sphere. Some residues are labeled for reference; solvents are depicted as green spheres. Liganding atoms are linked with a solid line and the respective binding distances (in Å) are indicated. (B) Same as (A) but for Ca<sup>998</sup>. (C) Same as (A) but for the two sodium sites Na<sup>997</sup> and Na<sup>996</sup> (blue spheres). (D) Detail in stereo of the semi-transparent surface of Kgp CD (solvent radius=1.4Å) and the solvent channel ranging from the bottom of the specificity pocket (pinpointed by the tip of the L-lysylmethyl group attached to the catalytic cysteine C<sup>477</sup>) to the opposite surface of the molecule (pinpointed by a solvent molecule in red). The solvent molecules found in the structure are depicted as green spheres. Protein atoms have been omitted for clarity. The orientation displayed corresponds to that of Fig. 2d after a vertical cw rotation of 120° and a ccw 90°-rotation in the plane.

**Figure 4 – Interactions in the S<sub>1</sub> pocket.** Scheme depicting relevant interactions (as curved lines) of the L-lysylmethyl group (bold trace) in the specificity pocket. The three solvent molecules in the pocket are shown as white spheres, the first solvent of the large inner solvent channel is shown as a grey sphere (see also Fig. 3d).

**Figure 5 – Structural similarities.** (A) Ribbon-plot in stereo showing the superposition of Kgp(CD+IgSF) (lilac+light pink) and RgpB(CD+IgSF) of strain HG66 (cyan+green; PDB 1CVR; (28)). The two calcium and two sodium ions of Kgp are depicted as red and blue spheres, respectively, and the three calcium ions of RgpB as green spheres. The catalytic active-site residues are shown as sticks for each structure, with carbons colored in the respective ribbon colors. (B) Superposition as in (A) but in standard orientation. (C) Close-up of (B) focusing into the active sites. RgpB is covalently modified at its catalytic cysteine Sy atom by a Phe-Phe-Arg-CH<sub>2</sub>- moiety. Shown are the catalytic cysteines (①; C<sup>477</sup> in Kgp, C<sup>473</sup> in RgpB) and histidines (②; H<sup>444</sup> in Kgp, H<sup>473</sup> in RgpB), and the putative histidine-polarizing acidic residues (③; D<sup>388</sup> in Kgp, E<sup>381</sup> in RgpB), as well as some residues engaged in P<sub>1</sub>-pocket framing (④, N<sup>510</sup> in Kgp and D<sup>510</sup> in RgpB; ⑤, T<sup>442</sup> in Kgp and T<sup>438</sup> in RgpB; ⑥, W<sup>513</sup> in Kgp and W<sup>513</sup> in RgpB; ⑦, S<sup>511</sup> in Kgp and Q<sup>511</sup> in RgpB), and the aspartates salt-bridging the basic residues in P<sub>1</sub> (⑧, D<sup>516</sup> in Kgp; ⑨, D<sup>392</sup> in RgpB). (D) Superposition in stereo of Kgp CD (NSD in tan, CSD in lilac) onto legumain (in turquoise; PDB 4AWA, (86)) in standard orientation.

**Table 1.** Crystallographic data.

Space group / cell constants (a,b, and c, in Å)	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub> / 56.64, 58.81, 135.50
Wavelength (Å)	0.9393
No. of measurements / unique reflections	303,645 / 46,309
Resolution range (Å) (outermost shell)	45.2 – 1.75 (1.79 – 1.75)
Completeness (%)	99.5 (93.9)
R <sub>merge</sub> <sup>a</sup>	0.100 (0.692)
R <sub>r.i.m.</sub> (= R <sub>meas</sub> ) <sup>a</sup> / CC(1/2) <sup>b</sup>	0.108 (0.821) / 0.998 (0.656)
Average intensity over stand. dev. (<[<I> / σ(<I>)]>)	16.5 (1.9)
B-Factor (Wilson) (Å <sup>2</sup> ) / Average multiplicity	22.0 / 6.6 (3.4)
Resolution range used for refinement (Å)	44.1 – 1.75
No. of reflections in working set / in test set	45,532 / 767
Crystallographic R <sub>factor</sub> (free R <sub>factor</sub> ) <sup>c</sup>	0.149 (0.172)
No. of protein atoms / solvent molecules / ions / ligands	3,534 / 533 / 2 Ca <sup>2+</sup> , 1 Ni <sup>2+</sup> , 3 Na <sup>+</sup> / 1 histidine, 2 acetate, 3 azide, 4 glycerol
<i>Rmsd</i> from target values	
bond lengths (Å) / bond angles (°)	0.010 / 1.00
Average B-factor (Å <sup>2</sup> ) protein atoms / CD only / IgSF only	17.1 / 13.6 / 34.5
Main-chain conformational angle analysis <sup>d</sup>	
Residues in favored regions / outliers / all residues	436 / 1 / 449
Values in parentheses refer to the outermost resolution shell.	
<sup>a</sup> R <sub>merge</sub> = $\sum_{hkl} \sum_i  I_i(hkl) - \langle I(hkl) \rangle  / \sum_{hkl} \sum_i I_i(hkl)$ ; R <sub>r.i.m.</sub> = $\sum_{hkl} (n_{hkl} / [n_{hkl} - 1])^{1/2} \sum_i  I_i(hkl) - \langle I(hkl) \rangle  / \sum_{hkl} \sum_i I_i(hkl)$ ; R <sub>p.i.m.</sub> = $\sum_{hkl} (1 / [n_{hkl} - 1])^{1/2} \sum_i  I_i(hkl) - \langle I(hkl) \rangle  / \sum_{hkl} \sum_i I_i(hkl)$ , where I <sub>i</sub> (hkl) is the i-th intensity measurement and n <sub>hkl</sub> the redundancy of reflection hkl—including symmetry-related reflections—and <I(hkl)> its average intensity. R <sub>r.i.m.</sub> ( <i>alias</i> R <sub>meas</sub> ) and R <sub>p.i.m.</sub> are improved multiplicity-weighted indicators of the quality of the data, the redundancy-independent merging R factor and the precision-indicating merging R factor. The latter is computed after averaging over multiple measurements (for details, see (87)).	
<sup>b</sup> According to Karplus & Diederichs (88). <sup>c</sup> Crystallographic R <sub>factor</sub> = $\sum_{hkl}   F_{obs}  - k  F_{calc}   / \sum_{hkl}  F_{obs} $ , where k is a scaling factor, and F <sub>obs</sub> and F <sub>calc</sub> are the observed and calculated structure factor amplitudes, respectively. This factor is calculated for the working-set reflections; free R <sub>factor</sub> <sup>c</sup> same for a test-set of reflections (>500) not used during refinement. <sup>d</sup> According to MOLPROBITY (47).	



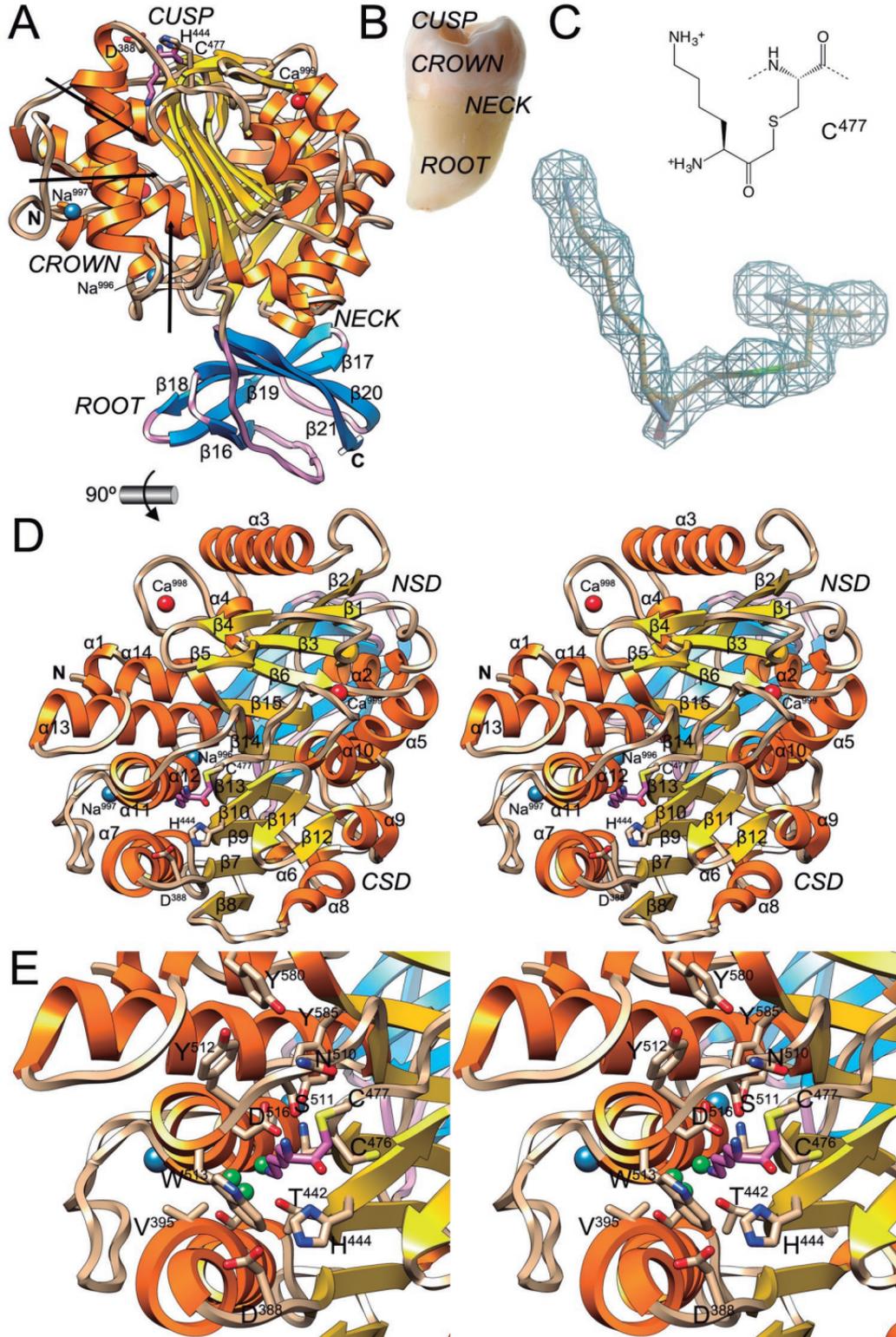


Fig. 2

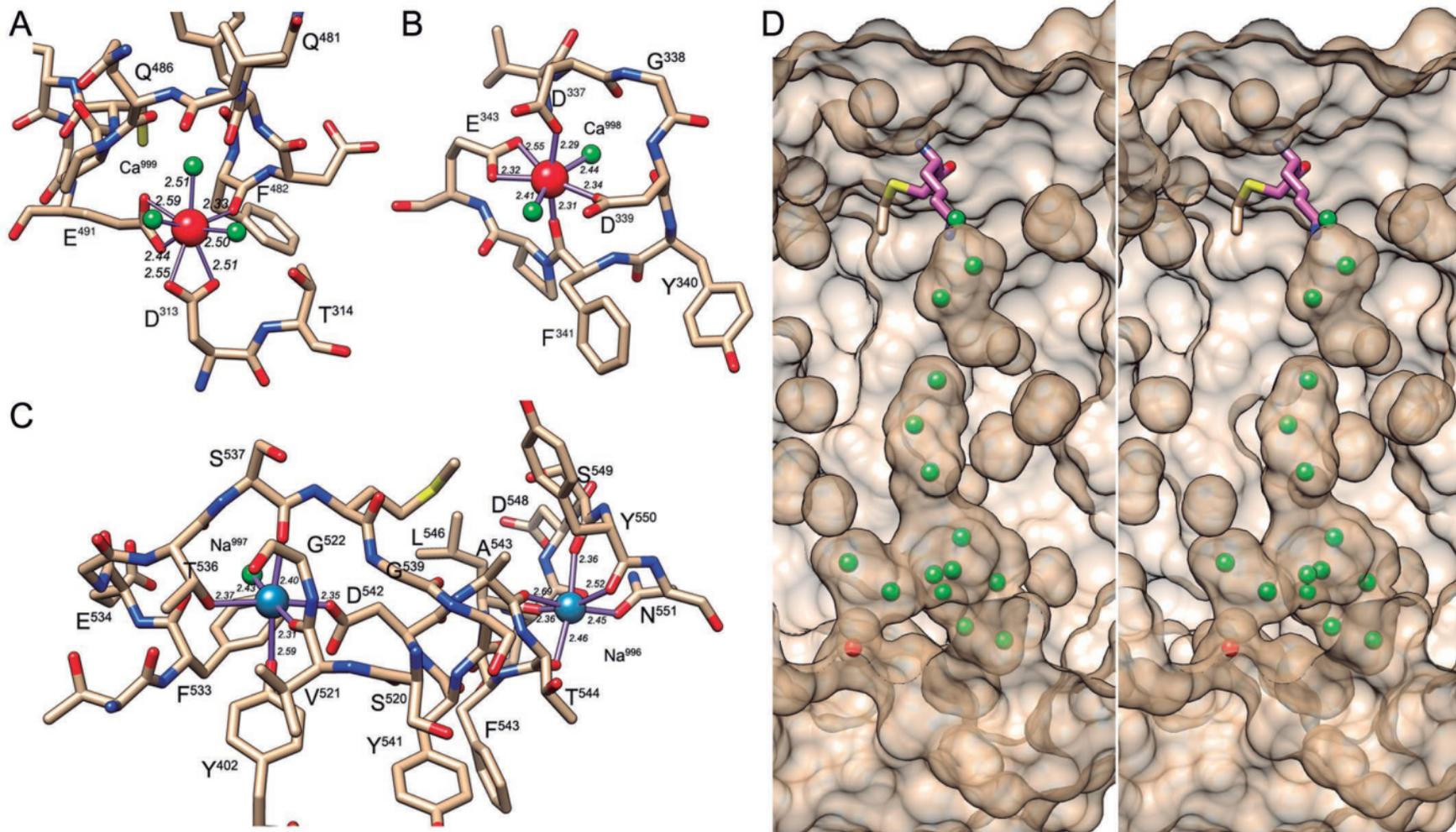


Fig. 3

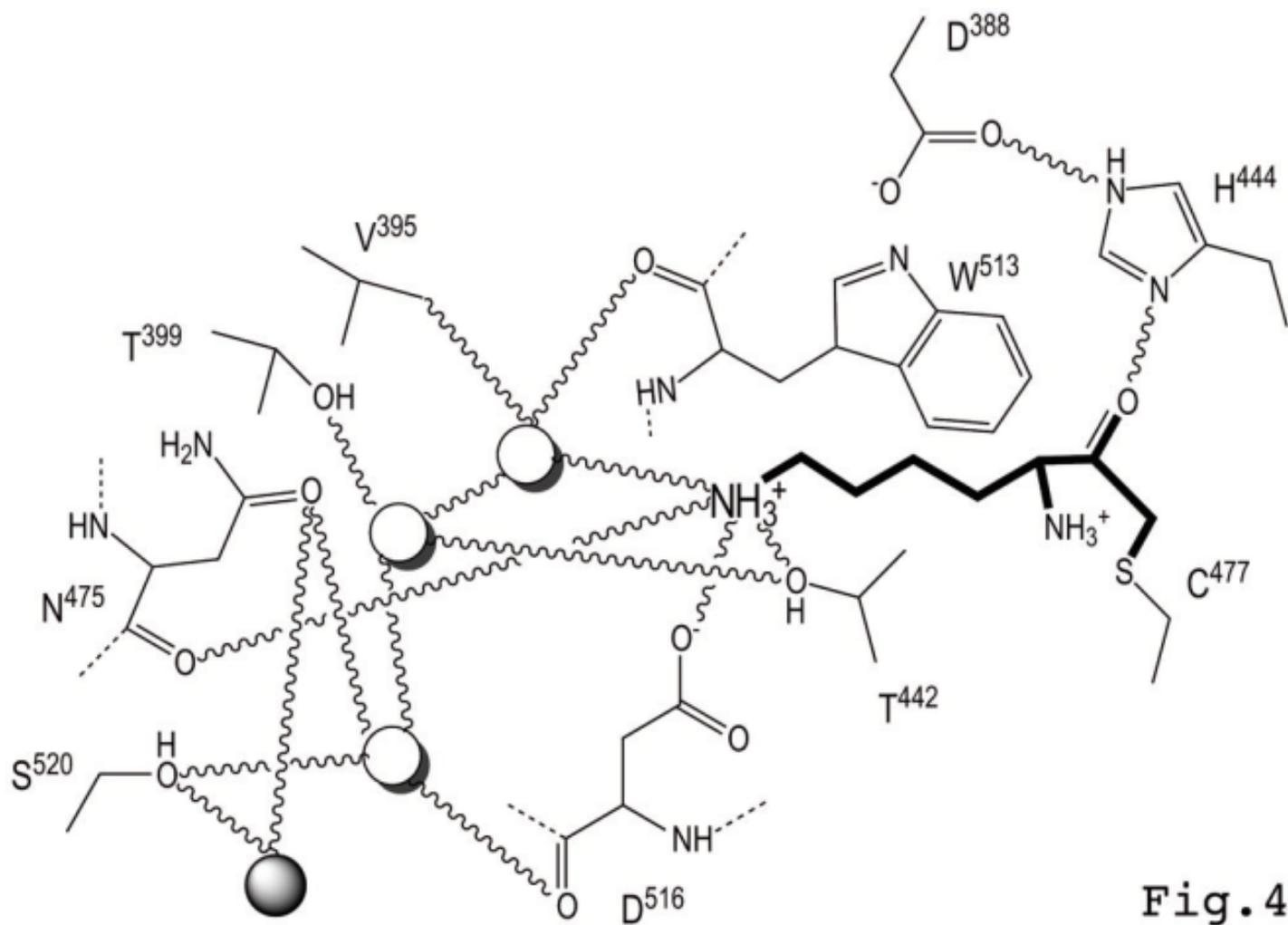


Fig.4

