Porphyromonas gingivalis virulence factor gingipain RgpB shows a unique zymogenic mechanism for cysteine peptidases **

Iñaki de Diego [1], Florian T. Veillard [2], Tibisay Guevara [1], Barbara Potempa [2], Maryta Sztukowska [2], Jan Potempa [2,3,*] and F.Xavier Gomis-Rüth [1,*]

Running title: Structure of the zymogenic complex of gingipain RgpB

[1] Molecular Biology Institute of Barcelona; Spanish Research Council CSIC; Barcelona Science Park; c/Baldiri Reixac,15-21; 08028 Barcelona, Catalonia (Spain).

[2] University of Louisville School of Dentistry; Louisville, KY (US).

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

[*] Corresponding authors. E-mail: xgrcri@ibmb.csic.es and jspote01@louisville.edu.

Keywords: latency; pro-peptidase; cysteine protease; inhibition; X-ray crystal structure

---

Background: The odontopathogenic virulence factor gingipain RgpB is produced as a zymogen to prevent intracellular activity prior to secretion.

Results: The structure of the complex between the pro-domain and the catalytic moiety of RgpB has been determined.

Conclusion: RgpB is kept latent by a novel molecular mechanism.

Significance: The structural details should enable to design small-molecule inhibitors to inhibit RgpB in a non-covalent manner**.**

SUMMARY

Zymogenicity is a regulatory mechanism which prevents inadequate catalytic activity in the wrong context. It plays a central role in maintaining microbial virulence factors in an inactive form inside the pathogen until secretion. Among these virulence factors is the cysteine peptidase gingipain B (RgpB), which is the major virulence factor secreted by the periodontopathogen *Porphyromonas gingivalis* that attacks host vasculature and defense proteins. The structure of the complex between soluble mature RgpB, consisting of a catalytic domain and an immunoglobulin-superfamily domain, and its 205-residue N-terminal pro-domain, the largest structurally characterized to date for a cysteine peptidase, reveals a novel fold for the pro-domain that is distantly related to sugar-binding lectins. It attaches laterally to the catalytic domain through a large concave surface. The main determinant for latency is a surface "inhibitory loop", which approaches the active-site cleft of the enzyme on its non-primed side in a substrate-like manner. It inserts an arginine (R^{126}) into the S_1 pocket, thus matching the substrate specificity of the enzyme. Downstream of R^{126}, the polypeptide leaves the cleft, thereby preventing cleavage. Moreover, the carbonyl group of R^{126} establishes a very strong hydrogen bond with the co-catalytic histidine, H^{440}, pulling it away from the catalytic cysteine, C^{473}, and towards E^{381}, which probably plays a role orienting the side chain of H^{440} during catalysis. The present results provide the structural determinants of zymogenic inhibition of RgpB by way of a novel inhibitory mechanism for peptidases in general and open the field for the design of novel inhibitory strategies in the treatment of human periodontal disease.

Periodontitis is a biofilm-associated chronic inflammatory disease of the gums caused by bacterial infection, which affects 10-15% of adults worldwide and may result in tooth loss (1,2). Of the several hundreds of bacterial species that colonize the oral cavity, the key disease pathogens are *Aggregatibacter (Actinobacillus) actinomycetemcomitans*, *Tannerella forsythia*, *Treponema denticola*, and *Porphyromonas gingivalis*, the latter three forming the red complex, which is implicated in severe forms of the disease (3-5). *P. gingivalis* is an opportunistic pathogen found in up to 85% of periodontitis cases and its presence at the infection site is indicative of disease progression (6,7). The pathogen requires nutrients, such as heme or vitamin K, and anaerobic conditions for growth (2). As part of the process of infection, *P. gingivalis* invades host epithelial cells and macrophages, where it affects cell-cycle pathways and suppresses apoptosis, thereby circumventing the host immune response and prolonging its survival (2,8). The pathogen possesses several factors which participate in infection, such as the lipopolysaccharide, the capsular polysaccharide, the fimbriae, and, most importantly, cysteine proteinases, viz. gingipains K (Kgp) and R (RgpA and RgpB) (2,9,10). While the former factors are intrinsic components of the outer membrane of the pathogen, gingipains are true cell-surface-anchored or soluble virulence factors that, when secreted, account for up to 85% of the total extracellular proteolytic activity of *P. gingivalis* (10,11). This activity is aimed at obtaining nutrients, cleavage of host-cell surface receptors, stimulation of protease-activated receptor

expression, and inactivation of cytokines and components of the complement system. These functions contribute to resistance of the pathogen to host bactericidal activity and maintenance of the chronic inflammatory condition at the site of infection (2). In addition, gingipains contribute to bleeding and vascular permeability by activating plasma kallikrein, degrading fibrinogen, and increasing the levels of thrombin and prothrombin, thus increasing the availability of heme needed for bacterial growth (12). These functions explain why gingipains are essential for bacterial survival and the pathological outcome of periodontitis (3,10,13).

As occurs with most proteolytic enzymes, the activity of gingipains must be regulated to prevent undesired intracellular proteolysis yet yield full activity once secreted (14). In general, such activity control occurs at the transcriptional level, through compartmentalization or allostery, or through inhibition by specific protein inhibitors. Another regulatory mechanism is zymogenic latency (15), which is observed for gingipains, and is carried out mostly by N-terminal pro-peptides or pro-domains (PDs). These usually prevent substrates from binding to the active-site cleft of the cognate catalytic domain (CD) and are mostly removed by limited proteolysis during maturation (15-17). Such PDs often fold independently and guide on their part the folding process of the CD (18); they may also act as intra-molecular chaperones or inhibitors of the mature enzymes *in trans*—as described for RgpB (19)—and in the intracellular sorting of the zymogen (15). Therefore, the study of the molecular mechanisms by which peptidases maintain latency is indispensable to the understanding of their basic mode of action. It also paves the way for the design of inhibitors that mimic the latent state so as to modulate proteolytic activity as part of a therapeutic approach. Detailed 3D structural information can contribute much to this understanding (20).

Among gingipains, Kgp is specific for peptide bonds after lysines, while RgpA and RgpB are arginine-specific (3,21). These enzymes are multi-domain proteins comprising at least a signal peptide, a PD, a CD, an immunoglobulin-superfamily domain (IgSF), and a C-terminal domain, as found in the 736-residue RgpB spanning, respectively, 24, 205, 351, 87, and 69 residues (3). RgpA has four additional hemagglutinin/adhesion domains (termed $RgpA_{A1}$-$RgpA_{A4}$) inserted between the IgSF and the C-terminal domain, thus totaling 1,706 residues. Kgp may have between three and five such domains (termed $KgpA_{A1}$-$KgpA_{A5}$) depending on the bacterial strain, thus totaling 1,723-1,732 residues (3). The CDs of gingipains distantly resemble caspase cysteine proteinases as revealed by the crystal structure of mature RgpB of *P. gingivalis* strain HG66 (22), and are grouped into MEROPS database family C25 (http://merops.sanger.ac.uk; (23)).

In order to understand the biochemical determinants of zymogenicity in gingipains, we analyzed the structure of the complex between the mature enzyme moiety (CD+IgSF domains) and the PD of RgpB from *P. gingivalis* strain W83. The results revealed a novel molecular mechanism of inhibition of peptidases and could thus pave the way for the design of novel inhibitory strategies that may help in palliating the effects of periodontal disease.

EXPERIMENTAL PROCEDURES

Protein production and complex formation — The wild-type PD of *Porphyromonas gingivalis* strain W83 gingipain R2 alias RgpB proteinase (see UniProt database access code P95493) was obtained as reported elsewhere. Briefly, the coding sequence (Q^{25}-R^{229}) was cloned into the pGEX-6P-1 expression vector using *Bam*HI/*Xho*I restriction sites, which attached an N-terminal glutathione-S-transferase-tag and a PreScission protease cleavage site. The vector was transformed into *Escherichia coli* BL21 (DE3) cells, and overexpression was induced with isopropyl-1-thio-β-D-galactopyranoside. The protein was purified in a glutathione-Sepharose High Performance column, cleaved with PreScission protease, and passed again through glutathione-Sepharose. The flow-through was concentrated by ultrafiltration. The final purified protein contained an N-terminal extension of GPLGS as the result of the cloning strategy. The CD plus the IgSF (residues Y^{230}-G^{662}), fused to a C-terminal hexahistidine tag for purification, was purified from culture medium of the *P. gingivalis* strain W83 bearing a modified *rgpB* gene, in which a sequence encoding six histidine residues had been inserted in frame at the junction between the IgSF and the C-terminal domain. This construct results in the secretion of soluble RgpB with a C-terminal His-tag as described (24). For complex formation, RgpB (15mg) was pre-activated in gel filtration buffer (50mM sodium phosphate, 0.15M sodium chloride, pH 7.2) freshly supplemented with 10mM L-cysteine for 10min. The activated RgpB-6xHis was then treated with 1mM N-[5-amino-L-1-(2-chloroacetyl)pentyl]-4-methyl-benzenesulfonamide (TLCK). After 10-15min preincubation, the PD was added in 1.5-molar excess with respect to RgpB and the reaction mixture was incubated for another 15min. All incubations were performed at room temperature. The complex was then separated from excess PD and TLCK by gel filtration on a Superdex 75 10/60 column (GE Health). The purity of the complex was evaluated by native PAGE and the protein concentration was determined by BCA assay (Sigma). The stability of the complex was assessed by following activity over time after incubation at 37ºC. No catalytic activity was observed after one week. This finding correlated well with purified full-length intact pro-RgpB undergoing auto-cleavage at the maturation site (R^{229}-Y^{230}) but showing no significant catalytic activity even after two weeks of incubation.

Crystallization and structure determination — 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) at the High-Throughput Crystallography Platform (PAC) at Barcelona Science Park. Plates were stored in Bruker steady-temperature crystal farms at 4°C and 20ºC. Successful

conditions were scaled up to the microliter range with 24-well Cryschem crystallization dishes (Hampton Research). Best crystals were obtained at 20ºC with protein complex solution (9.1mg/mL in 5mM Tris·HCl pH 7.4; 1mM 1,4-dithio-DL-threitol [DTT]; 1mM TLCK) and 14% polyethylene glycol 6000; 0.1M sodium acetate pH5.0; 0.2M calcium chloride as reservoir solution (with barium chloride as an additive) from 1:1μL or 2:1μL drops. Crystals were cryo-protected by immersion in harvesting solution (21% polyethylene glycol 6000; 0.1M sodium acetate pH5.0; 0.2M calcium chloride; 20%[v/v] glycerol). A complete diffraction dataset was collected from a liquid-N_2 flash-cryo-cooled crystal at 100K (provided 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 monoclinic and contained four PD/CD+IgSF complexes per asymmetric unit. Diffraction data were integrated, scaled, merged, and reduced with programs XDS and XSCALE (25) (see Table 1).

The structure of the PD/CD+IgSF RgpB complex was solved by likelihood-scoring molecular replacement with program PHASER (26) using the coordinates of the protein part only of mature RgpB of *P. gingivalis* strain HG66 (GenBank AAB41892; 97% sequence identity; Protein Data Bank [PDB] access code 1CVR; (22)). These calculations rendered four unambiguous solutions with values for the rotation/translation function Z-scores of 18.6/18.7, 18.5/40.0, 22.6/62.3, and 19.6/73.1, respectively, and confirmed space group $P2_1$ as the correct one. Subsequent density modification with program DM (27) under fourfold averaging rendered an electron-density map which enabled construction of most of the 205-residue PD for one of the four complexes on a Silicon Graphics Octane2 Workstation using program TURBO-FRODO (28). The position and orientation of the other three copies within the asymmetric unit were determined with PHASER. Subsequent model building alternated with crystallographic refinement with program BUSTER/TNT (29), which included TLS refinement and NCS restraints, until completion of the model. The final model contained four PDs (chains A [R^{31}-S^{204}+T^{210}-A^{227}], C [G^{30}-L^{205}+F^{211}-R^{229}], E [G^{30}-L^{205}+F^{211}-T^{228}], and G [R^{31}-S^{204}+F^{211}-E^{226}]) and four cognate CD+IgSF moieties (chains B [G^{239}-E^{661}], D, F, and H [all N^{238}-E^{661}]). Within each catalytic moiety, the CDs were much more rigid and better defined by proper electron density than the cognate IgSFs; within each CD, loop Lα7η3 (~530s) was flexible and traced based on weak electron density to preserve chain continuity. In addition to the protein chains, one Ba^{2+} (tentatively assigned based on the electron density map and presence in the crystallization conditions) and three Ca^{2+} cations were identified for each CD+IgSF moiety. Furthermore, in addition to a single magnesium and tris cation, two further tentatively-assigned calcium and three sodium cations, as well as one chloride anion, one glycerol molecule, and 858 solvent molecules completed the model. Each of the four catalytic cysteine residues (C^{473}) evinced extra electron density for its side chain, which we attribute to the purification strategy (see above) and conservatively interpreted as a methylsulfino group (residue type CSD). P^{532} of chain H and two residues of each CD (S^{449} and V^{474}) were the only Ramachandran outliers of the entire structure (see Table 1). The latter two were also outliers in the mature chloromethylketone-complex structure (22) and are unambiguously defined by proper electron density. Superposition of the three complexes CD, EF, and GH onto AB, respectively, revealed 612 common Cα atoms deviating less than 3Å, which gave rise to *rmsd* values of 0.75Å, 0.55Å, and 0.48Å, respectively, and indicated close similarity of the structures. This was confirmed by an independent superposition of the four CDs, which revealed very similar orientations and positions for the cognate PDs, with just marginal displacements on the proteinase distal surface (max. 3.8Å at E^{197}). Accordingly, the four complexes were considered equivalent and the results and discussion hereafter refer to complex AB, which had slightly lower overall thermal displacement parameters (A/B: 41.5Å^2/35.1Å^2, C/D: 46.2Å^2/38.1Å^2, E/F: 47.5Å^2/37.8Å^2, and G/H: 48.4Å^2/36.7Å^2), unless otherwise stated. Wherever distances, angles, etc. are mentioned, the range found in all four complexes is indicated.

Miscellaneous — Figures were prepared with program CHIMERA (30). Interaction surfaces (taken as half of the surface area buried at a complex interface) were calculated with CNS (31). Structure similarities were investigated with DALI (32). Model validation was performed with MOLPROBITY (33) and the WHATCHECK routine of WHATIF (34). The final coordinates of the complex between *P. gingivalis* RgpB PD and the CD+IgSF moiety have been deposited with the PDB at www.pdb.org (access code 4IEF).

RESULTS AND DISCUSSION

An experimental model of pro-RgpB — All attempts to crystallize intact full-length pro-RgpB for structural studies failed. Accordingly, the PD (residues Q^{25}-R^{229}) and the mature moiety (residues Y^{230}-G^{662}) of pro-RgpB from *P. gingivalis* strain W83 were produced separately and mixed to yield the zymogenic complex (see Experimental procedures). Four such complexes were found in the asymmetric unit of the crystal structure, which was determined by likelihood-scoring molecular replacement and averaging techniques, and refined with diffraction data to 2.3Å resolution (see Table 1). Superposition of the respective mature moieties revealed very similar relative arrangements of the four PDs. In addition, the distances (20-26Å) between the last residues of the PD moieties defined in the electron density (E^{226}/R^{229}) and the first of the respective CDs (N^{238}/G^{239}), both on the surface, could easily be bridged by the missing 8-11 residues running along the molecular surface. Accordingly, the present complex provides a *bona fide* model of the intact zymogen, which correlates well with the strong inhibitory capacity of the PD on CD+IgSF *in trans*. The biochemical and kinetic analysis of the interaction revealed a non-competitive mode of interaction leading to formation of the 1:1 stoichiometric complex, which was stable in native PAGE and size-exclusion chromatography.

The stability obeyed to an apparent inhibition constant in the low nanomolar range (K_i=6.2±1.0nM). The very tight inhibitory interaction between the PD and the mature enzyme suggests existence of a yet unknown mechanism facilitating dissociation of the complex and degradation of the PD *in vivo*. At present we can only speculate on the nature of this mechanism. It can be related to glycosylation of the gingipain during the secretion process. Alternatively, one of the components of the PorSS secretion system, which is engaged in secretion of virulence factors for periodontopathogenicity (35), could displace the PD during pro-RgpB translocation across the outer membrane and unleash activity.

A pro-domain with a novel fold — The PD is defined by the electron density for residues G^{30}/R^{31}-E^{226}/R^{229} and has the overall shape of a croissant with maximal dimensions ~60x35x20Å (Fig. 1A,B). It consists of a central eleven-stranded β-core (strands βI-βXI) divided in two antiparallel β-sandwiches, 1 and 2, and decorated on the surface with an α-helix (αI) and two 3_{10}-helices (ηI and ηII; see Fig. 1C,D). The two sandwiches are held together by a continuous hydrophobic core that reaches from L^{37}, L^{92}, and M^{223} on the right side of sandwich 1 to Y^{175}, P^{177}, and K^{180} on the left of sandwich 2 (view as in Fig. 1C). Sandwich 1 consists of a three-stranded back sheet (strands βV, βIX, and βVII+βVIII) and a four-stranded front sheet (strands βI, βII, βXI, and βVI; see Fig. 1C,D). Strands βXI and βVI are N- and C-terminally extended beyond the limits of the sandwich, respectively, and bent by ~50-60º. In this manner, they also contribute to the three-stranded front sheet of sandwich 2 (strands βVI, βIX, and βX), which is packed against a two-stranded back sheet (strands βIII and βIV). The two sandwiches are roughly perpendicular to each other, both for the direction of the contributing strands and the planes of the sheets (Fig. 1C,D). The strand connectivity of the two sandwiches is such that a first β-ribbon at the top of the front sheet of sandwich 1 (ribbon βIβII) is linked to the β-ribbon that creates the back sheet of sandwich 2 (βIIIβIV). This, in turn, ends in a loop connecting strands βIV and βV (LβIVβV), which is the top strand of the back sheet of sandwich 1. After βV, the chain enters the bottom strand of the front sheet of sandwich 1, βVI (Fig. 1C,D), the second half of which is the top strand of the front sheet of sandwich 2. After βVI, a 39-residue loop segment runs across the bottom surface of PD. This loop includes 3_{10}-helices ηI and ηII and encompasses a so-called "inhibitory loop" (see below; Fig. 1C). This long loop leads to the bottom strand of the back sheet of sandwich 1, which is split in two, βVII and βVIII. The latter contributes to a β-ribbon together with the central back-sheet strand of sandwich 1, βIX. Thereafter, the polypeptide enters a β-ribbon (βXβXI), which creates the bottom of the front sheet of sandwich 2. As described for βVI, βXI is also extended and gives rise to the third strand (top to bottom) of the front sheet of sandwich 1 before leaving the β-core of the PD (Fig. 1C,D). Thereafter, the chain passes through a flexible segment on the right surface, which is disordered at L^{205}/V^{206}-S^{209}/T^{210} and ends at the surface-located C-terminal helix αI on the back of sandwich 1, and finishes at E^{226}/R^{229}.

Structural similarity searches identified agglutinin from the roman snail, *Helix pomatia*, as the closest structural relative of the PD (PDB 2CE6; (36); Z-score=6.3; *rmsd*=2.7Å for 89 common residues according to program DALI (37)). This is a hexameric sugar-binding lectin from the albumen gland of the gastropod and part of its innate immune system. It belongs to a family of sugar-binding proteins mostly from invertebrates, which also includes discoidin C-terminal domain (PDB 2W94), and is a 3+3 pure β-sandwich. Upon superposition of the PD and agglutinin, it becomes evident that the latter resembles PD sandwich 1, both in connectivity and topology (Fig. 1E). However, the PD shows an additional β-strand, the N-terminal βI, and, most importantly, the second sandwich and the unique loop structures including the two 3_{10}-helices and the C-terminal α-helix. All-β domains similar to agglutinin are also present in glycolytic enzymes, i.e. sugar-binding proteins, such as endocellulase 9G (PDB 1G87), β-xylosidase (PDB 1W91), and endogluceramidase II (PDB 2OSX), as well as in the sugar-binding cellulosomal scaffolding protein A (PDB 4B9F), all of which were also identified as structurally related to the PD but match PD sandwich 1 only. We conclude that the PD of RgpB has a novel fold, hitherto unseen in peptidase zymogens and distantly related to functionally-unrelated sugar-binding proteins.

The mature enzyme moiety — The structure of the mature enzyme moiety within the zymogenic complex resembles a molar, with its crown, tooth body, and root. Its superposition onto the structure of a covalent complex of a closely-related mature enzyme moiety from a distinct bacterial strain with D-Phe-Phe-Arg-chloromethylketone (PDB 1CVR; (3,22)) reveals that, with the exception of some minor changes in side-chain conformations (see next section), the entire structure, including the active-site cleft and the surrounding moiety, is conserved, i.e. the zymogenic conformation of CD+IgSF induced by the PD reveals that the enzyme is probably in a competent conformation in the zymogen, as is often observed in peptidases, thus suggesting inhibition is mediated by competition with the substrate (38). Briefly, the CD (N^{238}/G^{239}-P^{580}) is subdivided into an N-terminal (or A-) sub-domain (NSD; N^{238}/G^{239}-E^{345}) and a C-terminal (or B-) sub-domain (CSD; S^{346}-P^{580}). Both sub-domains are α/β-moieties consisting of a central β-sheet, a four-stranded parallel one in the NSD and a six-stranded one parallel for all its strands except the outermost top one in the CSD (see Fig. 2B; NSD on top, CSD at bottom). This strand contacts the NSD β-sheet in an approximately perpendicular manner (see also Fig. 2 in (22)). NSD is decorated with one helix and two helices plus a short β-ribbon, respectively, on either side of the sheet. The CSD sheet has four helices and three helices plus a small three-stranded antiparallel β-sheet, respectively, on either side (Fig. 2B). One calcium ion is found in both the NSD and the CSD (the latter cation is most likely a barium in the present zymogenic structure due to the crystallization conditions; see Experimental procedures), and a further one is present at the sub-domain interface (Fig. 2B).

The active-site cleft is found at the "masticating surface" (22) of the molar crown and is formed by the CSD. As with enzymes with an α/β-hydrolase or PLEES fold (39,40), active-site residues are provided by loops and strands at the C-terminal edge of the central β-sheet, in this case that of the CSD: C^{473} is donated by the loop after the fourth strand (bottom to top in Fig. 2B), H^{440} by the first strand of the small three-stranded sheet inserted after the third strand, and E^{381} by a loop after the second strand. While cysteine-histidine dyads are common for cysteine proteinases, including MEROPS family C25 (41), the position and distance of E^{381} to H^{440} Nε2 in the mature enzyme (22) and in the present structure (Fig. 2B) suggests a role in protonation and, thus, side-chain orientation of the catalytic histidine during catalysis in RgpB, as described for an aspartate in the foot-and-mouth-disease-virus leader cysteine peptidase (42). In the present structure, the catalytic cysteine displayed extra density for its side chain beyond its Sγ atom possibly due to the purification procedure, which included reversible covalent-inhibition steps, and which we conservatively interpreted as a methylsulfino side chain (see Experimental procedures). On the opposite surface to the molar crown—where the active site is located—, the downstream IgSF (T^{581}-E^{661}) is inserted between the NSD and the CSD, thus mimicking the root of the molar. This domain is an antiparallel seven-stranded β-barrel or 3+4 β-sandwich, which contains a fourth calcium-binding site on the surface (Fig. 2A). Its fold corresponds to that of classic immunoglobulin-like domains (43) as found in, e.g., α_2-macroglobulin (44).

Interaction between the pro-domain and the catalytic domain — By contrast with other zymogens that are inhibited by large, structured, globular domains, e.g. those of the metallocarboxypeptidase class (38), the PD does not frontally cover and shield the CD active-site cleft but, rather, attaches laterally through its concave croissant surface to the enzyme moiety (Fig. 2A,B). The interaction of the PD with the CD (no interaction is observed between the PD and the IgSF) occludes a surface of ~1,650Å^2 (~16% of the total PD surface), which is within the range generally described for protein-protein complexes (1,250-1,750Å^2; (45)). It shows a surface complementarity (Sc=0.75) that is likewise within the range reported for protein oligomers and protein/protein inhibitor interfaces (0.70-0.76; (46)). The interaction results from 82 contacts (<4Å), among them three salt bridges (D^{216}-K^{536}, E^{73}-K^{553}, and R^{126}-D^{392}), one protein-metal interaction, 21 hydrogen bonds, and hydrophobic interactions between 11 PD and 13 CD residues. Participating segments of the CD are provided by the interface between the sub-domains and the active-site cleft: G^{239}, G^{296}-T^{300}, N^{331}-F^{338}, E^{381}-D^{392}, T^{438}-H^{440}, V^{471}-C^{473}, N^{510}-R^{518}, N^{533}-K^{536}, N^{545}-F^{548}, E^{552}-D^{557}. Further interactions include the calcium ion of the NSD. The PD contributes with segments V^{68}-I^{77} and S^{89}-S^{91} (from βIV, LβIVβV, and βV), S^{155}-R^{161} and V^{167}-N^{169} (from βVII, βVIII, LβVIIIβIX, and βIX), and N^{201}-I^{203} and F^{211}-V^{221} (from LβXIαI and αI). In particular, I^{159} O of PD replaces a solvent molecule binding the calcium ion of the NSD in the mature inhibitor-complex structure (PDB 1CVR; (22)). This ion has an overall octahedral co-ordination sphere and is further bound by NSD atoms V^{329} O, D^{332} Oδ2, Y^{334} O, and, bidentately, E^{336} Oδ1 and Oδ2, as well as by a solvent molecule.

Notwithstanding, the most relevant interaction of the pro-domain with the catalytic moiety is exerted by the inhibitory loop (K^{121}-Y^{135}), which is part of the segment flanked by strand βVI and the 3_{10}-helix ηII and is inserted like a stinger into the non-primed side of the active-site cleft, so that R^{126}-A^{127} occupies the position of a potential scissile bond. The inhibitory loop has a compact structure, which is contributed to by a tight 1,4-turn of type I (S^{125}-E^{128}), 3_{10}-helix ηI (N^{129}-I^{133}), and a small hydrophobic core created by P^{130}, I^{133}, I^{124}, and K^{121}, which interacts with W^{513} of the CD (Fig. 2D). The lateral ζ-ammonium group of K^{121} also contributes to the compact structure of the inhibitory loop through a double hydrogen bond with the main chain at P^{130} and I^{133} (Fig. 2D). Superposition of the present zymogenic complex and the reported chloromethylketone-bound enzyme, which mimics a substrate-bound form, reveals that segment I^{124}-R^{126} of the pro-domain binds to the active-site cleft in a substrate-like manner, i.e. in extended conformation and in the correct orientation. This binding places S^{125} in cleft sub-site S_2, I^{124} in S_3, and, most importantly, R^{126} in S_1, thus matching the specificity of RgpB (see Fig. 2C,D). The S_1 specificity pocket is lined by Q^{511}, V^{471}, M^{517}, H^{395}, T^{438}, and, at the bottom, D^{392}, which establishes a bidentate salt-bridge with R^{126}. In addition, the side chain of W^{513} closes the S_1 pocket like a lid. Moreover, R^{126} also interacts with the catalytic H^{440}: the carbonyl oxygen of R^{126} establishes a very strong hydrogen bond with H^{440} Nδ1 (2.3-2.6Å apart), causing the imidazole side chain to be rotated slightly around its χ_1-angle, away from the catalytic cysteine and towards E^{381} (Fig. 3). This rotation, in turn, gives rise to a second strong hydrogen bond between one of the E^{381} carboxylate oxygens and H^{440} Nε2 (2.6-2.9Å). This interference with the catalytic residues most likely prevents cleavage of the PD at R^{126}-A^{127}. In addition, the polypeptide chain folds back after R^{126} towards bulk solvent, so that the active-site cleft is free on its primed side, and this also leads the catalytic C^{473} Sγ atom to be too far apart (3.7-3.8Å) from R^{126}-A^{127}.

A novel mechanism of zymogenic inhibition — Latency maintenance in cysteine proteinases (17,47) has been structurally studied for MEROPS family C1 members such as *Carica papaya* plant papain (48) and caricain (49), and mammalian cathepsins B, K, L, S, and X (50-57). In these cases, the CD and the active-site cleft are, overall, in a preformed competent conformation in the zymogens and the ~60-100-residue PDs, which contain a globular part laterally attached to the CD, possess a C-terminal segment that runs across the entire active-site cleft in the opposite orientation to that of a substrate, thus blocking access to the cleft and preventing autolysis. Zymogens of family C14, in turn, include the structurally-studied mammalian caspases 1 (58); 3 (59); 7 (60,61), and 8 (62,63); and the insect *Drosophila* caspase-9 ortholog DRONC (64) and *Spodoptera frugiperda* caspase-1 ortholog ((65) and PDB 2NN3). Caspases are oligomeric functional enzymes and activation cleavage entails major rearrangement of the loops flanking the active-site cleft from an incompetent to a competent conformation. There are no truly globular PDs but, rather, short N-terminal

and/or internal peptides that are cleaved off during maturation, thus giving rise to two chains in the competent enzymes (60,61,66). Zymogenic activation has been also studied for *Staphylococcus aureus* staphopain B (67), a member of family C47. Here, a large 183-residue globular PD based on a barrel-sandwich hybrid possesses a loop in the middle of the structure that binds, in the opposite orientation to that of substrates, to the cleft of an overall competent CD but only blocks the primed side of the cleft. *Streptococcus pyogenes* exotoxin B (alias SpeB or streptopain) (68) and *Prevotella intermedia* interpain A (69) belong to family C10, and here a backing–helix of a ~115-residue α/β-sandwich PD is inserted laterally into the cleft in the zymogen but does not completely block substrate access. Activation entails major rearrangement of a zymogenic hairpin and a latency flap, which leads to a large displacement of the catalytic histidine (69).

In stark contrast, the description of the zymogenic complex of RgpB reported here demonstrates that the PD, the largest structurally-characterized to date for a cysteine peptidase with 205 residues, interacts laterally through a large surface with the CD. The distance from the activation cleavage site, R^{229}-Y^{230}, on the top back surface of the complex in Fig. 2B, to the catalytic cysteine, on the front bottom in Fig. 2B, is >40Å, strongly suggesting that cleavage and, thus, removal of the PD and activation of the CD occurs *in trans*, as previously suggested for the cysteine peptidase interpain A (69). The PD intrudes into the active-site cleft through a structurally-cohered inhibitory loop in the middle of the domain, thus blocking access to non-primed side of the cleft only. This mechanism is unlike previously reported ones and is thus unique for cysteine peptidases and peptidases in general. It will pave the way to designing small-molecule inhibitors that mimic the structure of the inhibitory loop and that inhibit RgpB in a non-covalent manner, and may contribute to the development of novel drugs to combat periodontitis.

REFERENCES

1. Petersen, P. E., and Ogawa, H. (2005) *J. Periodontol.* **76**, 2187-2193
2. Bostanci, N., and Belibasakis, G. N. (2012) *FEMS Microbiol. Lett.* **333**, 1-9
3. Yongqing, T., Potempa, J., Pike, R. N., and Wijeyewickrema, L. C. (2011) *Adv. Exp. Med. Biol.* **712**, 15-29
4. Paster, B. J., Olsen, I., Aas, J. A., and Dewhirst, F. E. (2006) *Periodontol. 2000* **42**, 80-87
5. Socransky, S. S., Haffajee, A. D., Cugini, M. A., Smith, C., and Kent, R. L., Jr. (1998) *J. Clin. Periodontol.* **25**, 134-144
6. Yang, H. W., Huang, Y. F., and Chou, M. Y. (2004) *J. Periodontol.* **75**, 1077-1083
7. van Winkelhoff, A. J., Loos, B. G., van der Reijden, W. A., and van der Velden, U. (2002) *J. Clin. Periodontol.* **29**, 1023-1028
8. Yilmaz, O. (2008) *Microbiology* **154**, 2897-2903
9. Lamont, R. J., and Jenkinson, H. F. (1998) *Microbiol. Mol. Biol. Rev.* **62**, 1244-1263
10. Guo, Y., Nguyen, K. A., and Potempa, J. (2010) *Periodontol. 2000* **54**, 15-44
11. Potempa, J., Pike, R., and Travis, J. (1997) *Biol. Chem.* **378**, 223-230
12. Schifferle, R. E., Shostad, S. A., Bayers-Thering, M. T., Dyer, D. W., and Neiders, M. E. (1996) *J. Endod.* **22**, 352-355
13. O'Brien-Simpson, N. M., Paolini, R. A., Hoffmann, B., Slakeski, N., Dashper, S. G., and Reynolds, E. C. (2001) *Infect. Immun.* **69**, 7527-7534
14. Holzer, H., and Heinrich, P. C. (1980) *Annu. Rev. Biochem.* **49**, 63-91
15. Khan, A. R., and James, M. N. (1998) *Prot. Sci.* **7**, 815-836
16. Bryan, P. N. (2002) *Chem. Rev.* **102**, 4805-4816
17. Wiederanders, B. (2003) *Acta Biochim. Pol.* **50**, 691-713
18. Tao, K., Stearns, N. A., Dong, J., Wu, Q. L., and Sahagian, G. G. (1994) *Arch Biochem Biophys* **311**, 19-27
19. Mikolajczyk, J., Boatright, K. M., Stennicke, H. R., Nazif, T., Potempa, J., Bogyo, M., and Salvesen, G. S. (2003) *J. Biol. Chem.* **278**, 10458-10464
20. Mittl, P. R., and Grütter, M. G. (2006) *Curr. Opin. Struct. Biol.* **16**, 769-775
21. Pike, R., McGraw, W., Potempa, J., and Travis, J. (1994) *J. Biol. Chem.* **269**, 406-411
22. Eichinger, A., Beisel, H. G., Jacob, U., Huber, R., Medrano, F. J., Banbula, A., Potempa, J., Travis, J., and Bode, W. (1999) *EMBO J.* **18**, 5453-5462
23. Rawlings, N. D., Barrett, A. J., and Bateman, A. (2012) *Nucl. Acids Res.* **40**, D343-D350
24. Skottrup, P. D., Leonard, P., Kaczmarek, J. Z., Veillard, F., Enghild, J. J., O'Kennedy, R., Sroka, A., Clausen, R. P., Potempa, J., and Riise, E. (2011) *Anal Biochem* **415**, 158-167
25. Kabsch, W. (2001) Chapter 25.2.9: XDS. in *International Tables for Crystallography. Volume F: Crystallography of Biological Macromolecules*. (Rossmann, M. G., and Arnold, E. eds.), 1st Ed., Kluwer Academic Publishers (for The International Union of Crystallography), Dordrecht (The Netherlands). pp 730-734
26. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J. Appl. Crystallogr.* **40**, 658-674
27. Cowtan, K. (2010) *Acta Crystallogr. sect. D* **66**, 470-478
28. 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
29. Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M., and Bricogne, G. (2004) *Acta Crystallogr. sect. D* **60**, 2210-2221
30. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) *J. Comput. Chem.* **25**, 1605-1612
31. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. sect. D* **54**, 905-921
32. Holm, L., Kaariainen, S., Wilton, C., and Plewczynski, D. (2006) *Curr. Protoc. Bioinformatics* **Chapter 5**, Unit 5 5
33. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Bryan Arendall, W., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) *Nucl. Acids Res.* **35 (Web Server issue)**, W375-W383
34. Vriend, G. (1990) *J. Mol. Graph.* **8**, 52-56
35. Sato, K., Naito, M., Yukitake, H., Hirakawa, H., Shoji, M., McBride, M. J., Rhodes, R. G., and Nakayama, K. (2010) *Proc. Natl. Acad. Sci. USA* **107**, 276-281
36. Sánchez, J. F., Lescar, J., Chazalet, V., Audfray, A., Gagnon, J., Alvarez, R., Breton, C., Imberty, A., and Mitchell, E. P. (2006) *J. Biol. Chem.* **281**, 20171-20180
37. Holm, L., and Rosenström, P. (2010) *Nucleic Acids Res.* **38**, W545-W549
38. Gomis-Rüth, F. X. (2008) *Crit. Rev. Biochem. Mol. Biol.* **43**, 319-345

39. 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) *Prot. Engng.* **5**, 197-211
40. Puente, X. S., and López-Otín, C. (1997) *Biochem. J.* **332**, 947-949
41. Chen, J. M., Rawlings, N. D., Stevens, R. A., and Barrett, A. J. (1998) *FEBS Lett.* **441**, 361-365
42. Guarné, A., Hampoelz, B., Glaser, W., Carpena, X., Tormo, J., Fita, I., and Skern, T. (2000) *J. Mol. Biol.* **302**, 1227-1240
43. Bork, P., Holm, L., and Sander, C. (1994) *J. Mol. Biol.* **242**, 309-320
44. Marrero, A., Duquerroy, S., Trapani, S., Goulas, T., Guevara, T., Andersen, G. R., Navaza, J., Sottrup-Jensen, L., and Gomis-Rüth, F. X. (2012) *Angew. Chem. Int. Ed.* **51**, 3340-3344
45. Janin, J., and Chothia, C. (1990) *J. Biol. Chem.* **265**, 16027-16030
46. Lawrence, M. C., and Colman, P. M. (1993) *J. Mol. Biol.* **234**, 946-950
47. Potempa, J., Golonka, E., Filipek, R., and Shaw, L. N. (2005) *Mol. Microbiol.* **57**, 605-610
48. Roy, S., Choudhury, D., Aich, P., Dattagupta, J. K., and Biswas, S. (2012) *Acta Crystallogr. sect. D* **68**, 1591-1603
49. Groves, M. R., Taylor, M. A., Scott, M., Cummings, N. J., Pickersgill, R. W., and Jenkins, J. A. (1996) *Structure* **4**, 1193-1203
50. Coulombe, R., Grochulski, P., Sivaraman, J., Menard, R., Mort, J. S., and Cygler, M. (1996) *EMBO J.* **15**, 5492-5503
51. Podobnik, M., Kuhelj, R., Turk, V., and Turk, D. (1997) *J. Mol. Biol.* **271**, 774-788
52. Cygler, M., Sivaraman, J., Grochulski, P., Coulombe, R., Storer, A. C., and Mort, J. S. (1996) *Structure* **4**, 405-416
53. Sivaraman, J., Lalumière, M., Ménard, R., and Cygler, M. (1999) *Protein. Sci.* **8**, 283-290
54. Sivaraman, J., Nagler, D. K., Zhang, R., Menard, R., and Cygler, M. (2000) *J. Mol. Biol.* **295**, 939-951
55. LaLonde, J. M., Zhao, B., Janson, C. A., D'Alessio, K. J., McQueney, M. S., Orsini, M. J., Debouck, C. M., and Smith, W. W. (1999) *Biochemistry* **38**, 862-869
56. Kaulmann, G., Palm, G. J., Schilling, K., Hilgenfeld, R., and Wiederanders, B. (2006) *Protein Sci.* **15**, 2619-2629
57. Turk, V., Stoka, V., Vasiljeva, O., Renko, M., Sun, T., Turk, B., and Turk, D. (2012) *Biochim. Biophys. Acta* **1824**, 68-88
58. Elliott, J. M., Rouge, L., Wiesmann, C., and Scheer, J. M. (2009) *J. Biol. Chem.* **284**, 6546-6553
59. Walters, J., Pop, C., Scott, F. L., Drag, M., Swartz, P., Mattos, C., Salvesen, G. S., and Clark, A. C. (2009) *Biochem. J.* **424**, 335-345
60. Chai, J., Wu, Q., Shiozaki, E., Srinivasula, S. M., Alnemri, E. S., and Shi, Y. (2001) *Cell* **107**, 399-407
61. Riedl, S. J., Fuentes-Prior, P., Renatus, M., Kairies, N., Krapp, S., Huber, R., Salvesen, G. S., and Bode, W. (2001) *Proc. Natl. Acad. Sci. USA* **98**, 14790-14795
62. Yu, J. W., Jeffrey, P. D., and Shi, Y. (2009) *Proc. Natl. Acad. Sci. USA* **106**, 8169-8174
63. Keller, N., Mares, J., Zerbe, O., and Grutter, M. G. (2009) *Structure* **17**, 438-448
64. Yan, N., Huh, J. R., Schirf, V., Demeler, B., Hay, B. A., and Shi, Y. (2006) *J. Biol. Chem.* **281**, 8667-8674
65. Forsyth, C. M., Lemongello, D., LaCount, D. J., Friesen, P. D., and Fisher, A. J. (2004) *J. Biol. Chem.* **279**, 7001-7008
66. Fuentes-Prior, P., and Salvesen, G. S. (2004) *Biochem. J.* **384**, 201-232
67. Filipek, R., Szczepanowski, R., Sabat, A., Potempa, J., and Bochtler, M. (2004) *Biochemistry* **43**, 14306-14315
68. Kagawa, T. F., Cooney, J. C., Baker, H. M., McSweeney, S., Liu, M., Gubba, S., Musser, J. M., and Baker, E. N. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 2235-2240
69. 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) *J. Biol. Chem.* **283**, 2871-2882
70. Schechter, I., and Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27**, 157-162
71. Weiss, M. S. (2001) *J. Appl. Cryst.* **34**, 130-135
72. Evans, P. (2006) *Acta Crystallogr. sect. D* **62**, 72-82
73. 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) *Acta Crystallogr. sect. D* **66**, 12-21

ACKNOWLEDGMENTS

We are grateful to the Automated Crystallography Platform at IBMB/IRB/PCB for assistance during crystallization experiments. We acknowledge the help provided by local contacts at the ESRF synchrotron. Funding for data collection was provided in part by ESRF.

FOOTNOTES

**This study was 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-F3-2009-223101 "AntiPathoGN"; FP7-HEALTH-2010-261460 "Gums&Joints"; FP7-PEOPLE-2011-ITN-290246 "RAPID"; BIO2009-10334; BFU2012-32862; CSD2006-00015; Fundació "La Marató de TV3" grant 2009-100732; and 2009SGR1036).

FIGURE LEGENDS

Figure 1 – General architecture of the RgpB pro-domain. (A) Ribbon-type plot of RgpB PD showing the regular secondary-structure elements (α- and 3_{10}-helices in magenta and labeled αI and ηI-ηII, respectively; β-strand as blue arrows and labeled βI-βXI) and the approximate overall dimensions of the molecule. The inhibitory loop, which includes the S_1-intruding residue, R^{126}, is also labeled. Two small black arrows pinpoint the residues flanking the disordered segment preceding the C-terminal helix αI. **(B)** and **(C)** depict orthogonal views of (A). **(D)** Topology scheme of RgpB PD roughly in the same orientation as in (C). Each regular secondary-structure element is labeled and marked with its limiting residues. **(E)** Superposition in wall-eye stereo of RgpB PD (turquoise) and *Helix pomatia* agglutinin (magenta; PDB 2CE6; (36)).

Figure 2 – The zymogenic complex. (A) Ribbon-type plot in wall-eye stereo of the complex between the RgpB PD (in blue/magenta) and the mature RgpB moiety in front view. The latter consists of domains CD (in yellow/orange) and IgSF (in green/brown). The three calcium and the barium ions are depicted as red and magenta spheres, respectively. The inhibitory loop and the respective N- and C-termini are labeled, in turquoise for PD and in brown for CD+IgSF. R^{126} from the PD inhibitory loop and the active-site residues of CD—C^{473}, H^{440}, and E^{381}—are further shown as sticks for reference of the active site. **(B)** Orthogonal view of (A) showing the CD in standard orientation (22,70), i.e. with the view into the active-site cleft, which runs horizontally from left (non-primed side) to right (primed side). **(C)** Cartoon of the complex in the orientations of (A)(left) and (B)(right) showing the regions of the PD and CD engaged in binding in dark blue and orange, respectively. The rest of each molecule is shown in turquoise and yellow, respectively. **(D)** Close-up view in wall-eye stereo of the area around the inhibitory loop delimited by a black rectangle in (C). The CD moiety is shown as a tan ribbon and selected residues are labeled and shown for their side chains as sticks with tan carbons. The inhibitory loop (K^{121}-Y^{135}) is shown as a stick model with carbons in turquoise. Selected residues are also labeled. The Ba^{2+} ion of the CD is depicted as a magenta sphere. Note that the catalytic cysteine, C^{473}, is oxidized to 3-sulfino-L-alanine (residue name CSD).

Figure 3 – The preformed catalytic moiety. Ribbon-plot in wall-eye stereo showing the superposition of the CD in the present zymogenic complex (yellow) and its mature inhibitor-bound form (purple; PDB 1CVR; (22)) in standard orientation. The two calcium ions (red spheres) and the barium ion (magenta sphere) found in the CD correspond to the zymogenic complex structure. Selected active-site residues are shown as sticks for each structure, as is the covalent inhibitor, atom-colored with green carbons.

Table 1. Crystallographic data.

Space group / cell constants (a,b, and c, in Å; β in º)	$P2_1$ / 84.4, 133.1, 109.8, 90.5
Wavelength (Å)	0.9393
No. of measurements / unique reflections	395,643 / 106,820
Resolution range (Å) (outermost shell)	47.3 – 2.30 (2.36 – 2.30)
Completeness (%)	99.2 (91.2)
R_{merge} [a]	0.070 (0.436)
$R_{r.i.m.}$ (= R_{meas}) [a]	0.082 (0.540)
Average intensity over stand. dev. (<[<I> / σ(<I>)]>)	15.2 (2.6)
B-Factor (Wilson) (Å^2) / Average multiplicity	40.9 / 3.7 (2.8)
Resolution range used for refinement (Å)	∞ – 2.30
No. of reflections in working set / in test set	101,461 (5,323)
Crystallographic R_{factor} (free R_{factor}) [b]	0.189 (0.225)
No. of protein atoms / solvent molecules / ligands / ions	19,119 / 858 / 1 $(CH_2OH)_2CHOH$ / 14 Ca^{2+}, 4 Ba^{2+}, 1 Mg^{2+}, 3 Na^+, 1 Cl^-, 1 $(OHCH_2)_3C(NH_3^+)$
Rmsd from target values	
bond lengths (Å) / bond angles (°)	0.008 / 1.04
Overall average B-factor (Å^2)	39.7
Main-chain conformational angle analysis [c]	
Residues in favored regions / outliers / all residues	2,395 / 9 / 2,438

Values in parentheses refer to the outermost resolution shell.

[a] $R_{rmerge} = \Sigma_{hkl}\Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$; $R_{r.i.m.} = \Sigma_{hkl}(n_{hkl}/[n_{hkl}-1])^{1/2} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$; $R_{p.i.m.} = \Sigma_{hkl}(1/[n_{hkl}-1])^{1/2} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl}\Sigma_i I_i(hkl)$, where $I_i(hkl)$ is the i-th intensity measurement and n_{hkl} the redundancy of reflection hkl—including symmetry-related reflections—and $\langle I(hkl)\rangle$ its average intensity. $R_{r.i.m.}$ (*alias* R_{meas}) and $R_{p.i.m.}$ 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 (71,72)).

[b] Crystallographic $R_{factor} = \Sigma_{hkl} ||F_{obs}| - k |F_{calc}|| / \Sigma_{hkl} |F_{obs}|$, where k is a scaling factor, and F_{obs} and F_{calc} are the observed and calc structure factor amplitudes, respectively. This factor is calculated for the working-set reflections; free R_{factor}, same for a test-set of refl (>500) not used during refinement.

[c] According to MOLPROBITY (73).

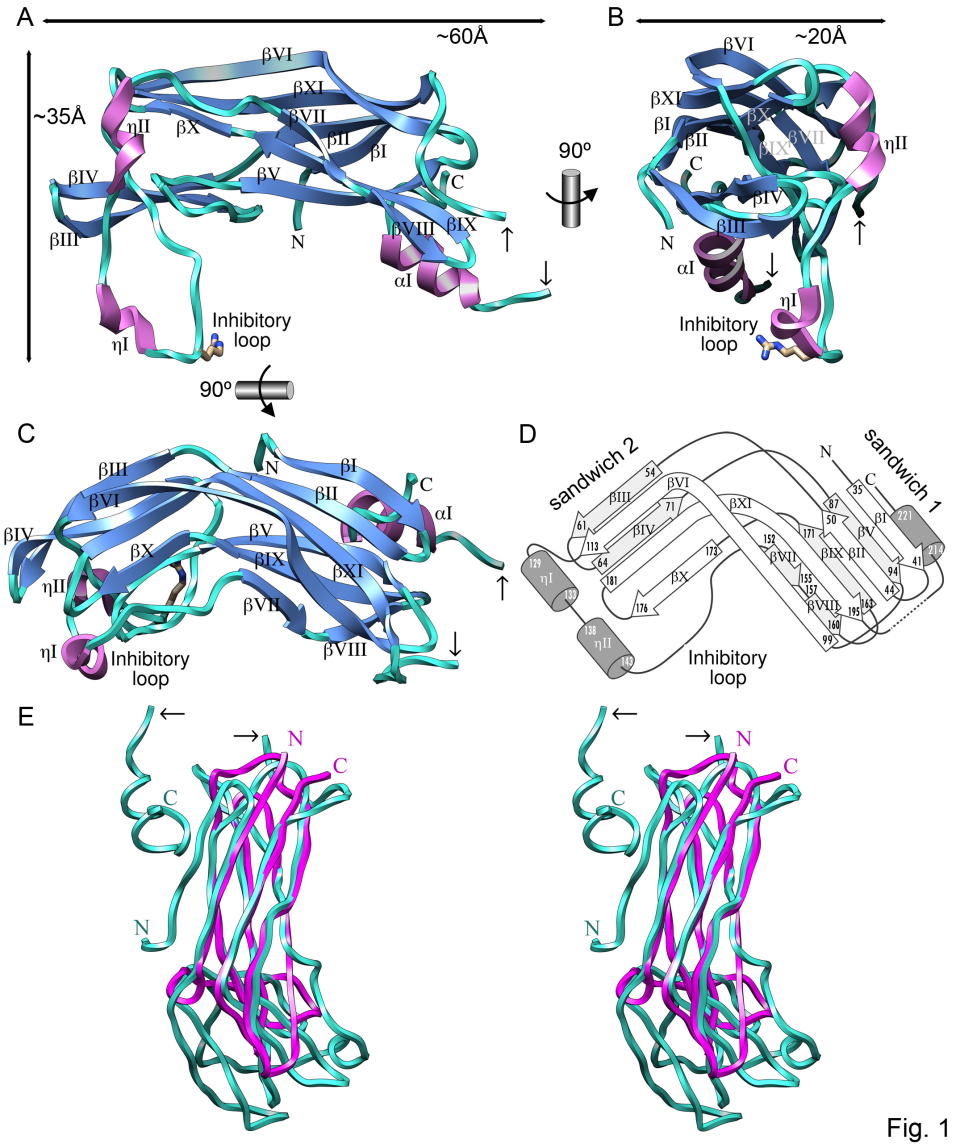
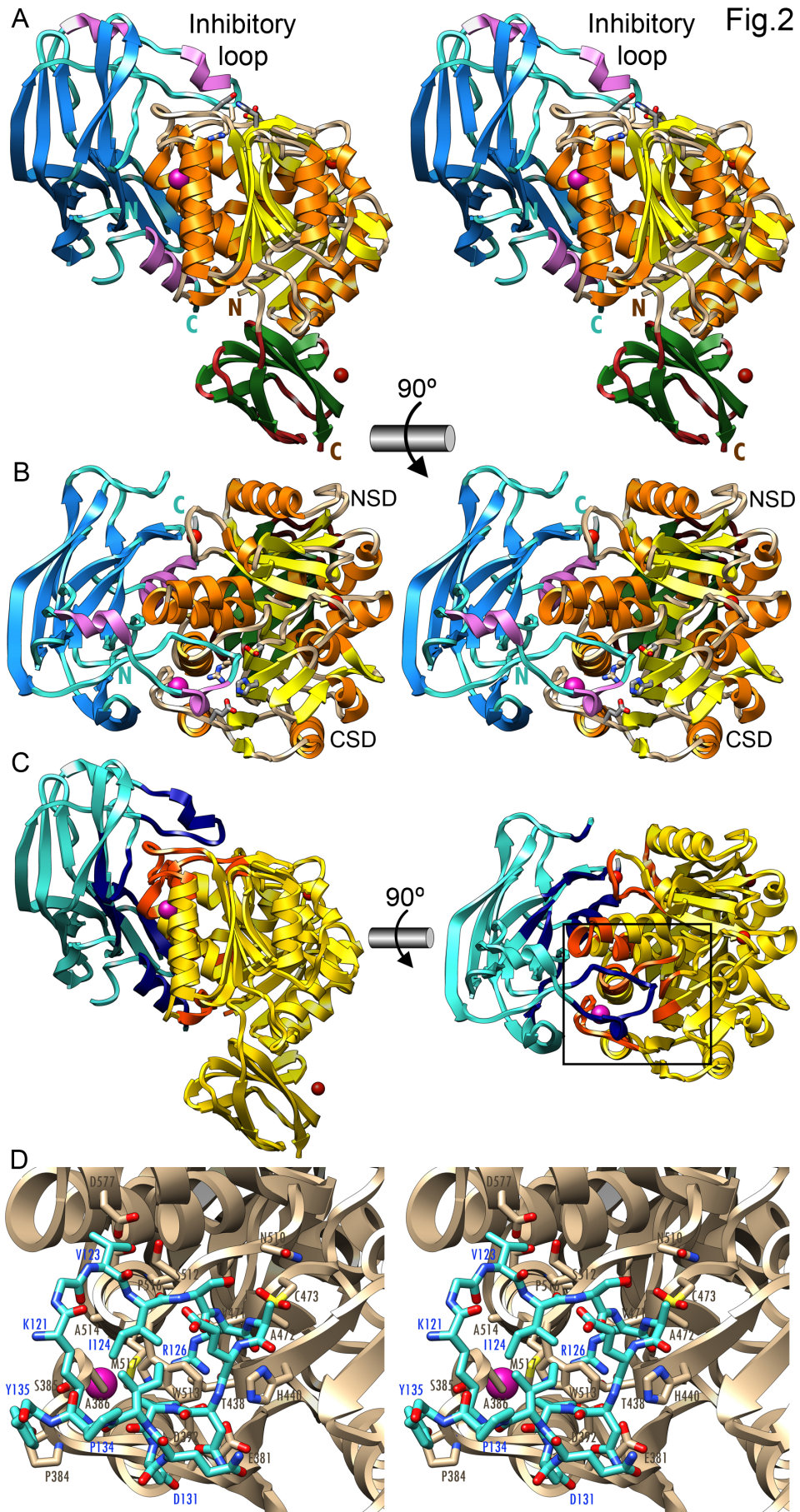

A
~60Å
~35Å
βVI
βXI
βVII
ηII
βX
βII
βI
βIV
βV
C
βIII
N
βVIII
βIX
αI
ηI
Inhibitory loop
90°
B
~20Å
βVI
βXI
βI
βII
C
βIV
βIII
N
αI
ηI
ηII
Inhibitory loop
90°
C
βIII
N
βI
βVI
βII
C
αI
βIV
βV
βX
βIX
βXI
ηII
βVII
ηI
Inhibitory loop
βVIII
D
sandwich 2
sandwich 1
N
C
54
βVI
βIII
71
61
βIV
113
64
181
βX
176
βXI
152
βVII
155
157
βVIII
160
99
87
50
βV
βI
βIX
βII
171
195
94
44
41
221
214
35
129
ηI
133
138
ηII
142
Inhibitory loop
E
N
C
C
N
N
C
C
N

Fig. 1


A
Inhibitory loop
Inhibitory loop
Fig.2
N
C
N
C
90°
C
B
NSD
NSD
N
CSD
CSD
C
D
D577
N510
V123
S512
C473
K121
A514
I124
A472
R126
M517
Y135
S383
A386
W511
T438
H440
P134
D397
E381
P384
D131

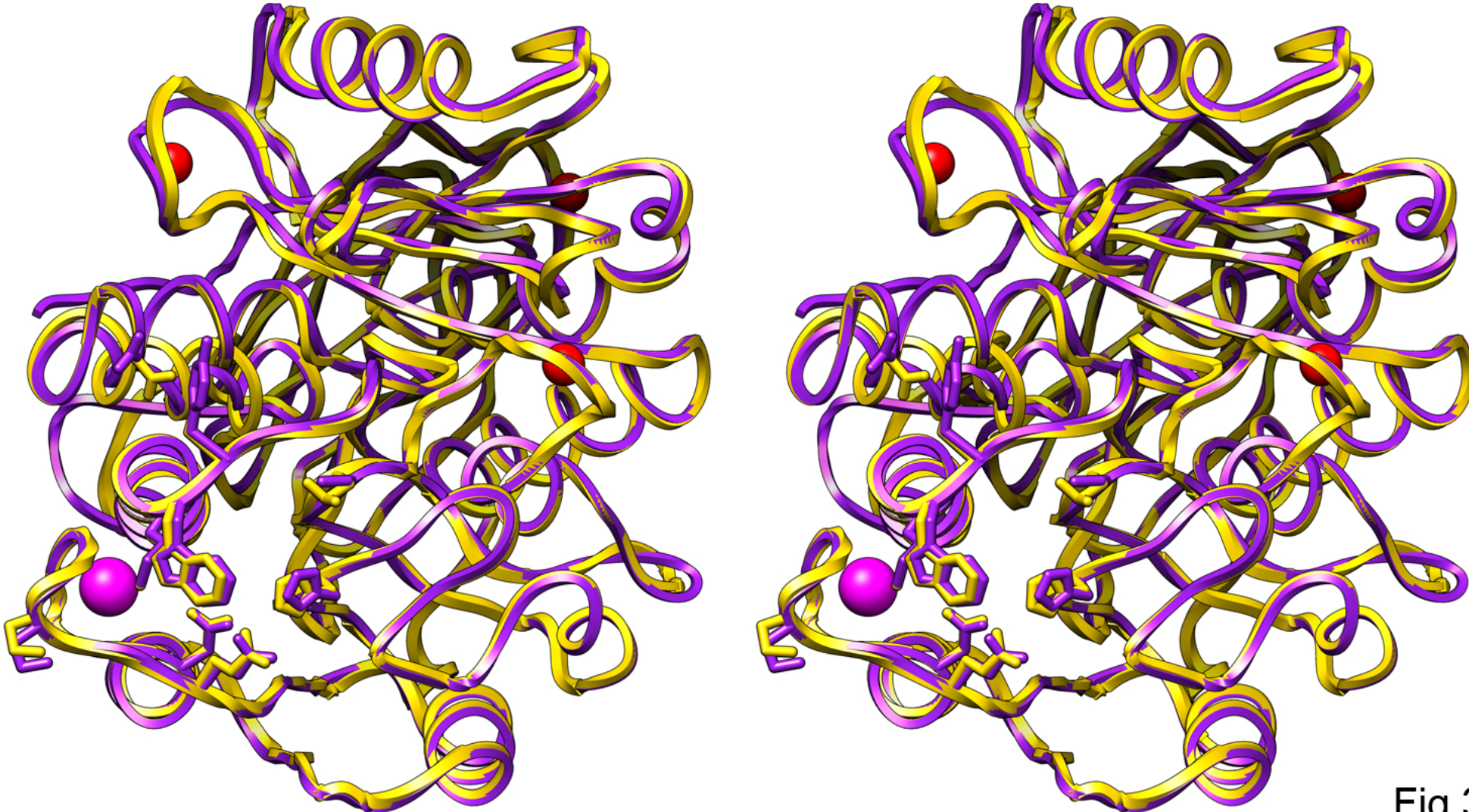


Fig.3