JBC Papers in Press. Published on September 30, 2005 as Manuscript M506766200
The latest version is at http://www.jbc.org/cgi/doi/10.1074/jbc.M506766200

EXPRESSION, CRYSTALLIZATION AND THREE-DIMENSIONAL STRUCTURE OF THE CATALYTIC DOMAIN OF HUMAN PLASMA KALLIKREIN

Jie Tang, Christine Luong Yu, Steven R. Williams, Eric Springman[1], Douglas Jeffery, Paul Sprengeler, Alberto Estevez, Jun Sampang, William Shrader, Jeff Spencer, Wendy Young, Mary McGrath and Bradley A. Katz*

Running Title: Structure of Plasma Kallikrein Protease Domain

Address correspondence to Bradley A. Katz, Department of Structural Chemistry, Celera Genomics, 180 Kimball Way, South San Francisco, 94080, Tel. 650 866-6270; Fax. 650 866-6654; E-Mail: brad.katz@celera.com

Current address: Locus Pharmaceuticals, Inc., 4 Valley Square, 512 E. Township Line Road, Blue Bell, PA 19422

Plasma kallikrein is a serine protease that has many important functions including modulation of blood pressure, complement activation, and mediation and maintenance of inflammatory responses. Although plasma kallikrein has been purified for 40 years, its structure has not been elucidated. In this report, we describe two systems (*Pichia pastoris* and baculovirus/Sf9 cells) for expression of the protease domain of plasma kallikrein, along with the purification and high-resolution crystal structures of the two recombinant forms. In the *Pichia pastoris* system, the protease domain was expressed as a heterogeneously glycosylated zymogen that was activated by limited trypsin digestion and treated with Endo H deglycosidase to reduce heterogeneity from the glycosylation. The resulting protein was chromatographically resolved into four components, one of which was crystallized. In the baculovirus/Sf9 system, homogeneous, crystallizable, non-glycosylated protein was expressed after mutagenizing three asparagines (the glycosylation sites) to glutamates. When assayed against the peptide substrates, Pefachrome PK and oxidized insulin B chain, both forms of the protease domain were found to have catalytic activity similar to that of the full-length protein. Crystallization and x-ray crystal structure determination of both forms have yielded the first three-dimensional views of the catalytic domain of plasma kallikrein. The structures, determined at 1.85 Å for the Endo H-deglycosylated protease domain produced from *P. pastoris*, and at 1.40 Å for the mutagenically deglycosylated form produced from Sf9 cells, show that the protease domain adopts a typical chymotrypsin-like serine protease conformation. The structural information provides insights into the biochemical and enzymatic properties of plasma kallikrein and paves the way for structure-based design of protease inhibitors that are selective either for or against plasma kallikrein.

The name "kallikrein" was originally introduced in the 1930's to describe the proteolytic activity from the pancreas responsible for cleavage of kininogen (1,2). The kallikreins are now divided into two major categories: plasma kallikrein (EC 3.4.21.34) and tissue kallikreins (3,4). Tissue kallikreins were initially discovered as an activity from pancreatic or renal tissues effecting the release of kinin from kininogens. Plasma kallikrein was discovered from a hereditary abnormality termed Fletcher trait, characterized by a prolonged partial thromboplastin time that becomes normal if the plasma is exposed to a clot-promoting surface (5). Prekallikrein (PK) was identified as the protein whose absence is responsible for Fletcher trait (6).

Since its discovery in 1965, plasma kallikrein has been biochemically characterized in great detail. The human cDNA of PK encodes a signal peptide of 19 amino acids and a mature prekallikrein of 619 amino acids (7). Activation of PK to kallikrein occurs through a cleavage at the Arg371-Ile372 bond (based on the numbering of the mature protein), producing a two-subunit protein involving a heavy chain and a light chain (6), linked through a disulfide bond between Cys364 and Cys484 (8). The NH_2-terminal heavy chain of 371 amino acids contains four Apple domains whose homologues are also found in factor XI (fXI) (9), while the COOH-terminal light chain forms the protease domain. The two chains have been isolated through reduction and alkylation of the purified enzyme and through affinity capture of the heavy chain on immobilized high molecular weight kinin-Sepharose (10). The purified light chain has amidolytic activity similar to that of the native enzyme, but is less active in kaolin-dependent coagulation assays. Both the light and heavy chains are glycosylated; PK from human blood contains 15% carbohydrate and has five potential sites for N-glycosylation. Due to heterogeneity in the carbohydrate moiety, the light chain can exist as 36 kDa and 33 kDa isoforms.

Although plasma and tissue kallikreins are related, and share the historical name, "kallikrein", for kininogen degrading activity, plasma kallikrein differs from tissue kallikrein members in gene structure, amino acid sequence, protein subunit structure, substrate specificity, and physiological functions (3,4,8,11). Tissue kallikreins currently comprising 15 identified members (hK1 through hK15), are encoded by genes clustered at 19q13.3-19q13.4 of the human genome, while plasma kallikrein is encoded by a single gene on human chromosome 4q35. In contrast to the multi-domain structure of plasma kallikrein (homologous to that of fXI), each tissue kallikrein is comprised of a protease domain alone. The degree of sequence identity between the protease domain of plasma kallikrein and individual members of the tissue kallikrein family ranges from 30% to 36%, whereas it is much higher between the protease domains of plasma kallikrein and fXI (66 %). Thus, plasma kallikrein bears more resemblance to the coagulation factor, fXI, than to members of the tissue kallikrein family.

Plasma kallikrein plays a central role in the kinin generating pathways (PK system) and in the surface-mediated "contact system" (8, 12). It is synthesized predominantly in the liver as a proenzyme, prekallikrein, also known as prokallikrein (13). In plasma, approximately 90% of the circulating PK is complexed with high-molecular-weight kininogen (HK) (14). On endothelial cells, the PK-HK complex binds to a multi-protein complex that consists of cytokeratin 1 (CK1), urokinase plasminogen activator receptor (uPAR) and gC1qR, a ubiquitously expressed, multicompartmental cellular protein involved in modulating complement, coagulation, and kinin cascades (15-21). The binding of PK-HK to the multi-protein complex results in the activation of PK by prolylcarboxypeptidase (PRCP), a constitutively active serine protease located on endothelial membranes (22-23). The activated plasma kallikrein in turn cleaves HK to liberate bradykinin, activates factor XII (fXII) to factor XIIa (fXIIa) and converts prorenin to renin, the central component in the renin-angiotensin system (RAS) (8,23). By controlling the release of bradykinin (a potent vasodialator) and the activation of renin (the protease converting angiotensinogen to angiotensin I), plasma kallikrein is deeply involved in blood pressure regulation. Plasma kallikrein also participates in fibrinolysis by activating prourokinase and plasminogen to urokinase and plasmin, respectively (24-27). Finally, plasma kallikrein stimulates neutrophils to aggregate and release their lysosomal contents such as elastase (28,29). Thus, plasma kallikrein has many important physiological functions including modulation of blood pressure, complement activation, mediation of inflammatory responses, and maintenance of the balance between fibrin deposition and fibrinolysis (8).

In recent years, tremendous efforts to design specific protease inhibitors for treating diseases have yielded many life-saving drugs (30-33). One such field receiving considerable attention involves the blood coagulation

pathways (34). Anticoagulant therapies are in great demand to block undesired thrombosis of surgery patients and to treat many life-threatening diseases such as stroke, acute coronary syndrome, and deep vein/pulmonary embolism (35). Current therapies with heparin or coumarins have limited efficacy and safety records (36,37), prompting synthesis of many small molecules intended to inhibit specific target proteases, such as thrombin, factor VIIa (fVIIa), factor IXa (fIXa), and factor Xa (fXa), within the coagulation pathways (34). However, because of the similarity among plasma kallikrein and the coagulation factors, including a preference for arginine at the S1 substrate-binding site, many inhibitors designed for other coagulation factors also inhibit plasma kallikrein.

The therapeutic efficacy of a protease inhibitor relies not only on high affinity for the disease target, but also on attenuated affinity for anti-targets, whose vital functions need to be maintained. Thus, structural information both for the protease target and for related anti-targets often provide the key for designing potent, selective, safe inhibitors. High-resolution structures can be used to enhance inhibitor selectivity by identifying major structural differences among protease homologues at distal binding sites, such as S4, S3, S2, S1', S2', S3' or S4' (38), or by harnessing more subtle differences at the S1 site. Thus, the determined crystal structures of proteases such as thrombin, fVIIa and fXa have provided a wealth of information for increasing inhibitor potency and selectivity toward the particular protease target (33,39-43). However, one important anti-target for these and other serine protease targets is plasma kallikrein, whose structure has been unavailable. As a step toward the design of inhibitors with enhanced selectivity for a target protease like fVIIa and against other anti-targets, including plasma kallikrein, we initiated efforts to establish high-level expression systems for plasma kallikrein, to purify and to crystallize the protein, and to determine its structure.

Plasma kallikrein is not only an important anti-target for therapeutic inhibitors designed toward other serine proteases, it is also emerging as an important target itself for some disease states (8,22,44,45). For example, abnormal plasma kallikrein activity is implicated in the symptoms of hereditary angioedema (44,45), a disease caused by decreased functional levels of the major physiological inhibitor of kallikrein, C1 Inhibitor. During acute attacks of this disease, increased levels of bradykinin cause an increase in vascular permeability. Plasma kallikrein inhibitors suppress the defect in vascular permeability in C1 inhibitor-knockout mice (46), and thus may constitute treatment for hereditary angioedema. Other disorders involving plasma kallikrein are septicemia and septic shock (23,47). The over-production of bradykinin is implicated in the pathogenesis of septic shock through its ability to lower blood pressure in both humans and animals. Because of the important role of plasma kallikrein in these disease states, a structure of this enzyme would be a critical asset for structure-based design of a potent, selective, small molecule therapeutic.

High-resolution structures of human tissue kallikreins, hK1 (48,49), hK6 (50), and pro-hK6 (51) have been determined. However, the degree of sequence identity of these tissue kallikrein structures to that of plasma kallikrein is at best 34 % (for hK6). More importantly, the loop sequences that form distal binding sites (such as S1') deemed important for inhibitor design differ significantly between tissue and plasma kallikrein. Thus the available structures of tissue kallikrein (or of other proteases) can provide only rough guidance in the design of potent, selective inhibitors of plasma kallikrein.

In this report, we describe the expression of the protease domain of plasma kallikrein in the methylotrophic yeast *Pichia pastoris*, and the production of crystallizable protein by removal of Endo H-sensitive glycosylation. We also describe the expression of the protease domain in the baculovirus system as a non-glycosylated form with all three N-linked glycosylation sites removed through site-directed mutagenesis. Both forms were purified and crystallized to yield the first reported crystal structures of the protease domain of human plasma kallikrein (at 1.85 Å for the enzymatically deglycosylated form, and at 1.40 Å for the mutagenically deglycosylated

form) The structures show that the protease domain of plasma kallikrein adopts a typical chymotrypsin-like serine protease fold, and provide insight into substrate selectivity related to the function of this important protease. In addition, well-defined structural features including the unique S1' site pave the way for structure-based design of protease inhibitors that are selective either for or against plasma kallikrein.

MATERIALS AND METHODS

Expression of plasma kallikrein protease domain in Pichia pastoris - A cDNA sequence encoding the plasma kallikrein protease domain, from Asn357 to Ala619 (based on the numbering system of mature prekallikrein, signal sequence not included), in which Cys364 and Cys484 were mutagenized to serines, was inserted into pPIC9. The pPIC9 plasmid containing the insert was linearized with SalI and electroporated into *Pichia pastoris* strain KM71. *Pichia pastoris* clones were analyzed for expression of the plasma kallikrein protease domain, and an expression clone was inoculated into the complex media BMMY. After overnight incubation at 30°C with shaking, the cells were then used to inoculate a 15 liter bioreactor containing ~8 liters of BMMY media. Fermentation with glycerol feed continued for 24 hours to a cell density whose absorbance was 500 at 600 nm. The expression of recombinant plasma kallikrein protease domain was induced by switching the carbon source to methanol for 48-60 hours. The fermentation was harvested by centrifuging the culture in a Sorvall RC3B at 5000 rpm for 30 min. The supernatant was recovered and filtered through a 0.22 um filter to remove particles.

Expression of mutagenically deglycosylated plasma kallikrein in Sf9 cells - The non-glycosylated plasma kallikrein protease domain was expressed in the baculovirus system. The gp67 signal peptide sequence (MLLVNDSHQGFNKEHTSKMVSAIVLYVL LAAAAHSAFA) was synthesized from overlapping oligonucleotides and cloned into pVC19. The cDNA fragment encoding Asn357-Ala619 of plasma kallikrein (numbering system based on the mature protein sequence of prekallikrein) was fused to the gp67 signal peptide sequence by standard techniques. Sequences encoding the glycosylation asparagines, Asn377, Asn434 and Asn475, and the cysteines, Cys364 and Cys484, were mutated to encode Glu377, Glu434, Glu475, Ser364 and Ser484, respectively through site-directed mutagenesis using a Stratagene Quikchange Multi Site-directed Mutagenesis kit. The mutant cDNA was subcloned into pFastBac1 (Invitrogen). The expression baculovirus stock was obtained by transposition in *E.coli* strain DH10Bac and by transfection of Sf9 cells with the recovered bacmid. To express mutagenically deglycosylated plasma kallikrein protease domain, Sf9 cells were infected with the recombinant baculovirus at a MOI of 5. After infection for 65-72 hours, the culture was harvested by centrifugation in a Sorvall RC3B at 5000 rpm for 30 min. The supernatant was recovered and filtered through a 0.22 um filter.

Purification and characterization of recombinant Endo H-deglycosylated plasma kallikrein protease domain - The recombinant plasma kallikrein protease domain expressed in *P. pastoris* was secreted into the culture media as a heterogeneously glycosylated zymogen which could be activated through limited trypsin digestion. A typical purification process is summarized in Table 1. The *P. pastoris* media (6 liters) was concentrated and buffer exchanged against 50 mM Na_3PO_4, pH 6.5 with a 10 K cutoff ultrafiltration system (Millipore). To remove contaminating proteins, the processed *Pichia* media (600 ml) was applied twice to a 300 ml Q-Sepharose Fast Flow (GE Healthcare) column that had been equilibrated with 50 mM Na_3PO_4, pH 6.5. Flow-through fractions from the Q-Sepharose column were combined and diluted with two volumes of 5 mM 2-morpholinoethanesulfonic acid (MES), pH 5.5, and the pH adjusted to 5.5. This solution was then applied to a Source 15S (15 μm resin) column (1.6 cm x 10 cm, GE Healthcare) that had been equilibrated with 30 mM MES, pH 5.5. After washing with 10 column volumes of 30 mM MES, pH 5.5, the protein was eluted from the Source S column with 30 mM MES, 350

mM NaCl, pH 5.5. To remove the Endo-H-sensitive glycosylation from the recombinant plasma kallikrein, Endo H (Roche) was added to the Source S pool at a ratio of 30 mU per mg of total protein, and the reaction mixture incubated at 37 ºC for 100 minutes. The reaction mixture was then diluted with an equal volume of 100 mM Tris-HCl, 300 mM NaCl, pH 7.6, and the final pH adjusted to 7.6. The recombinant plasma kallikrein in the diluted Endo H reaction mixture was activated by adding freshly dissolved sequencing grade trypsin (Roche) at a ratio of 2 ug trypsin per mg of total protein and by incubating at 37 ºC for 60 minutes. The activated, deglycosylated plasma kallikrein protease domain was then loaded onto a 16 ml benzamidine-Sepharose column (high substitute, GE Healthcare) that had been equilibrated with 50 mM Na_3PO_4, pH 7.0. After washing the column with 5 volumes of 50 mM Na_3PO_4, 200 mM NaCl, pH 7.0, and then with 3 volumes of 50 mM Na_3PO_4, pH 7.0, the kallikrein was eluted with 50 mM Na_3PO_4, 50 mM benzamidine, pH 7.0. Fractions containing protein were pooled and dialyzed against 30 mM MES, pH 5.5. The dialyzed protein solution was further fractionated with a Source 15S column (1 cm x 10 cm), and eluted with 30 column volumes of 30 mM MES, pH 5.5, containing a linear NaCl gradient from 0 to 400 mM. This Endo H-treated plasma kallikrein protease domain was resolved into four peaks (presumably due to differential, residual Endo H-resistant glycosylation), the first of which eluted at a conductivity of ~ 8 ms/cm. Dynamic light scattering (DLS) experiments indicated that the protein of each peak was monodisperse. The protein from peak 1 (hereafter referred to as Endo H-deglycosylated plasma kallikrein) was used for structural studies.

Purification and characterization of recombinant mutagenically deglycosylated plasma kallikrein protease domain - Mutagenically deglycosylated plasma kallikrein protease domain was purified from 3.0 liters of conditioned media of infected Sf9 cells (Table 2). After slow addition of $(NH_4)_2SO_4$ to the media (with stirring) to bring the final concentration to 1.8 M, the media was loaded onto a Phenyl-Sepharose High Performance column (GE Healthcare, 2.6 cm x 15 cm) that had been equilibrated with 30 mM Na_3PO_4, pH 7.0 containing 2 M $(NH_4)_2SO_4$. The Phenyl-Sepharose column was then washed with the equilibration buffer until a baseline absorbance was achieved. Plasma kallikrein protease domain was eluted with 30 mM Na_3PO_4, pH 7.0 and then dialyzed overnight against 50 mM Na_3PO_4, pH 6.5. The dialyzed protein was applied to a 150 ml Q-Sepharose Fast Flow column that had been equilibrated with 50 mM Na_3PO_4, pH 6.5. Flow-through fractions from the Q-Sepharose column (containing the plasma kallikrein) were recovered and diluted with an equal volume of 50 mM Na_3PO_4, pH 7.6, 300 mM NaCl, and the pH adjusted to 7.6. Activation of the plasma kallikrein protease domain was achieved by adding sequencing grade trypsin at the ratio of 2 ug trypsin per mg of total protein and by incubating the reaction mixture at 37 ˚C for 60 minutes. The activated protein was further purified with a benzamidine-Sepharose column using conditions similar to those described for purification of Endo-H deglycosylated plasma kallikrein. After concentration by ultrafiltration with a 10 K membrane (Millipore), the protein was further purified with a high-load Superdex 200 column (1.6 cm x 60 cm, GE Healthcare) using 30 mM Na_3PO_4, 140 mM NaCl, pH 7.4 as the eluent. The fractions containing plasma kallikrein were combined and dialyzed against 30 mM MES, pH 5.5. The dialyzed sample was finally purified on a Source 15S column using conditions similar to those described for purification of Endo-H deglycosylated plasma kallikrein. From 3 liters of conditioned media, 0.7 mg of purified plasma kallikrein protease domain was recovered.

Enzymatic activity of full-length plasma kallikrein and of the recombinant forms of the protease domain - Amidolytic activities of full-length plasma kallikrein and of the recombinant forms of the protease domain were assayed at room temperature in 100 ul reaction mixtures containing 50 mM Tris-Hcl, 150 mM NaCl, 2.5% DMSO, pH 7.6 and 250 uM Pefachrome-PK (2AcOH-H-D-But-CHA-Arg-*p*NA,

Centerchem, Inc) in 96 well microtiter plates. The release of *p*-nitroaniline was monitored at 405 nm with a SpectraMax plate reader controlled by SoftMax Pro software. Enzyme concentration was determined by bicinchoninic acid assays (Pierce-Endogen). Kinetic parameters were determined by non-linear least squares regression analysis of the absorbance versus time data for a series of assays involving substrate concentrations ranging from 2 uM to 500 uM. The values were calculated by fitting the data-points to the Michaelis-Menten equation using Kaleidagraph (Synergy Software).

The peptidase specificities of full-length plasma kallikrein and of the recombinant forms of the protease domain were compared with one another by assaying the rate of cleavage of oxidized insulin B chain and the nature of the cleavage products. In each reaction, 75 ug of oxidized insulin B chain (Sigma) in 50 mM Tris-HCl, 150 mM NaCl, pH 7.6 was incubated with 7.5 ug full length plasma kallikrein or with 1.5 ug recombinant protease domain in a 75 ul reaction. In control reactions, no protease was included. After incubation for 1 hr or for 24 hr, 25 ul of the reaction mixture was removed and the reaction stopped by adding 25 ul 8 M guanidine HCl, pH 3.0. The digestion mixture was loaded onto a Zobax C18 column (4.6 mm x 15 cm, Agilent Technologies) with a HP1090 HPLC system, and the column then washed with 95 % mobile phase A (H_2O, 0.1% TFA), 5% mobile phase B (ACN, 0.08% TFA) for 5 min at a flow rate of 1.0 ml/min. The peptides were eluted from the column with a linear gradient from 5% to 95 % mobile phase B over a 45 min time period. Peptides in the chromatogram were monitored at 220 nm and their identities established by collision-induced dissociation followed by peptide sequencing using a Q-Star mass spectrometer from Applied Biosystems. The oxidized insulin B chain (Peak IV) eluted at 22.68 min, the Phe1-Arg22 fragment (Peak III) at 20.56 min, Gly23-Ala30 (Peak II) at 17.65 min, and Gly23-Lys29 (Peak 1) at 17.26 min. The areas of these peaks were integrated to yield the relative amounts after 1 hr incubation (Table 3a), and after 24 hr incubation (Table 3b).

Determination of inhibition constants for full-length plasma kallikrein and for other serine proteases - The K_i values involving small molecule inhibitors toward a panel of serine proteases including plasma kallikrein, were determined at pH 7.4 as described (52). Apparent inhibition constants, K_i' values, were calculated from the enzyme hydrolysis rate data collected at various inhibitor concentrations with *BatchKi* (developed and provided by Dr. Petr Kuzmic, Biokin Ltd., Pullman, MA) using methodology similar to that described for tight-binding inhibitors (53). K_i' values were converted to K_i values by the formula, $K_i = K_i'/(1 + S/K_m)$.

Crystallization and x-ray diffraction data collection of recombinant plasma kallikrein - Human plasma kallikrein (either Endo H- or mutagenically deglycosylated), in 30 mM MES, 75 mM NaCl, pH 5.5, was concentrated to 12-18 mg/ml in the presence of 20 mM benzamidine (Sigma) using the Amicon Centricon-10 system (Millipore). The benzamidine complex of the enzymatically deglycosylated form was crystallized from sitting drops, composed of 0.5 µl of the protein solution and 0.5 µl of the reservoir solution (25% PEG 6K, 0.10 M MES pH 6.5). The benzamidine complex of the mutagenically deglycosylated form was crystallized from sitting drops composed of 0.5 µl of the protein solution and 0.5 µl of the reservoir solution (0.20 M potassium dihydrogen phosphate, 20 % PEG 3350). Crystals of 0.1 mm in each dimension grew within 2 weeks at 17°C.

X-ray diffraction datasets for plasma kallikrein crystals were collected at –160° C at the Berkeley ALS Synchrotron (λ = 1.000 Å). For the endo H-deglycosylated plasma kallikrein protease domain (space group = $P2_12_12$, a = 79.89, b = 63.19, c = 50.31 Å), the Quantum 315 CCD detector on beam-line 5.0.2 was used with the following parameters (distance = 235 cm; 24 sec per 1.0° frame; 180° total). For the other construct, in which all three asparagines involved in N-glycosylation were changed to glutamates (space group $P2_12_12_1$, a = 55.19, b = 57.36, c = 79.85 Å), data were collected with the Quantum 210 CCD detector (distance = 140 cm;

15 sec per 1.0° frame; 180° total) at beam-line 5.0.1. The datasets were reduced with Denzo (54).

A Blast search of the RCSB for the structure most homologous to plasma kallikrein identified the extra-cellular region of hepsin (39 % sequence identity), as an appropriate probe for molecular replacement. (Recently the structure of a much more homologous protease, fXIa (identity = 66 %) has been deposited, entry code 1XX9, reference 55). The molecular replacement program written by Liang Tong as implemented in the X-sight subroutine licensed from Accelrys was used to solve the structure of enzymatically deglycosylated plasma kallikrein, using the hepsin probe structure (entry code 1P57 (56) obtained from the RCSB (57). The refined structure of enzymatically deglycosylated plasma kallikrein was then used to solve the structure of the mutagenically deglycosylated plasma kallikrein by molecular replacement.

To help determine the structures of some of the loops, automated fitting was carried out with the ARP/wARP package (warpNtrace mode) (58). Structures were refined manually by analysis of ($|F_o|$-$|F_c|$), α_c and of ($2|F_o|$-$|F_c|$), α_c maps, followed by refinement with X-PLOR (59,60). Map analysis and X-PLOR refinement were iterated until no further improvement in R_{cryst} or in the models could be achieved. X-ray diffraction and refinement statistics and RCSB PDB accession codes are provided in Table 4.

Synthesis of inhibitors- Synthesis of inhibitors is described in detail elsewhere (61-63).

RESULTS

Purification and characterization of the Endo H- and mutagenically deglycosylated plasma kallikrein protease domain - The protease domain of plasma kallikrein (Asn357-Ala619) expressed in *P. pastoris* was heterogeneously glycosylated and secreted as a zymogen. The zymogen was activated by trypsin treatment, which removes the N-terminal peptide. Heterogeneity in the recombinant protease domain results from incomplete carbohydrate processing during biosynthesis. Endoglycosidase H (EndoH) was used to remove most of the carbohydrate and to convert the protease domain to a single band on SDS-PAGE. This Endo H-deglycosylated protein could be further fractionated into three to four species with cation-exchange chromatography using Mono S or Source S columns (data not shown). These species differed in molecular weight by 1000-2000 Da, most likely due to differential glycosylation. Each was monodisperse as analyzed by dynamic light scattering (DLS). The K_m values of the various species were similar to one another, as were the K_{cat}/K_m values, using Pefachrome-PK as the substrate (data not shown). We used the first Source S peak for crystallization experiments because the extent of post-translational modification was the least and the purification yield was the highest.

To obtain a nonglycosylated plasma kallikrein protease domain and to confirm that the differences in molecular weight among Source S peaks were from N-glycosylation, we mutagenized all three N-glycosylation sites (Asn377, Asn434 and Asn475) to glutamates and expressed the mutant in a baculovirus system. The mutant kallikrein protease domain was secreted into culture media as a zymogen and could be activated by trypsin digestion. The purified mutant protease was homogenous on cation exchange chromatography and was resolved as a single peak on Mono S (data not shown). LC-MS analysis of the purified nonglycosylated mutant plasma kallikrein protease domain indicated that it included residues from Ile372 to Lys612 (Ile16 to Lys248 basing on chymotrypsin numbering scheme) without any posttranslational modification (Fig. 1). This protein, hereafter referred to as mutagenically deglycosylated plasma kallikrein, was found to be monodisperse by DLS analysis, and was used for crystallization and for structural studies.

Comparison of activity of recombinant forms of plasma kallikrein protease domain and of the full-length protein from blood - The activities of purified full-length plasma kallikrein and of the two recombinant deglycosylated protease domains were compared

in substrate specificity analysis using oxidized insulin B chain and in kinetic analysis using Pefachrome-PK. In the insulin B chain cleavage experiment (Table 3), the native full-length plasma kallikrein from blood and the two recombinant forms of the protease domain each cleaved peptide bonds with arginine or lysine at the S1 site. Each of the three purified kallikreins first cleaved the oxidized insulin B chain between Arg22-Gly23, thus converting the 30-residue insulin B chain into two fragments, an N-terminal 22-mer and a C-terminal 8-mer. In longer digestions, each protease then removed the last alanine residue from the insulin B chain C-terminal 8 residue peptide by cleaving the Lys29-Ala30 peptide bond (Table 3). To compare the substrate affinity and enzymatic activity of the purified full-length protein and of the recombinant forms of the protease domain, Pefachrome-PK was used to measure K_m and K_{cat}/K_m (Table 5). All three purified proteases had a similar K_m towards Pefachrome PK (~ 100 uM); the K_{cat}/K_m values were also similar to one another (~ 1 x 10^6). Thus the recombinant forms of the plasma kallikrein protease domain from *P. pastoris* or from Sf9 cells have catalytic activities similar to that of the full-length native plasma kallikrein.

Crystal Structures of Endo H- and mutagenically deglycosylated human plasma kallikrein protease domain - There is extensive sequence conservation between the protease domain of plasma kallikrein and of other trypsin-like serine proteases. The sequence encoding the protease domain of plasma kallikrein is 39 % identical with that of the protease domain of hepsin, and 38 % with that of bovine trypsin. Thus it is not surprising that the structures of both mutagenically- and Endo H-deglycosylated plasma kallikrein (Fig. 2) are typical of the chymotrypsin serine protease family. Core main residues of plasma kallikrein superimpose well on the corresponding residues in bovine trypsin-benzamidine (PDB Entry code 1C5P), while major differences in main chain occur in loop regions (Fig. 3a,b).

The plasma kallikrein protease shares lower sequence identity with the tissue kallikreins (33 % on average) than it does with trypsin (38 %) or hepsin (39%), but all of these proteases adopt similar three-dimensional structures. The superposition of the main-chain of plasma kallikrein protease onto that of the tissue kallikrein, hK6 (PDB Entry Code 1LO6) (50) is shown in Fig 3c). Of the available crystal structures of tissue kallikrein (hK1, Entry Code 1SPJ (48)**,** hK6, Entry Code 1LO6 (50), and pro-hk6, Entry Code 1GVL (51)), hK6 exhibits an S1 site that is more similar to that of plasma kallikrein than does hK1. The S1 site of hK1 contains a serine at position 226 whose side chain is hydrogen-bonded to one of the carboxylate oxygens of Asp189. The presence of serine at 226 enlarges the S1 site of hK1 compared to that of plasma kallikrein, as the latter has a glycine at this position.

The structures of the Endo H- and mutagenically-deglycosylated forms of the protease domain of human plasma kallikrein are very similar to one another (Fig. 4a). The protease domain has two six-stranded β barrels forming the substrate binding sites and the catalytic triad (His57, Asp102, and Ser195) is in a typical arrangement (Fig. 2). Benzamidine is bound at the S1 site, making hydrogen-bonded salt bridges with Asp189 (Fig. 5), as in other trypsin-like serine proteases (52).

The three N-linked glycosylation sites, Asn21, Asn72, and Asn113 occur within loops on the surface of the molecule. In agreement with the observation that the Endo-H- and mutagenically deglycosylated forms of the plasma kallikrein protease domain have similar proteolytic activities (Table 5), their core structures superimpose well on one another, with an rms deviation of 0.17 Å for the 61 Cα atoms used for superposition (Fig. 4a). However there are seven loops whose structures differ significantly between Endo-H- and mutagenically deglycosylated kallikrein (Fig. 4a,b). These differences occur in regions of crystal packing constraints. Despite the looser packing of mutagenically deglycosylated kallikrein (42 % solvent) compared with that of the EndoH-deglycosylated counterpart (28 % solvent), there are nevertheless about as many regions (six) influenced by packing in the latter

as in the former (five) (Fig. 4b). None of the major differences between the structures of Endo H- and mutagenically deglycosylated plasma kallikrein occur at the glycosylation sites (Fig 4a,b), suggesting that these replacements cause minor structural changes.

Inhibition of plasma kallikrein by short hydrogen bond-mediated serine protease inhibitors - Previously, we developed a large suite of small molecule inhibitors (2-(2-phenol)-benzimidazoles and 2-(2-phenol)-indoles) as serine protease inhibitors (61-67). Inhibition by such scaffolds is typically mediated by short hydrogen bonds at the active site, often via a multi-centered short hydrogen bond array involving the active site Ser195, a water in the oxyanion hole, the phenolate of the inhibitor and the benzimidazole (or indole) nitrogen of the inhibitor (see diagram in Table 6). The K_i values for a set of such inhibitors towards plasma kallikrein and other proteases are provided in Table 6a.

Plasma kallikrein stands out from the other trypsin-like serine proteases in the degree of inhibition by these simple short hydrogen bond-mediated inhibitors; the inhibitors on average are more potent for plasma kallikrein than for the other proteases. For example, the simplest scaffold, CG-02, is more potent for plasma kallikrein (K_i = 0.020 μM) than for any of the remaining proteases (K_i = 0.041 μM to 1.7 μM). Similarly, the benzimidazole isostere, CG-01, is more potent for plasma kallikrein (K_i = 0.22 μM) than for the other proteases (K_i = 0.55 μM to 24 μM). The average increase in potency toward plasma kallikrein over the other proteases is significant for both CG-01 (29-fold) and CG-02 (24-fold). Although these inhibitors were synthesized in previous programs without the help of the plasma kallikrein structure, the newly solved crystal structures presented here indicate many of the structural features in the substrate binding sites that are potentially useful in further improving this class of inhibitors, or in the design of inhibitors with different scaffolds.

DISCUSSION

Expression, purification and activity of recombinant forms of plasma kallikrein protease domain - The first step in determining the structure of plasma kallikrein was to establish an abundant source of crystallizable protein. The catalytic subunit of plasma kallikrein (the light chain), isolated through reduction and alkylation of the inter-chain disulfide bond, retains full enzymatic activity (10). However, because of the reduction and alkylation of disulfides and because of heterogeneity in glycosylation (involving 33 kDa and 36 kDa forms), the purified light chain is not an ideal source for crystallization studies. Thus, we set out to establish expression and purification systems to obtain sufficient amounts of recombinant plasma kallikrein protease domain for structural studies. In the first system (Table 1), the protease domain was expressed as a zymogen in *Pichia pastoria* and converted to an active protease by trypsin treatment. In the second system (Table 2), all three N-glycosylation sites, Asn377, Asn434, and Asn475, were mutagenized to glutamates and the mutant protease domain expressed in the baculovirus/Sf9 system as an unglycosylated zymogen, subsequently activated by trypsin treatment. In both systems, Cys364 and Cys484 were also changed to serines to prevent disulfide formation between the protease and non-protease domains. Each form of plasma kallikrein has enzymatic activity similar to that of native plasma kallikrein when assayed with HMW kininogen (unpublished results), with insulin B chain (Table 3) or with synthetic peptide substrates (Table 5 and unpublished results), in agreement with the previous observation that the light chain has activity similar to that of the full-length protein (10). The purified proteins are readily crystallizable, providing good systems for plasma kallikrein structural studies.

N-linked oligosaccharides affect the activity of some serine proteases (68) but not others (50,56). In the protease subunit of plasma kallikrein, all three N-glycosylation sites are on the surface of the protein, removed from substrate binding pockets (Fig 2), and thus glycosylation should not affect the enzymatic

activity. The similarity in enzymatic activities among full-length plasma kallikrein and the two recombinant forms of the protease domain support this expectation. The N-linked glycosylation of plasma kallikrein may facilitate its biosynthesis and/or stability; the detailed function(s) of glycosylation remain to be elucidated.

Structural features of substrate binding site S1 and implications in structure-based drug design - At the bottom of the S1 site of plasma kallikrein lies negatively-charged Asp189, which forms hydrogen-bonded salt bridges with positively-charged Arg or Lys residues that are the P1 residues of protein or peptide substrates for this protease. This interaction with Asp189 constitutes the structural basis for the selectivity of plasma kallikrein for Arg or Lys at the S1 site (Fig. 5 and reference 52). The role of Asp189 in recognition at the S1 site of plasma kallikrein is apparent in the structure of the benzamidine complex, in which Asp189 makes hydrogen-bonded salt bridges with the amidino group (Fig 5), as in other trypsin-like serine proteases (52). Other hydrogen bonds to benzamidine involve the carbonyl group of Gly219 and a water molecule. Although the structures of the S1 sites of trypsin-like serine proteases resemble another, one important difference often occurs at residue 190 (52,62,66). Trypsin-like serine proteases can be classified into three subgroups: Ala190, Ser190, or Thr190 proteases. The Ser190 and Ala190 group proteases each constitute about 32% of the more than 100 trypsin-like serine proteases in the human genome (67). Plasma kallikrein is an Ala190 protease, with no hydrogen-bonding capacity for the Ala190 side chain. In Ala190 protease-amidine complexes (such as plasma kallikrein-benzamidine, Fig. 5), the sole hydrogen-bonding partner to one of the amidine protons is a water molecule, whereas in the corresponding Ser190 protease complexes (such as trypsin-benzamidine), there are two hydrogen-bonding acceptors for the amidine proton, $O\gamma_{Ser190}$ and the water. Thus, inhibitors designed to displace the water are often highly selective for Ser190 proteases over Ala190 counterparts (such as plasma kallikrein) because of the resulting hydrogen-bonding deficit at the S1 site (62,66,67). Thus, viewed as an anti-target, plasma kallikrein offers the S1 site as a potential locus for engineering selectivity against this protease, if the target of interest is a Ser190 protease, such as urokinase. For example, CG-06 (Y=Cl, X=N, R=H, see Table 6c), whose chloro ortho to the amidine displaces the water from the S1 site, is 90-fold selective for urokinase (K_i = 0.009 μM) over plasma kallikrein (K_i = 0.80 μM) (62).

Another important structural feature bordering the S1 site of plasma kallikrein is Lys192. Sequence analysis of the trypsin-like proteases of the human genome shows that the majority of proteases contain glutamine at residue 192 (60 %), followed by lysine (14 %), and glutamate (5.4 %). As proteases with Lys192 are a minority, a potential strategy for the development of inhibitors selective for plasma kallikrein could rely on specific interactions with Lys192. For example, in the serine protease inhibitor series that make short hydrogen bonds at the active site (64-67) (Table 6a), substituents para to the phenol interact with residue 192. In a model of the plasma kallikrein-bound structure of CG-05, one of the oxygens of the succinate makes a hydrogen-bonded salt bridge with Lys192 (Fig. 6). The role in potency and selectivity of residue 192 is revealed in the inhibition data in Table 6a,b in which both residue 192 of the protease as well as the interacting group of the inhibitor are varied.

The data in Table 6a show that addition of a succinate para to the phenol of the inhibitor results in a significant increase in potency toward serine proteases with a lysine at position 192 (plasma kallikrein and fVIIa). In the case of plasma kallikrein, addition of a succinate para to the phenol incurs a 27-fold increase in potency for the benzimidazole (CG-04, K_i = 0.0081 μM) over that of the parent (CG-01, K_i = 0.215 μM), or a 29-fold increase for the indole (CG-05, K_i = 0.0007 μM) over that of the parent (CG-02, K_i = 0.020 μM). Similarly, for fVIIa the succinate results in increases in potency of 12- and 5.6-fold. The increase in potency toward plasma kallikrein and fVIIa imparted by succinate reflects a favorable interaction of this

group with Lys192. Sizeable increases are also observed with acetate para to the phenol (data not shown).

The effect of incorporating a succinate into the inhibitor scaffold is generally minor for the Gln192 proteases (Table 6a). Changes in potency resulting from the succinate range from 0.26-fold to 4.3-fold, with an average change of 1.4 ± 0.8 for eight Gln192 proteases (not all of which are shown in Table 6). For the Glu192 protease, thrombin, the succinate group decreases potency by as much as 8.6-fold (for CG-04, K_i = 210 μM compared with CG-01, K_i = 24 μM), probably reflecting an unfavorable electrostatic repulsion between this group and the Glu192 side chain. Thus, the presence of Lys192 of plasma kallikrein provides a means for introducing selectivity toward plasma kallikrein and against other non-Lys192 serine proteases (Table 6b). Conversely, it might be expected that incorporation of an appropriate basic group ortho to the phenol would allow development of selectivity against plasma kallikrein because of the resulting unfavorable electrostatic repulsion. Note that features other than the Lys192 side chain must also play important roles in selectivity of the compounds in Table 6. For example, the succinate-containing CG-05 shows substantial selectivity, 84-fold, toward plasma kallikrein against fVIIa, even though both are Lys192 proteases.

Implications for enzymatic specificity of structural features of substrate binding site S2 - The S2 site of plasma kallikrein is bordered by the His57 and Trp215 side chains and by the main chain of residues Ser97 to Gly99 (Fig. 7a). The physiological substrate of plasma kallikrein (kininogen) is cleaved in two positions, in which either the Phe or Met side chain occupies the S2 site (69). This preference for Phe or Met at S2 is in contrast to that of thrombin or of trypsin, which prefer a Pro at S2 (70). Substrates with Pro at the S2 site are sub-optimal substrates for plasma kallikrein (71).

In order to investigate the structural basis for the S2 site selectivity of plasma kallikrein *versus* that of trypsin, tri-peptide portions (P1, P2 and P3) of substrates were docked into these proteases, and the resulting models of the catalytic complexes compared. Key differences in the S2 site of trypsin (or of thrombin) versus that of plasma kallikrein is the positions of loop residues 94 to 99 and the identity of residue 99, Leu in trypsin (Fig. 7b) *versus* Gly in plasma kallikrein (Fig. 7a). The lack of a side chain at position 99 creates space at the S2 site of plasma kallikrein (Fig. 7a) that is absent in trypsin (Fig. 7b) or in thrombin. In addition, loop residues 95 through 99, defining a wall of the S2 site of plasma kallikrein, are shifted toward the substrate in plasma kallikrein compared to their position in trypsin, thus creating a cavity for accommodation of a Phe or Met side chain at the S2 site of plasma kallikrein (Fig. 7a). This cavity is absent in trypsin (Fig. 7b). Thus the structure of plasma kallikrein affords a convincing explanation for the S2 preference of this protease.

Structural features of substrate binding site S1' to harness for drug design - The S1' site of trypsin, formed from residues 35 to 42 and 57 to 60, is relatively shallow. The location of the S1' site on the superimposed Cα structures of trypsin and plasma kallikrein is shown in Fig. 3a. In many important protease targets like tryptase, thrombin, hepsin, fVIIa, uPA, tPA and plasmin, additional loop residues (the 60's loop) inserted after residue 59 in trypsin, provide another wall to the S1' site. This additional boundary creates a well-defined S1' binding pocket with distinct structural features that can be exploited in the development of selective inhibitors. Plasma kallikrein is one such protease that contains a "60's loop", and an important difference between plasma kallikrein and trypsin is the insertion of residues Gly60A, Leu60B, and Pro60C at the S1' site. This difference is also apparent in the comparison of plasma kallikrein *versus* tissue kallikrein, hK6 (Fig 3c). The location of this loop is shown in Fig. 3a,c, in which the first residue, Gly60A, is labeled. The S1' site region of plasma kallikrein is further elaborated from that of trypsin or of hK6 due to insertion into plasma kallikrein of additional loop residues after residue 35.

In our other serine protease inhibitor development programs, the S1' site has been used extensively for engineering selectivity

towards particular protease targets into inhibitors that make short hydrogen bonds at the active site (61-66). In plasma kallikrein, the side chains of residues Leu41, Leu60B, and Trp60D form a binding pocket that is readily accessible from the para position of the distal phenyl group of a typical, simple short hydrogen bond-mediated scaffold, like CG-01 (Fig. 8). In addition, both the main-chain and side chain of Asp60 are accessible for interactions from substituents on the *meta* position of the distal phenyl of bound CG-01. Fortunately, this region is free of crystal packing interactions in the Endo-H-deglycosylated kallikrein, and should therefore reflect the true solution structure. The unique S1' site of plasma kallikrein is thus an attractive locus for engineering selectivity into plasma kallikrein inhibitors using previously developed, active site-directed, short hydrogen bond-mediated inhibitor scaffolds.

ACKNOWLEDGEMENT
We are grateful to the staff at the Advanced Light Source, Berkeley for help with crystallographic data collection. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, Materials Sciences Division, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098 at Lawrence Berkeley National Laboratory.

REFERENCES

1. Kraut, H., Frey, E.K. and Werle, E. (1930) *Hoppe-Seyler's Zeitschrifft fur Physiologische Chemie* 189, 97-106
2. Werle, E. (1934) *Biochem. Z.* 269, 415-434
3. Clements, J.A., Hooper, J.D. and Odorico, D.M. (2004) in *Handbook of Proteolytic Enzyme 2nd Edition* (Barret, A.J., Rawlings, N.D. & Woessner, J.I. eds) pp 1569-1577, Elsevier, London
4. Yousef, G.M. and Diamandis, E.P. (2001) *Endocrine Reviews* 22, 184-204
5. Hathaway, W.E., Belhasen, L.P., Hathaway, H.S. (1965) *Blood* 26, 521-532
6. Wuepper, K.D. and Cochrane, C.G. (1972) *J. Exp. Med.* 135, 1-20
7. Chung, D.W., Fujikawa, K., McMullen, B.A. and Davie, E.W. (1986) *Biochemistry* 25, 2410-2417
8. Colman, R.W. (2004) In *Handbook of Proteolytic Enzymes, 2 edn* (Barrett, A.J., Rawlings, N.D. and Woessner, J.F. eds), pp 1644-1651, Elsevier, London
9. McMullen, B.A., Fujikawa, K., and Davie, E.W. (1981) *Biochemistry* 26, 2050-2056.
10. van der Graaf, F., Tans, G., Bouma, B.N., and Griffin, J.H. (1982) *J. Biol. Chem.* 257, 14300-14305
11. Borgono, C.A. and Diamandis, E.P. (2004) *Nat Rev Cancer* 4, 876-890
12. Kitchens, C.S. (2002) *Arch. Pathol. Lab. Med.* 126, 1382-1386
13. Ciechanowicz, A., Bader, M., Wagner, J., and Ganten, D. (1993) *Biochem. Biophys. Res. Commun.* 197, 1370-1376
14. Mandle, R.J., Colman, R.W., and Kaplan, A.P. (1976) *Proc. Natl. Acad. Sci. U. S. A.* 73, 4179-4183
15. Hasan, A.A., Zisman, T., and Schmaier, A.H. (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 3615-3620
16. Shariat-Madar, Z., Mahdi, F., and Schmaier, A.H. (1999) *J. Biol. Chem.* 274, 7137-7145
17. Colman, R.W., Pixley, R.A., Najamunnisa, S., Yan, W., Wang, J., Mazar, A., and McCrae, K.R. (1997) *J. Clin. Invest.* 100, 1481-1487
18. Herwald, H., Dedio, J., Kellner, R., Loos, M., and Muller-Esterl, W. (1996) *J. Biol. Chem.* 271, 13040-13047
19. Joseph. K., Ghebrehiwet, B., Peerschke, E.I., Reid, K.B., and Kaplan, A.P. (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 8552-8557
20. Mahdi, F., Shariat-Madar, Z., Todd, R.F., 3rd, Figueroa, C.D., and Schmaier, A.H. (2001) *Blood* 97, 2342-2350
21. Schmaier, A.H. (2002) *J. Clin. Invest.* 109, 1007-1009
22. Shariat-Madar, Z., Mahdi, F., and Schmaier, A.H. (2002) *J. Biol. Chem.* 277, 17962-17969
23. Shariat-Madar, Z., and Schmaier, A.H. (2004) *J. Endotoxin. Res.* 10, 3-13
24. Colman, R.W. (1969) *Biochem. Biophys. Res. Commun.* 35, 273-279
25. Ichinose, A., Fujikawa, K., and Suyama, T. (1986) *J. Biol. Chem*. 261, 3486-3489
26. Hauert, J., Nicoloso, G., Schleuning, W.D., Bachmann, F., and Schapira, M. (1989) *Blood* 73, 994-999
27 Schmaier A.H. (2000) *Current Opinion in Hematology* 7(5), 261-265
28. Kaplan, A.P., Kay, A.B., and Austen, K.F. (1972) *J. Exp. Med.* 135, 81-97
29. Wachtfogel, Y.T., Kucich, U., James, H.L., Scott, C.F., Schapira, M., Zimmerman, M., Cohen, A.B., and Colman, R.W. (1983) *J. Clin. Invest.* 72, 1672-1677
30. Ni, Z.J. and Wagman, A.S. (2004) *Curr. Opin. Drug Discov. Devel.* 7, 446-459
31. Adams, J., and Kauffman, M. (2004) *Cancer Invest.* 22, 304-311

32. Docherty, A.J., Crabbe, T., O'Connell, J.P., and Groom, C.R. (2003) *Biochem. Soc. Symp.* 70, 147-161
33. Maryanoff, B.E., (2004) *J. Med. Chem*. 47, 769-787
34. Wieland, H.A., Laux, V., Kozian, D., Lorenz, M. (2003) *Curr. Opin. Investig. Drugs.* 4, 264-271
35. Kyrle, P.A., and Eichinger, S. (2002) *Vasa.* 31, 163-166
36. Yu, A., Jacobson, S.H., Bygden, A., and Egberg, N. (2002) *Clin. Chem. Lab. Med.* 40, 21-26
37. Levine, M.N., Raskob, G., Landefeld, S., and Kearon, C. (2001) *Chest* 119, (1 Suppl.), 108S-121S
38. Schechter, I. and Berger, A. (1967) *Biochem. Biophys. Res. Commun*. 27, 157-162
39. Banner, D. W., D'Arcy, A., Chène, C., Winkler, F. K., Guha, A., Konigsberg, W. H., Nemerson, Y., and Kirchhofer, D. (1996). *Nature* 380, 41-46
40. Parlow, J. J., Case, B. L., Dice, T. A., Fenton, R. L., Hayes, M. J., Jones, D. E., Neumann, W. L., Wood, R. S., Lachance, R. M., Girard, T. J., Nicholson, N. S., Clare, M., Stegeman, R. A., Stevens, A. M., Stallings, W. C., Kurumbail, R. G., and South, M. S. (2003) *J. Med. Chem.* 46, 4050-4062
41. Brandstetter, H., Kühne, A., Bode, W., Hubert, R., von der Saal, W., Wirthensohn, K., and Engh, R. A. (1996). *J. Biol. Chem.* 271, 29988-29992
42. Wiley, M. R., Weir, L. C., Briggs, S., Bryan, N. A., Buben, J., Campbell, C., Chirgadze, N. Y., Conrad, R. C., Craft, T. J., Ficorilli, J. V., Franciskovich, J. B., Froelich, L. L., Gifford-Moore, D. S., Goodson, T. Jr., Herron, D. K., Klimkowski, V. J., Kurz, K. D., Kyle, J. A., Masters, J. J., Ratz, A. M., Milot, G., Shuman, R. T., Smith, T., Smith, G. F., Tebbe, A. L, Tinsley, J. M., Towner, R. D., Wilson, A., and Yee, Y. K. (2000) *J. Med. Chem.* 43, 883-899
43. Maignan, S. and Mikol, V. (2001) *Curr. Top. Med. Chem.* 1, 161-174
44. Walport, M.J. (2001) *N. Engl. J. Med.* 344, 1058-1066
45. Colman, R.W. (1999) *Immunopharmacology* 43, 103-108
46. Han, E.D., MacFarlane, R.C., Mulligan, A.N., Scafidi, J., and Davis, A.E. 3rd. (2002) *J. Clin. Invest.* 109, 1057-1063
47. De La Cadena, R. A., Suffredini, A. F., Page, J. D., Pixley, R. A., Kaufman, N., Parrillo, J. E., and Colman, R. W. (1993) *Blood* 81, 3313-3317
48. Laxmikanthan, G, Blaber, S. I., Bernett, M. J., Scarisbrick, I. A., Juliano, M. A and Blaber, M. (2005) *Proteins: Structure, Function, and Bioinformatics* 58, 802-814
49. Katz, B. A., Liu, B., Barnes, M. and Springman, E. (1998) *Protein Science* 7, 875-885
50. Bernett, M. J., Blaber, S. I, Scarisbrick, I. A., Dhanarajan, P., Thompson, S. M, and Blaber, M. (2002). *J. Biol. Chem. 277*, 24562-24570
51. Gomis-Ruth, F. X., Bayes, A, Sotiropoulou, G., Pampalakis, G., Tsetsenis, T., Villegas, V., Aviles, F. X., Coll, M. (2002) *J. Biol. Chem*. 277, 27273-27281
52. Katz, B. A., Mackman, R., Luong, C., Radika, K., Martelli, A., Sprengeler, P. A., Wang, J., Chan, H., and Wong, L. (2000) *Chem. Biol.* 7, 299-312
53. Kuzmic, P., Sideris, S., Cregar, L. M., Elrod, K. C., Rice, K. D. and Janc, J. W. (2000). *Anal. Biochem.* 281, 62-67
54. Otwinowski, Z. and Minor, W. (1993) In *Proceedings of the CCP4 Study Weekend*, SERC Daresbury Laboratory, Warrington, pp. 56–62
55. Pandey, P., Babine, R. E., Gorga, J. C., Seidl, K. J., Gelfand, E., Weaver, D. T., Abdel-Meguid, S. S., and Strickler, J. E. (2005) *J. Biol. Chem*. 280, 4704-4712
56. Somoza, J. R., Ho, J. D., Luong, C., Ghate, M., Sprengeler, P. A., Mortara, K., Shrader, W. D., Sperandio, D., Chan, H., McGrath, M. E. and Katz, B. A. (2003) *Structure* 11, 1-20
57. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne P. E. (2000) *Nucleic Acids Res.* 28, 235-242
58. Perrakis, A., Harkiolaki, M., Wilson, K. S. and Lamzin, V. S. (2001) *Acta Cryst.* D57, 1445-1450
59. Brünger, A. T. (1990). X-PLOR Manual, Version 3.1: *A system for X-ray Crystallography and NMR*; Yale University: New Haven, CT; pp 187-206

60. Brünger, A. T. (1992). *Nature* 355, 472-474.
61. Verner, E., Katz, B. A., Spencer, J., Allen, D., Hataye, J., Hruzewicz, W., Hui, H., Kolesnikov, A., Li, Y., Luong, C., Martelli, A., Radika, K., Rai, R., She, M., Shrader, W., Sprengeler, P., Trapp, S., Wang, J., Young, W. B., and Mackman, R. L. (2001) *J. Med. Chem.*, 44, 2753-2771.
62. Mackman, R. L., Katz, B. A., Breitenbucher, J. G., Hui, H. C., Verner, E., Luong, Liu, L. and Sprengeler, P. A. (2001). *J. Med. Chem.* 44, 3856-3871
63. Mackman, R. L., Hui, H. C., Breitenbucher, J. G., Katz, B. A., Luong, C., Martelli, A., McGee, D., Radika, K., Sendzik, M., Spencer, J. R., Sprengeler, P. A., Tario, J., Verner, E. and Wang, J. (2002) *Bioorg. Med. Chem. Lett.* 12, 2019-2022
64. Katz, B. A., Spencer, J. R., Elrod, K., Luong, C., Mackman, R. L., Rice, M., Sprengeler, P. A., Allen, D., Janc, J. (2002) *J. Am. Chem. Soc.* 124, 11657-11668
65. Katz, B. A., Elrod, K., Verner, E., Mackman, R. L., Luong, C., Shrader, W. D., Sendzik, M., Spencer, J. R., Sprengeler, P. A., Kolesnikov, A., Tai, V. W., Hui, H. C., Breitenbucher, J. G., Allen, D., Janc, J. W., (2003) *J. Mol. Biol.* 329, 93-120
66. Katz, B. A., Sprengeler, P. A., Luong, C., Verner, E., Elrod, K., Kirtley, M., Janc, J., Spencer, J., Breitenbucher, J. G., Hui, H., McGee, D., Allen, D., Martelli, A. and Mackman, R. L. (2001) *Chem. Biol.* 8, 1107-1121
67. Katz, B. A., Luong, C., Ho, J. D., Somoza, J. R., Gjerstad, E., Tang, J., Williams, S. R., Verner, E., Mackman, R. L., Young, W., Sprengeler, P. A., Chan, H., Mortara, K., Janc, J. W., and McGrath, M. E. (2004) *J. Mol. Biol.* 344, 527-547
68. Oka, T., Hakoshima, T., Itakura, M., Yamamori, S., Takahashi, M., Hashimoto, Y., Shiosaka, S., and Kato, K. (2002) *J. Biol. Chem.* 277, 14724-14730
69. Kitamura, N., Takagaki, Y., Furuto, S., Tanaka, T., Nawa, H., and Nakanishi, S. (1983) *Nature* 305, 545-549
70. Harris, J. L., Backes, B. J., Leonetti, F., Mahrus, S., Ellman, J. A., Craik, C. S. (2000) *Proc. Natl. Acad. Sci. U. S. A.* 97, 7754-7759
71. McRae, B. J., Kurachi, K., Heimark, R. L., Fujikawa, K., Davie, E. W., and Powers, J. C. (1981) *Biochemistry* 20, 7196-7206.

FIGURE LEGENDS:

Fig 1: Characterization of non-glycosylated recombinant plasma kallikrein protease domain from conditioned Sf9 cell media. The purified protein was analyzed with a reverse phase Zobax CN column using water/0.1% TFA (mobile phase A) and ACN/0.08% TFA (mobile phase B), and with electro-spray mass spectrometry using a Q-Star Mass Spectrometer. The amino acid sequence of the mutant plasma kallikrein construct is shown. The Asn377->Glu, Asn434->Glu and Asn475->Glu mutations at the N-glycosylation sites are colored green, and the Cys364->Ser and Cys484->Ser mutations purple. The N-terminal residues (Asn357 through Arg371) and C-terminal residues (Ala613 through Ala619) residues, absent in the final purified protein, are colored red. (These C-terminal residues are removed during trypsin treatment). Each of the cysteine pairs involved in the 4 disulfides (5) is color coded, and given the appropriate superscript (S1, S2, S3, or S4).

Fig 2: Secondary structure of plasma kallikrein. β-sheet, α-helix and remaining structural elements are yellow, red, and cyan, respectively. Bound benzamidine, the active site residues (His57, Asp102, Ser195) and the glycosylation sites (Asn21, Asn72, and Asn113) are shown.

Fig 3: (a): Superposition of main-chains of trypsin (orange) and of enzymatically deglycosylated plasma kallikrein protease domain (blue) based on 61 core residues (16-21, 28-33, 40-46, 51-57, 101-108, 139-143, 155-158, 189-197, 212-217, 227-229) (rms = 0.33 Å). The active site triad (His57, Asp102, and Ser195) is shown along with Asp189 at the base of the S1 site. (b): Sequence alignment of trypsin and plasma kallikrein based on structural superposition involving the enzymatically deglycosylated form (pkal 1). For un-italicized sequence segments, the differences between corresponding Cα coordinates of the superimposed trypsin and plasma kallikrein structures are less than 1.0 Å. Compared sequences in italics separated by dashes usually involve different loop sizes, adopt different conformations, and involve corresponding residues with differences in coordinates greater than 1.0 Å. Compared sequences in italics not separated by dashes correspond to segments of structure of the same sequence length, and with similar conformations, but with global translational or rotational differences. Active site residues are in red font, glycosylation sites in blue, and the Cys122->Ser mutation in blue. (c): Superposition of main-chain of tissue kallikrein, hK6, PDB code 1LO6 (50) (orange) onto that of enzymatically deglycosylated plasma kallikrein protease domain (blue) based on 68 core residues (16-21, 32-33, 39-46, 51-57, 87-91, 102-111, 136-143, 155-158, 189-197, 209-215, 227-230 (rms = 0.29 Å).

Fig 4: (a). Superposition of main-chains of mutagenically deglycosylated plasma kallikrein protease domain (cyan) onto the enzymatically deglycosylated form (orange) based on 61 core residues (16-21, 28-33, 40-46, 51-57, 101-108, 139-143, 155-158, 189-197, 212-217, 227-229) (rms = 0.17 Å). The active site triad (His57, Asp102, and Ser195) is shown along with Asp189 at the base of the S1 site. The Asn->Glu glycosylation site mutations are labeled in blue. Positions where the structures of the two forms of the plasma kallikrein protease domain differ significantly from one another are labeled in red. (b). Summary of the loop differences between the two recombinant plasma kallikrein forms.

Fig 5: S1 site of plasma kallikrein-benzamidine superimposed on a $|F_o|-|F_c|$ omit map contoured at 2.75 σ.

Fig 6: Model of CG-05 bound to plasma kallikrein, showing expected interaction of the diacid group of the inhibitor with Lys192 of the enzyme, along with the multi-centered short hydrogen bond array observed at the active site of the trypsin complex. Short hydrogen bonds are cyan.

Fig 7. (a): Model of tetrahedral intermediate complex of plasma kallikrein bound by P1, P2, P3 portion of substrate with Arg at S1 site and Met at S2 site. (b): Model of tetrahedral intermediate complex of trypsin bound by P1, P2, P3 portion of a substrate with Arg at S1 site and Pro at S2 site.

Fig 8: Model of CG-01 bound to plasma kallikrein, showing the expected location of the terminal phenyl group of the inhibitor in the S1' site.

Table 1. Purification of recombinant plasma kallikrein from conditioned *P. pastoris* media

Step	Sample	Total protein (mg)	Specific activity (mOD/(min*ug))	Total activity (mOD/min)	Activity yield (%)
1	Concentrated Media	1200	186	223×10^6	100
2	Q Flow-through	474	286	135×10^6	60.8
3	Source S pool	94.0	591	55.5×10^6	24.5
4	EndoH treatment and activation	59.0	750	44.3×10^6	19.9
5	Benzamidine Sepharose pool	30.3	1010	30.5×10^6	13.7
6	Source S Peak1	10.5	1090	11.5×10^6	5.20

Table 2. Purification of recombinant kallikrein from conditioned Sf9 media

Step	Sample	Total protein (mg)	Specific activity (mOD/(min*ug))	Total activity (mOD/min)	Activity Yield%
1	Conditioned media (3 l)	213	56.3	12.0×10^6	100
2	Phenyl Sepharose pool	39.0	231	9.01×10^6	75
3	Q Flow-through	13.2	488	6.49×10^6	54
4	Activation	12.0	375	4.50×10^6	38
5	Benzamide Sepharose pool	3.60	289[a]	1.04×10^{6a}	
6	Superdex 200 pool	1.28	1286	1.65×10^6	14
7	Source S pool	0.70	1288	0.902×10^6	7.6

[a]The apparent specific activity is low because the protein was eluted with benzamidine and then assayed in the presence of 250 μM benzamidine.

Table 3: Relative amounts (percentages) of peptide products produced by incubating oxidized insulin B chain with various forms of plasma kallikrein for 1 hour (a) or 24 hours (b).

(a):	Plasma Kallikrein (1 hour incubation)			
Peptide	No Enzyme	Full-Length Kallikrein	Endo H-deglycosylated	Mutagenically deglycosylated
Peak IV	100%	58 %	64 %	27 %
Peak III	0	23 %	20 %	41 %
Peak II	0	16 %	14 %	27 %
Peak I	0	2.3%	1.3 %	4.9 %

(b):	Plasma Kallikrein (24 hour incubation)			
Peptide	No Enzyme	Full-Length Kallikrein	Endo H-deglycosylated	Mutagenically deglycosylated
Peak IV	100%	5.6%	1.7 %	1.1 %
Peak III	0	55 %	61 %	59 %
Peak II	0	25 %	1.7 %	9.7 %
Peak I	0	15 %	36 %	30 %

Peak IV: FVNQHLC(SO3H)GSHLVEALYLVC(SO3H)GE**R**GFFYTP**K**A (oxidized insulin B chain)

Peak III: FVNQHLC(SO3H)GSHLVEALYLVC(SO3H)GER

Peak II: GFFYTPKA

Peak I: GFFYTPK

The cleavage sites of oxidized Insulin B chain are shown by highlighting the P1 residue (R or K) in bold.

Table 4: Crystallography of the two recombinant forms of the protease domain of human plasma kallikrein

Parameters[a]		
codename	pkal_bnz_1	pkal_bnz_2
Deglysosylation form	enzymatic	mutagenic
RCSB access code	2ANW	2ANY
pH	6.5	6.5
space group, lattice parameters[b]	$P2_12_12$	$P2_12_12_1$
No. atoms (including disorder)[a]	2982	3658
No. solvent molecules / ions	214	430
No. discretely disordered groups[c]	2	8
No. residues / groups with refined occs[d]	6	24
Diffraction Statistics		
ADSC CCD Detector	Q-315	Q-210
ALS Beam-line	5.0.2	5.0.1
Date	7/23/04	9/27/04
Highest resolution (Å)	1.85	1.40
No. observations	154041	332035
No. merged reflections	22313	50133
average redundancy	6.9	6.6
I/σ cutoff for reflection retention	0.0	0.0
<I>/<σ> (overall)	24.4	24.6
<I>/<σ> (highest resolution shell)[e]	6.1	4.0
R_{merge} (%)[f]	4.9	4.3

R_{merge} (%) (highest resolution shell)[e]	34.0	20.6
Refinement Statistics		
Refinement range (Å)	7.0-1.85	7.0-1.40
R_{cryst}, completeness[g]	22.1, 99.4	19.0, 99.1
highest resolution shell (Å)	1.93-1.85	1.46-1.40
completeness at highest resolution	99.1	94.8
free R_{cryst}[h]	28.3	21.2
RMS Deviations[i]		
bond lengths (Å)	0.018	0.016
bond angles (°)	3.2	2.9
torsion angles (°)	26.6	25.6
average protein B value (Å^2)		
Protein	30(12)	15(6)
Inhibitor	16(1)	13(1)
Ramachandran analysis		
No. (and percentage) of residues in:		
(a) most favored regions	179 (88.2)	181 (90.5)
(b) additional allowed regions	22 (10.8)	19 (9.5)
(c) generously allowed regions	2 (1.0)	0 (0.9)
(d) disallowed regions	0	0 (0.5)

[a]Restrained, isotropic temperature factors were refined. The CHARMM polar atom force field (toph19xb2.pro) of the XPLOR package (59) was used in final refinements. The atom count includes polar hydrogens. The four terminal residues are not visible in either structure. [b]The lattice parameters for the structures

are: pkal-bnz 1: a = 79.89, b = 63.19, c = 50.31 Å; pkal-bnz 2: a = 55.19, b = 57.36, c = 79.85 Å. [c]Not including waters. [d]Density for all side chain atoms or for terminal atoms in these groups was weak or absent and temperature factors were high. Discretely disordered groups are not included in this category. Occupancies for poorly defined groups of atoms were refined. [e]These statistics refer to the highest resolutions shell used for refinement. [f]R_{merge} = $\sum_h\sum_i |I(h)_i - \langle I(h)_i \rangle| / \sum_h\sum_i I(h)_i$. [g]The first value is the conventional R-factor (in percent), R_{cryst} = $\Sigma(||F_o|-|F_c||)/\Sigma|F_o|$ x 100 (for reflections from 7.0 Å to the upper resolution). The second value is the overall completeness in percent. [h]Cross validation R-factor (60) using 10% of the data withheld from the refinement. [i]Root mean square deviations from ideal bond lengths and bond angles.

Table 5. Enzyme kinetic parameters of full-length plasma kallikrein and of the two recombinant forms of the protease domain

Substrate (Pefachrome-PK): 2AcOH-H-D-But-CHA-Arg-pNA			
	Full-length kallikrein	Endo H-deglycosylated	Mutagenically deglycosylated
K_m (μM)	108 ± 7.4	117 ± 12	91 ± 15
K_{cat}/K_m (1/sec*M)	0.70×10^6	1.7×10^6	2.4×10^6

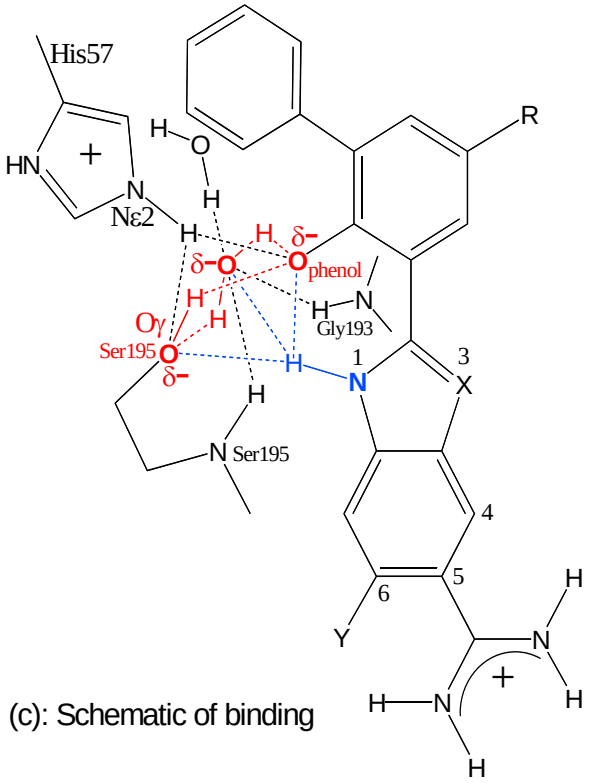
Table 6: Inhibition constants of inhibitors for full length plasma kallikrein and for other proteases

(a): K_i values (μM) for binding of inhibitors (Y=H) to human plasma kallikrein and to a panel of other trypsin-like serine proteases

CG:	X	R[a]	Lys192 pKal	Lys192 fVIIa	Gln192 uPA	Gln192 tPA	Gln192 trypsin	Gln192 fXa	Glu192 thrombin
01	N	H	0.215	0.88	0.605	0.55	8.7	3.2	24
02	CH	H	0.020	0.34	0.041	0.070	0.25	0.36	1.7
04	N	suc	0.0081	0.074	0.24	2.1	7.3	2.4	210
05	CH	suc	0.0007	0.061	ND	0.17	0.89	0.21	6.4

(b): Selectivities of inhibitors for plasma kallikrein against other trypsin-like proteases (K_i(other protease)/K_i(kallikrein))

CG:	X	R[a]	Lys192 fVIIa	Gln192 uPA	Gln192 tPA	Gln192 trypsin	Gln192 fXa	Glu192 thrombin
01	N	H	4.1	2.8	2.6	41	15	110
02	CH	H	17	2.1	3.5	13	18	88
04	N	suc	9.1	30	260	900	300	26000
05	CH	suc	84	ND	240	1200	290	8800



(c): Schematic of binding

(c): Schematic showing the interactions between serine proteases and small molecule inhibitors with scaffolds: 2-(2-phenol)-benzimidazoles and 2-(2-phenol)-indoles. Inhibition by such scaffolds is typically mediated by short hydrogen bonds at the active site, often via a multi-centered short hydrogen bond array. Short H-bonds involving only oxygens are red; those involving a nitrogen are blue.

[a]suc = succinate = $CH(COOH)CH_2COOH$.

Fig.1

NTGDNSV**S**^{364}TTKTSTRIVGGT**E**^{377}SSWGEWPWQVSLQVKLTAQRH
LC^{S1}GGSLIGHQWVLTAAHC^{S1}FDGLPLQDVWRIYSGIL**E**^{434}LSDITKDT
PFSQIKEIIIHQNYKVSEGNHDIALIKLQAPL**E**^{475}YTEFQKPI**S**^{484}LPSKG
DTSTIYTNC^{S2}WVTGWGFSKEKGEIQNILQKVNIPLVTNEEC^{S3}QKRYQ
DYKITQRMVC^{S3}AGYKEGGKDAC^{S4}KGDSGGPLVC^{S2}KHNGMWRLVGI
TSWGEGC^{S4}ARREQPGVYTKVAEYMDWILEKTQSSDGKAQMQSPA

Calculated MW (linear I^{372}-A^{619}): 27827

Calculated MW (I^{372}-K^{612}): 27105
(with 4 disulfide bonds)

Purified kallikrein: I^{372}-K^{612}

Observed MW: 27105

Fig. 2

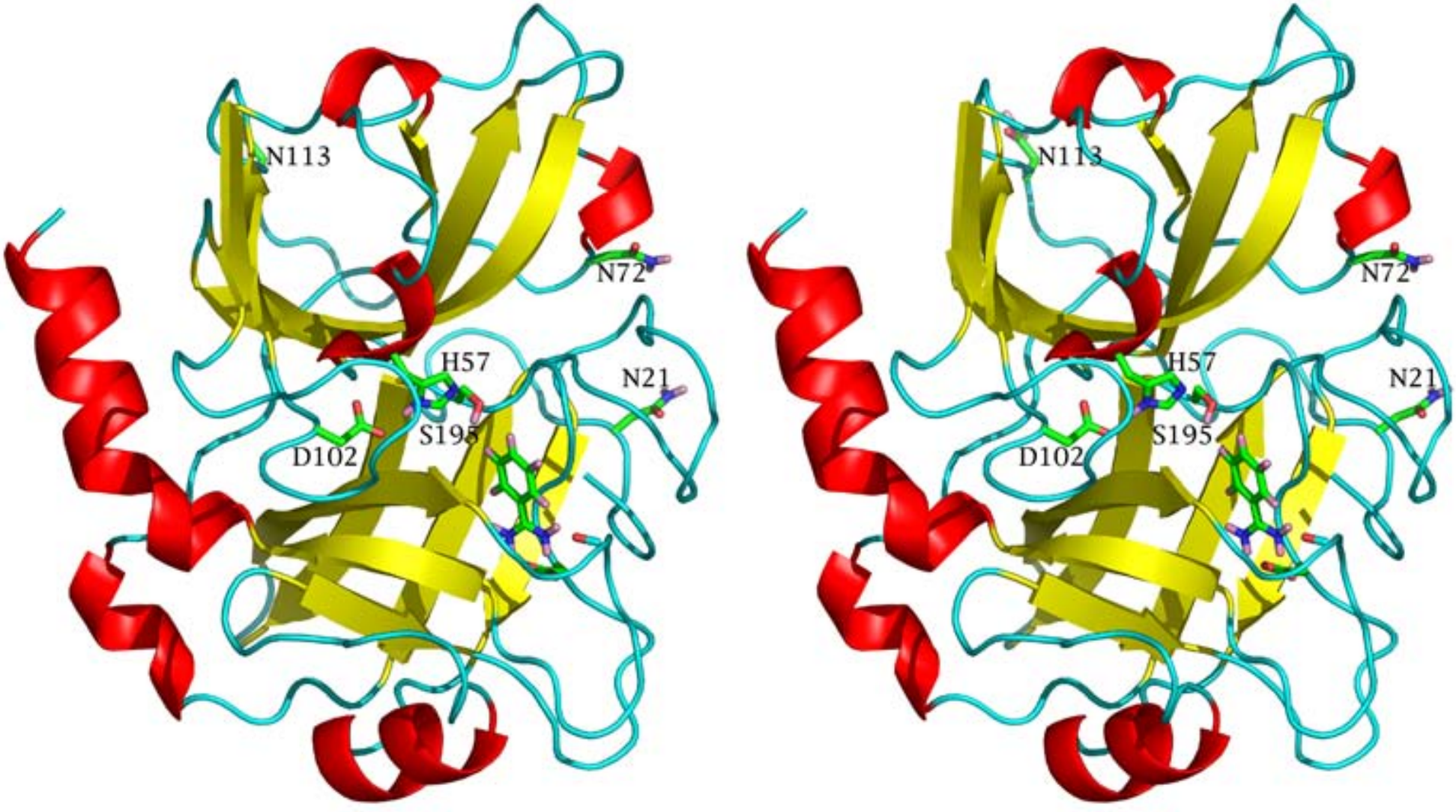
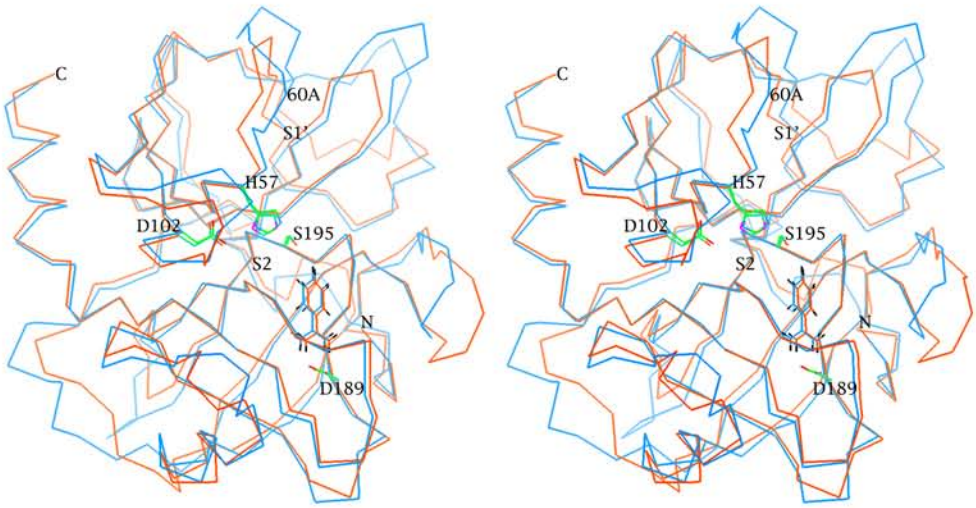
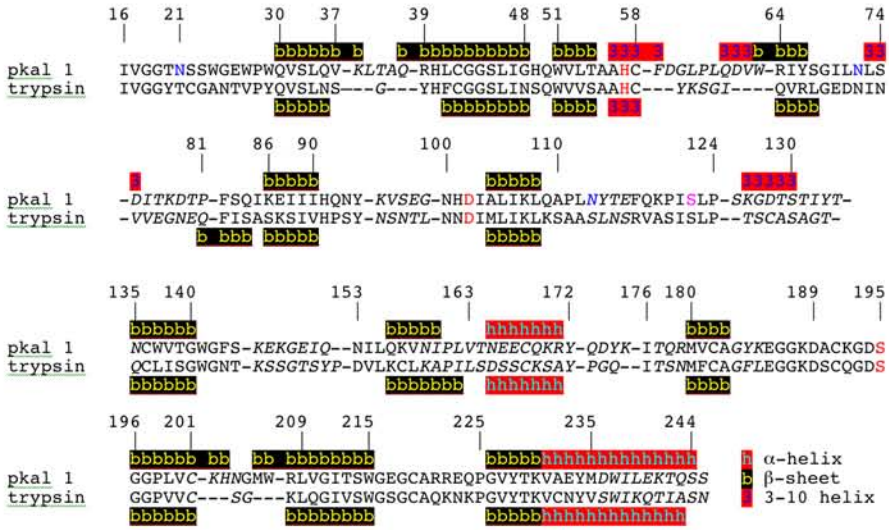


Fig. 3

(a)


C
60A
S1'
H57
D102
S195
S2
N
D189
C
60A
S1'
H57
D102
S195
S2
N
D189

(b)


pkal 1 IVGGTNSSWGEWPWQVSLQV-KLTAQ-RHLCGGSLIGHQWVLTAAHC-FDGLPLQDVW-RIYSGILNLS
trypsin IVGGYTCGANTVPYQVSLNS---G---YHFCGGSLINSQWVVSAAHC---YKSGI----QVRLGEDNIN
pkal 1 -DITKDTP-FSQIKEIIIHQNY-KVSEG-NHDIALIKLQAPLNYTEFQKPISLP-SKGDTSTIYT-
trypsin -VVEGNEQ-FISASKSIVHPSY-NSNTL-NNDIMLIKLKSAASLNSRVASISLP--TSCASAGT-
pkal 1 NCWVTGWGFS-KEKGEIQ--NILQKVNIPLVTNEECQKRY-QDYK-ITQRMVCAGYKEGGKDACKGDS
trypsin QCLISGWGNT-KSSGTSYP-DVLKCLKAPILSDSSCKSAY-PGQ--ITSNMFCAGFLEGGKDSCQGDS
pkal 1 GGPLVC-KHNGMW-RLVGITSWGEGCARREQPGVYTKVAEYMDWILEKTQSS
trypsin GGPVVC---SG---KLQGIVSWGSGCAQKNKPGVYTKVCNYVSWIKQTIASN
α-helix
β-sheet
3-10 helix

(c)

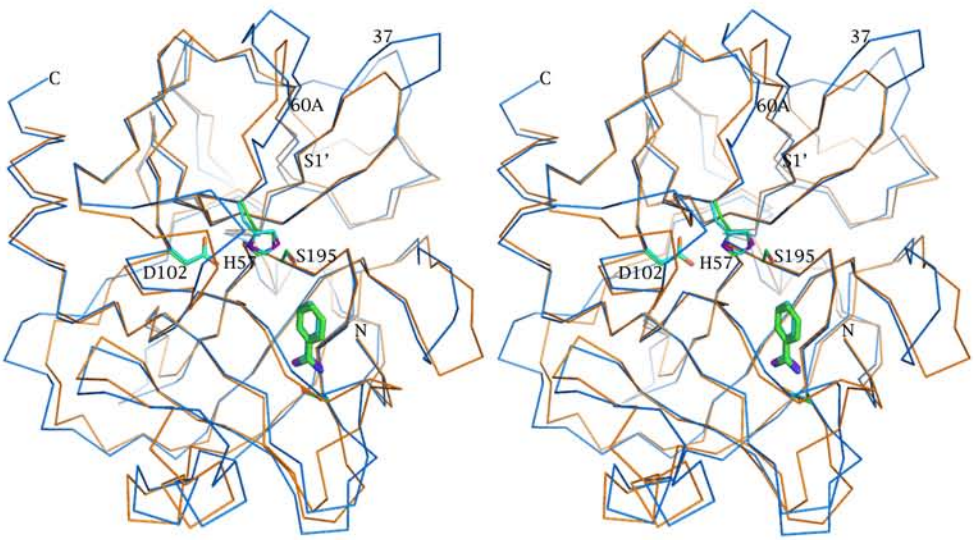
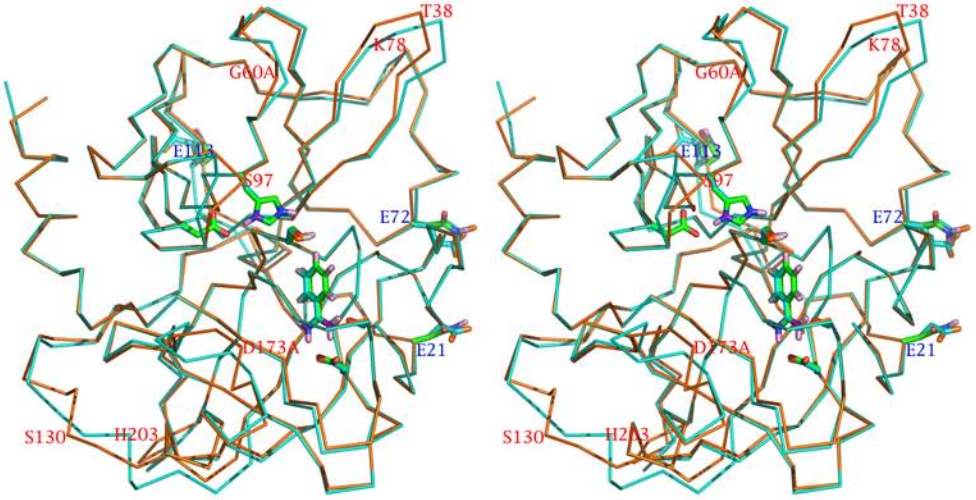

C
37
60A
S1'
D102
H57
S195
N
C
37
60A
S1'
D102
H57
S195
N

Fig 4

(a)



(b) Differences between Endo-H- and mutagenically deglycosylated kallikrein

Residues	Sequence	Packing problems EndoH-	mutagenically-	dif.[a] (Å)	res.[b]
34-38B	QVKLT	yes	no	2.8	38
60-65C	DGLPLQDV	no	yes	1.9	60B
74-82	SDITKDTPFSQ	yes	possible	1.2	78
96-99	VSEG	no	yes	3.4	97
127-135	GDTSTIYTN	yes	yes	4.7	129
173-175	QDYK	yes	yes	5.0	174
203-205	HNG	yes	yes	1.7	204

[a] Maximum difference in Cα coordinates (Å) for this loop region after global superposition of structures.

[b] Residue in loop region where maximum difference occurs.

Fig 5

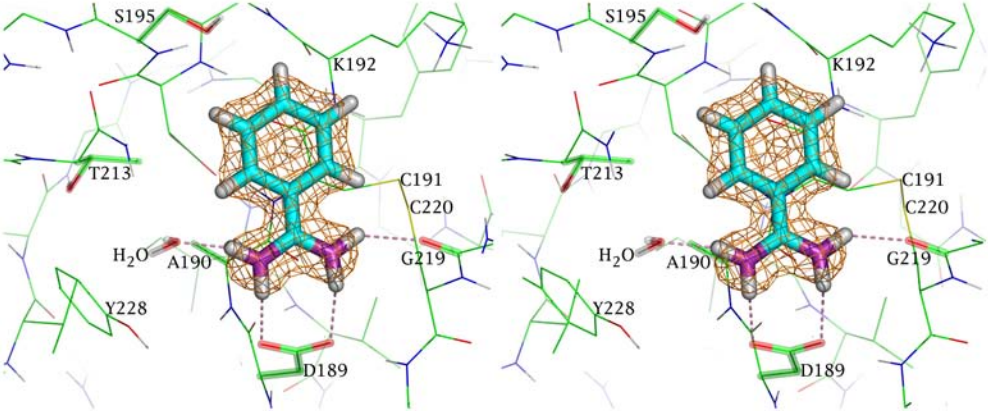

S195
K192
T213
C191
C220
H_2O
A190
G219
Y228
D189
S195
K192
T213
C191
C220
H_2O
A190
G219
Y228
D189

Fig 6

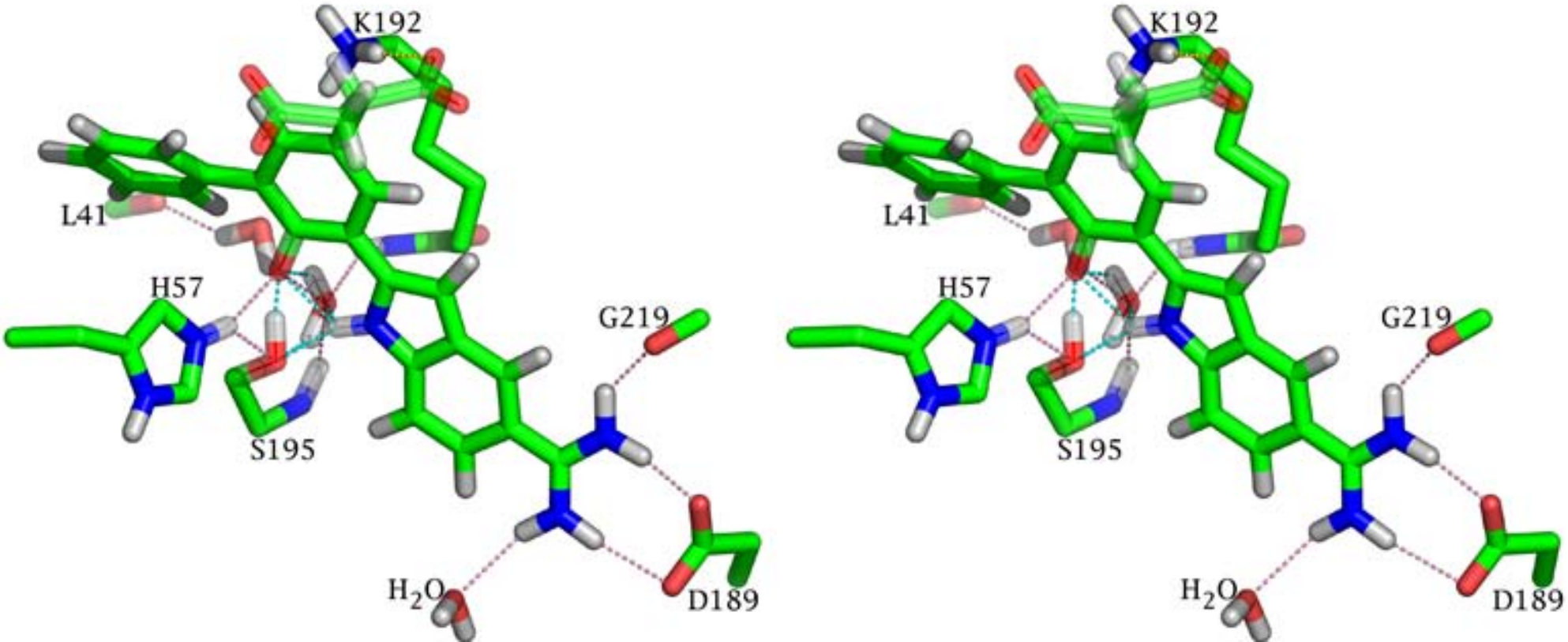

K192
L41
H57
S195
G219
H_2O
D189
K192
L41
H57
S195
G219
H_2O
D189

Fig 7

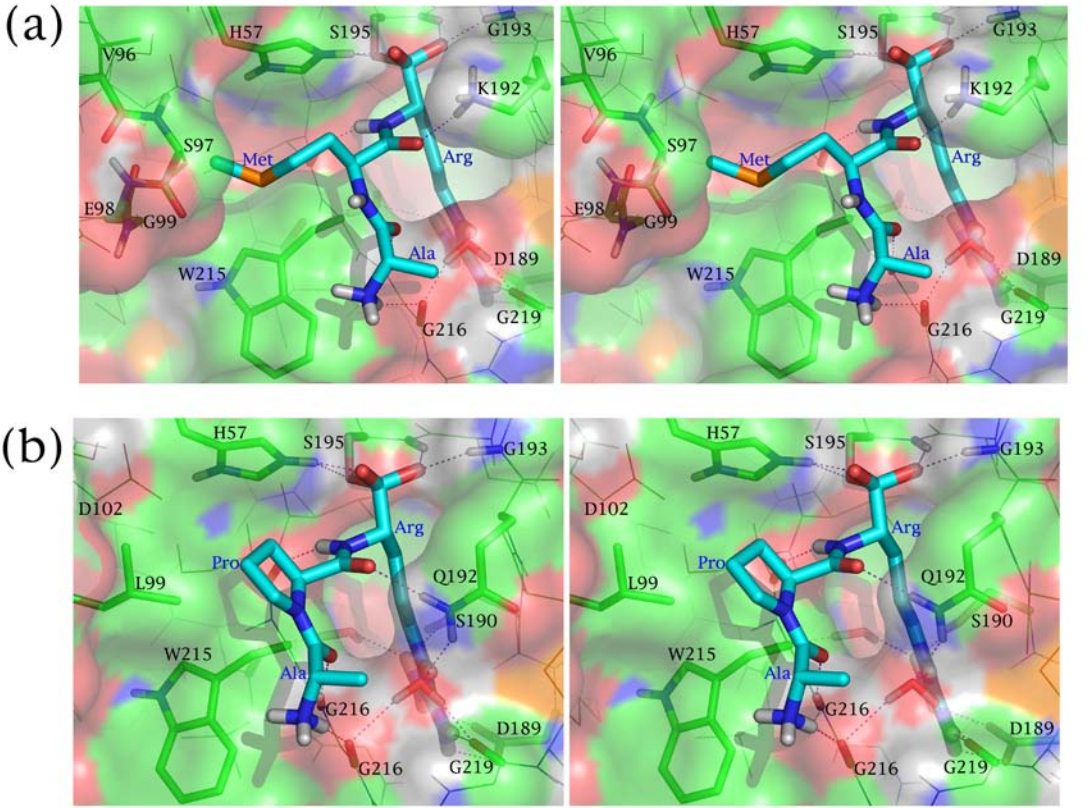
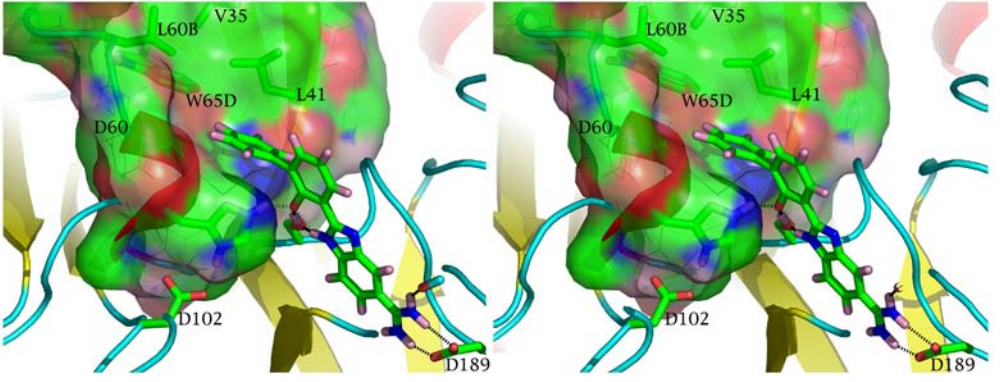

(a)
H57
S195
G193
V96
K192
S97
Met
Arg
E98
G99
Ala
D189
W215
G216
G219
(b)
H57
S195
G193
D102
Arg
Pro
Q192
L99
S190
W215
Ala
G216
D189
G216
G219

Fig 8


L60B
V35
W65D
L41
D60
D102
D189
L60B
V35
W65D
L41
D60
D102
D189

Expression, crystallization and three-dimensional structure of the catalytic domain of human plasma kallikrein

Jie Tang, Christine Luong Yu, Steven R. Williams, Erik Springman, Douglas Jeffery, Paul Sprengeler, Alberto Estevez, Jun Sampang, William Shrader, Jeff Spencer, Wendy Young, Mary McGrath and Bradley A. Katz

J. Biol. Chem. published online September 30, 2005
