*Protein Science* (1997), *6*:1418–1425. Cambridge University Press. Printed in the USA.


# The crystal structure of the mouse glandular kallikrein-13 (prorenin converting enzyme)

DAVID E. TIMM

ICRF Unit, Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX

(RECEIVED November 18, 1996; ACCEPTED April 4, 1997)

**Abstract**

A crystal structure of the serine protease, mouse glandular kallikrein 13 (mGK-13) has been determined at 2.6-Å resolution. This enzyme, isolated from the mouse submandibular gland, is also known as prorenin-converting enzyme and cleaves submandibular gland Ren-2 prorenin to yield active renin. The mGK-13 structure is similar to other members of the mammalian serine protease family, having five conserved disulfide bonds and an active site located in the cleft between two $\beta$-barrel domains. The mGK-13 structure reveals for the first time an ordered kallikrein loop conformation containing a short $3_{10}$ helix. This loop is disordered in the related porcine pancreatic kallikrein and rat submandibular tonin structures. The kallikrein loop is in close spatial proximity to the active site and is also involved in a dimeric arrangement of mGK-13. The catalytic specificity of mGK-13 for Ren-2 prorenin was studied by modeling a prorenin-derived peptide into the active site of mGK-13. This model emphasizes two electronegative substrate specificity pockets on the mGK-13 surface, which could accommodate the dibasic P2 and P1 residues at the site of prorenin cleavage by mGK-13.

**Keywords:** crystallography; glandular kallikrein; protein structure; serine protease

Interest in the mouse glandular kallikrein-13 arose from it's identification in isolations of the high molecular weight (74 kDa) form of epidermal growth factor from mouse submandibular glands (Taylor et al., 1970; Anundi et al., 1982). Precursors to EGF and nerve growth factor are processed by specific glandular kallikreins that are also abundant in these glands and the mature growth factors are stored complexed with these enzymes (Varon et al., 1967; Taylor et al., 1970). The EGF binding protein binds and processes pro-EGF with a 1:1 stoichiometry (Frey et al., 1979), but the identity of the EGF-BP has been controversial. Two distinct polypeptides, EGF-BP type A and type B, were identified initially from the glands of adult male NMRI mice (Anundi et al., 1982; Ronne et al., 1983; Lundgren et al., 1984), and a third EGF-BP type C was later identified from the glands of male Swiss Webster mice (Blaber et al., 1987; Isackson et al., 1987). EGF-BPs A, B, and C have been correlated, respectively, with the mGK-22, mGK-13, and mGK-9 through sequencing studies of BALB/c mouse cDNA (Drinkwater et al., 1987). mGK-9 and mGK-22 share 71% and 76% sequence identity with mGK-13, respectively. To date, only mGK-9 has been shown to bind EGF in vitro and is favored as the authentic EGF-BP (Blaber et al., 1987; Isackson et al., 1987).

The kallikreins are a group within the structurally conserved mammalian serine protease family. Serine protease enzymes are involved in numerous processes, such as digestion, blood coagulation, blood pressure regulation, and protein maturation. The high degree of structural conservation among the members of this family is well documented by numerous structural studies, including a number of high-resolution structures (Cohen et al., 1981; Bode et al., 1983, 1989; Marquart et al., 1983; Blevins & Tulinski, 1985; Tsukada & Blow, 1985; Fujinaga & James, 1987; Meyer et al., 1988; Bartunik et al., 1989). Features of this family include the overall fold of two $\beta$-barrel domains, the presence of four to six conserved disulfide bonds, and the conserved geometry of the catalytic triad residues (His 57, Asp 102, and Ser 195).

The mGK-13 is composed of a 17-kDa and a 10-kDa polypeptide chain that are joined by disulfide bonds at cystine 22–157 and cystine 136–201. This chain arrangement is the result of proteolytic processing of a 261-residue proenzyme (zymogen), a common feature of mammalian serine protease activation, which results in the removal of 24 residues from the N-terminus of mGK-13 and a single internal cleavage between residues Arg 148 and Trp 149 (Blaber et al., 1987). Enzymatically, mGK-13 has been shown to convert submandibular gland Ren-2 prorenin to activated renin by cleavage at a specific Lys-Arg site (Kim et al., 1990, 1991; Nakayama et al., 1990) and has, therefore, been referred to as prorenin-converting enzyme. However, mGK-13 does not convert mouse

Reprint requests to David E. Timm at his present address: Biology Division, Oak Ridge National Laboratory, P.O. Box 2009, M.S. 8077, Oak Ridge, Tennessee 37831; e-mail: timmde@bioax1.bio.ornl.gov.

*Abbreviations:* EGF, epidermal growth factor; EGF-BP, epidermal growth factor binding protein; HMWEGF, high molecular weight EGF; mGK, mouse glandular kallikrein; NAG, N-acetyl glucosamine; PKA, porcine pancreatic kallikrein; RMSD, RMS deviation.

kidney type Ren-1 prorenin or human prorenin (Kim et al., 1990; Nakayama et al., 1990). The mGK-13 structure provides a basis for the catalytic specificity of mGK-13 for mouse Ren-2 prorenin. A preliminary crystallographic analysis of mGK-13 has also been reported in abstract form (Blaber, 1994).

The kallikrein loop is a characteristic insertion occurring between the sixth and seventh antiparallel $\beta$-sheets in the kallikrein group of serine protease enzymes. The mGK-13, mGK-9, the prostate-specific antigen, and tonin contain 11-residue insertions relative to chymotrypsin, whereas mGK-22 and PKA contain contain 9- and 4-residue insertions, respectively. This region of the PKA and tonin structures is disordered, with residues Asn 95A–Lys 97 missing from the PKA structure (Bode et al., 1983; Chen & Bode, 1983) and residues Ile 95C–Gln 95J missing from the tonin structure (Fujinaga & James, 1987). The mGK-13 structure presented here describes an ordered kallikrein loop conformation that is near the active site and involved in a dimeric arrangement of mGK-13.

## Results

Purified mGK-13 crystallized as tetragonal bipyramids in the space group $P4_32_12$ using a precipitant solution containing PEG 8000 and $Li_2SO_4$ at pH 6.5. The mGK-13 crystal structure was determined at 2.6 Å resolution using the coordinates of the porcine pancreatic kallikrein (Bode et al., 1983; Chen & Bode, 1983) as a search model for molecular replacement. The mGK-13 model, containing 236 of the 237 residues present in each of two protein molecules, a total of 119 water molecules, and four NAG residues, has been refined to an $R$-factor of 19.7%, with a free $R$-factor of 27.1%. The crystallographic data is summarized in Table 1. A representative section of electron density calculated in the vicinity of the active site and the ordered kallikrein loop (residues 95A–95K) is given in Figure 1.

The mGK-13 has the conserved mammalian serine protease fold consisting of two $\beta$-barrel domains with an active site located in the cleft between the two domains (Fig. 2A). Two $\alpha$-helices occur in the structure between residues 165 and 171 (residue numbering is relative to chymotrypsin) and between residues 231 and 244. In addition to the regular $\beta$-strands and $\alpha$-helices, 33 $\beta$-turns were identified in the mGK-13 structure (Table 2). Thus, close to half of the molecule exists in ordered turn and loop conformations, which are particularly important in the positioning of the catalytic triad (His 57, Asp 102, and Ser 195) and in forming the enzyme's substrate binding site.

**Table 1.** *Crystallographic data and refinement statistics*

| | |
|---|---|
| Space group | $P4_32_12$ |
| Unit cell | $a = b = 73.75$ Å; $c = 192.71$ Å |
| Maximum resolution | 2.55 Å |
| Unique reflections | 17,034 |
| Completeness | 95.3% |
| Multiplicity | 2.5 |
| $R_{merge}$ | 0.065 |
| $R$ ($R_{free}$) | 19.7% (27.1%) |
| RMSD bonds | 0.008 Å |
| RMSD angles | 1.740 deg |

[a] $R_{merge} = \Sigma\langle I\rangle - Ii/\Sigma\langle I\rangle$; crystallographic $R$-factor and free $R$-factor calculated using 15,150 and 785 reflections, respectively, to 2.60 Å.

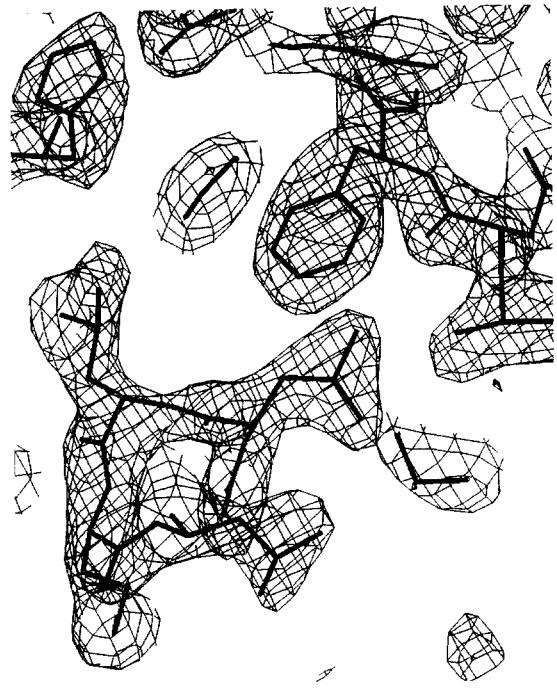

**Fig. 1.** Electron density. A section of the refined $2F_o - F_c$ electron density map is presented in the vicinity of the kallikrein loop residues Leu 95C and Leu 95D at the bottom and His 57 at the upper left. The section is contoured at a level of 1.3 sigma.

The kallikrein loop of mGK-13 is in an ordered conformation that is clearly defined by the electron density (Fig. 1). The kallikrein loop overhangs the active site cleft of mGK-13, with two consecutive type IV $\beta$-turns forming a $3_{10}$ helix between residues 95 and 95E (Fig. 2; Table 2). The loop conformation is stabilized by intramolecular hydrophobic packing of Met 95A, Leu 95C, Leu 95D, Leu 95F, Ile 95I, and Pro 95J in the vicinity of His 57, Tyr 59, Pro 90, Phe 94, Phe 99, and Met 104. The loop is in close spatial proximity to the active site cleft, with the CD1 atom of Leu 95D located at a distance of 4.1 Å from the CB atom of His 57. Two ordered NAG residues N-linked to Asn 95 are also apparent in the vicinity of the kallikrein loop (Fig. 2). These residues make intramolecular contacts with Ser 95B, Leu 95C, Leu 95F, and Pro 95J and intermolecular contacts with Tyr 59 and Asp 61.

The mGK-13 crystallized with two molecules in the asymmetric unit related by a noncrystallographic dyad axis (Fig. 2B,C). The two molecules have an RMSD of 0.42 Å upon superposition of all alpha carbon atoms. The intersubunit contacts are made by residues Tyr 35, Gln 36, His 57, Cys 58, Val 60, and Asp 61 and the kallikrein loop residues Met 95A, Leu 95D, Met 95E, Leu 95F, Gln 95G, Thr 95H, Pro 95J, and Pro 95K. Contacts between the twofold related molecules result in about 840 Å$^2$ of buried surface area per subunit, which is close to twice the surface area buried in alternative dimeric arrangements of the asymmetric unit. This value is less than that generally associated with proteins having a significant dissociation constant in solution (Janin & Chothia, 1990). A void space, having a volume of 1,825 Å$^3$, occurs between the two mGK-13 molecules (Fig. 2B).

The enzymatic specificity of mGK-13 was studied by modeling a hexameric peptide substrate (Thr-Lys-Arg*Ser-Ser-Leu), derived

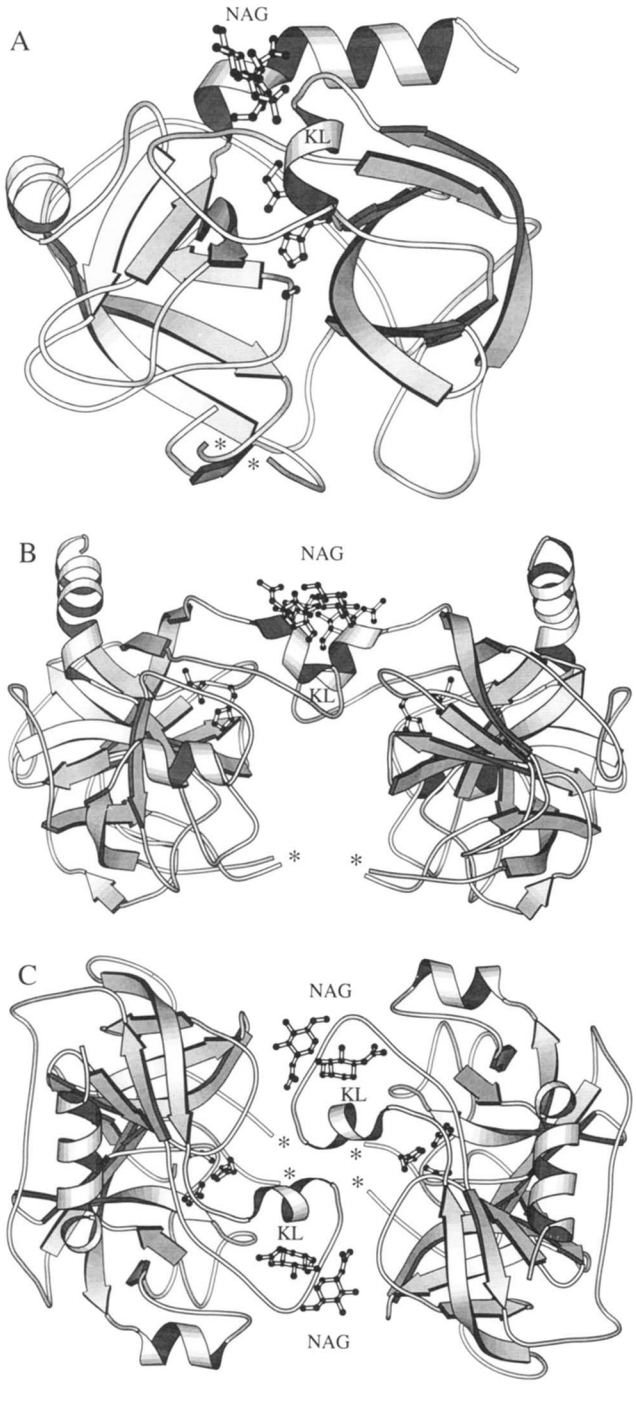


**Fig. 2.** The mGK-13 structure. **A:** Tertiary structure of mGK-13 is viewed looking onto the active site cleft. **B:** Dimeric arrangement of mGK-13 in the crystallographic asymmetric unit is shown viewed perpendicular to the noncrystallographic dyad axis. **C:** Same as B, shown parallel to the noncrystallographic dyad axis. The kallikrein loop (KL) and proteolytic break points (*) are also indicated. The ribbon diagrams were generated using RASMOL (R. Sayle) and MOLSCRIPT (Kraulis, 1991).

from the site of mouse Ren-2 prorenin processing, into the mGK-13 active site (see Materials and methods). The mGK-13 substrate binding site appears as a lengthy, branching groove on the surface of the molecule (Fig. 3A). The main features of the modeled complex are the interactions of the basic P1 Arg and P2 Lys substrate residues with the electronegative S1 and S2 binding pockets (Fig. 3A,B). The P1 Arg side chain fits into a deep primary specificity pocket formed by the main-chain atoms of Tyr 215 and Cys 191 and side-chain atoms from Thr 213, Tyr 228, and Ala 226. The P1 Arg side-chain nitrogen atoms are positioned to allow hydrogen bonding to the Asp 189 side-chain oxygens and to the side-chain and main-chain oxygens of Thr 190. The S2 pocket is comprised of His 57, Phe 99, and Tyr 215, with the P2 Lys side chain positioned such that it has the potential to hydrogen bond with the side chain of the catalytic Asp 102 and both side-chain and main-chain oxygens of the highly conserved Ser 214.

The tertiary structure of mGK-13 is most similar to PKA, despite having a slightly higher sequence identity with tonin (Fig. 4; Table 3). The higher RMSDs between mGK-13 and tonin are primarily due to the distortion of the tonin active-site and primary specificity binding loop (residues Gly 215–Pro 219) by a $Zn^{2+}$ ion coordinated by the active site His 57 imidazole group (Fujinaga & James, 1987). The major regions of variation between mGK-13, PKA, and tonin are confined to surface loops, with the kallikrein loop and the loops between residues 34–40, 57–63, 74–78, 146–149, and 169–178 showing the largest structural differences (Fig. 4). Superposition of the catalytic triad alpha carbon atoms reveals an RMSD of 0.31 Å between mGK-13 and PKA, compared to 0.37 Å between tonin and PKA. These values are probably less than the overall error in the mGK-13 coordinates refined at 2.6-Å resolution and having a 0.42-Å RMSD between the two mGK-13 molecules in the asymmetric unit. However, distortion of the serine protease active site may occur in precipitant solutions containing sulfate below pH 6.8 (Bartunik et al., 1989).

## Discussion

As a member of the kallikrein group of serine protease enzymes, the mGK-13 structure shares the structural features common to the family. The most striking feature of the mGK-13 structure is the ordered kallikrein loop, which is missing from the PKA and tonin structures (Bode et al., 1983; Chen & Bode, 1983; Fujinaga & James, 1987). Proteolysis may contribute to the disorder of the kallikrein loop in the PKA structure (Bode et al., 1983). However, because most of the contacts between the two mGK-13 molecules in the asymmetric unit are made by residues in the kallikrein loop, the observed loop conformation could be induced by crystallization, with the loops in both molecules assuming nearly identical, complementary conformations. Contacts made by the NAG residues linked to Asn 95 may also contribute to the observed loop conformation. The PKA also has carbohydrate linked to Asn 95 (Bode et al., 1983), but carbohydrate was not apparent in the electron density. The consensus Asn 95 glycosylation sequence is also present in mGK-9 and mGK-22. However, the carbohydrate content of mGK-9 and mGK-22 have not been reported.

The proximity of the kallikrein loop to the active site (Figs. 2, 3) suggests that the loop could have a role in substrate binding and catalysis. Based on the peptide modeling studies, residues Leu 95D and Ile 95I are part of the proposed S′3 and S3 binding sites (Fig. 3). The kallikrein loop is defined slightly better in the PKA complex with pancreatic trypsin inhibitor (Chen & Bode, 1983), and PKA residues 89–95, immediately adjacent to the kallikrein loop, contact the inhibitor. The kallikrein loop's involvement in binding and catalysis is also suggested by a prostate specific antigen modeling study (Villoutreix et al., 1994), where it is noted

**Table 2.** *β-Turn types identified in mGK-13*[a]

| Residues | Turn type | Residues | Turn type | Residues | Turn type |
|---|---|---|---|---|---|
| 16–19 | II | 77–80 | I | 116–119 | VIII |
| 23–26 | II | 91–94 | I | 131–134 | II |
| 25–28 | IV | 95–95C | IV | 177–180 | I |
| 27–30 | I | 95A–95D | I | 185–186 | I |
| 34–38 | IV | 95B–95E | IV | 191–194 | II |
| 35–39 | II′ | 95C–95F | I | 194–197 | II |
| 48–51 | I | 95F–95I | VIII | 200–207 | IV |
| 55–58 | I | 95J–97 | II | 201–208 | I′ |
| 56–59 | IV | 99–102 | IV | 217–220 | VIII |
| 70–73 | IV | 108–111 | VIII | 219–222 | II |
| 72–75 | I | 115–118 | I | 222–224 | II |

[a]Identified using PROMOTIF (Hutchinson & Thornton, 1996).

that molecular dynamics simulations led to a covering of the active site by a "closed" kallikrein loop conformation having the lowest potential energy, whereas a defined "open" loop conformation had the lowest solvation energy. Homology models of this clinically important serum marker for prostate carcinoma have been reported by a number of groups (Vihinen, 1994; Villoutreix et al., 1994, 1996; Bridon & Dowell, 1995), with modeling of the 11-residue kallikrein loop noted to be a major problem. The mGK-13 kallikrein loop structure presented here may represent a similar "open" conformation with an accessible active site.

The electronegative character of the S1 and S2 binding sites (Fig. 3A) on the mGK-13 surface are consistent with its preference for dibasic P1 and P2 substrate residues (Kim et al., 1990; Nakayama et al., 1990). The S1 Asp 189 is conserved in PKA and trypsin, which cleave substrates having basic P1 residues. The negative potential in the S2 pocket is primarily due to the catalytic Asp 102 side-chain oxygens and to both side-chain and main-chain oxygens of the highly conserved Ser 214. Therefore, the electronegative character of the S2 pocket is not unique to mGK-13 and is apparently less important than S1 in determining enzymatic specificity, because mGK-13 will cleave monobasic sites with reduced efficiency compared to the dibasic Ren-2 prorenin site (Nakayama et al., 1990). The length of the cleft on the mGK-13 surface (Fig. 3A) suggests that a more extended area of contact may be involved in the binding of prorenin to mGK-13. An extended binding area has also been proposed for tonin, which apparently derives some specificity from substrate residues as distant as P7 and P8 (Fujinaga & James, 1987).

The conditions of crystallization (see Materials and methods) make the physiological relevance of the mGK-13 dimer unclear. Although mGK-13 is monomeric in solution at a concentration of 1 mg/mL, the yield of purified mGK-13 indicates that a concentration in excess of 10 mg/mL occurs in the mouse submandibular glands (see Materials and methods). The local concentration of mGK-13 in secretory granules is likely to be higher still. Based on the identification of mGK-13 in isolations of the high molecular weight form of EGF (Anundi et al., 1982; Ronne et al., 1983; Lundgren et al., 1984), it is tempting to speculate that the mGK-13 dimer might be related to the arrangement of EGF-BP within the HMWEGF. Residues shown to be important in the binding of mGK-9 to EGF (Blaber et al., 1993a, 1993b), including the substrate binding pockets and side chains of residues 38–41, are located on the perimeter of the void. However, the binding and processing of pro-EGF by mGK-13 remain to be determined.

## Materials and methods

### *Purification and crystallization*

The mGK-13 was purified from male mouse submandibular glands (Pelfreeze) with slight modification of the procedure for isolating high molecular weight EGF (Taylor et al., 1974). Briefly, 20 g of submandibular glands were homogenized in phosphate-buffered saline containing 10 $\mu$M $ZnCl_2$. Following centrifugation in an SS34 rotor at 13,000 rpm, the supernatant was applied to a S-200 Sephacryl column (Pharmacia) equilibrated in 50 mM Tris-HCl, 10 $\mu$M $ZnCl_2$, pH 7.4. Fractions were monitored by absorbance readings at 280 nm and were pooled from the beginning to end of the red hemoglobin peak. The sample was then concentrated by ultrafiltration (Amicon) and dialyzed against 10 mM sodium acetate, pH 5.9. The material was further purified by passing first over a CM52 (Whatman) column, then a DE52 column (Whatman) equilibrated with the sodium acetate buffer. As with high molecular weight EGF (Taylor et al., 1974), the mGK-13 does not bind to either of the ion exchange columns under these conditions. The procedure yielded 180 mg of mGK-13.

The purified mGK-13 ran as a single band on SDS-PAGE with an apparent molecular weight of 27 kDa under oxidizing conditions, and as bands of 27, 17, and 10 kDa of approximately equal staining intensity under reducing conditions. The material was further characterized by 14 cycles of N-terminal amino acid sequencing (data not shown), which revealed two sequences consistent with the mature N-terminus of mGK-13 and the single internal cleavage site between Arg 148 and Trp 149 (Blaber et al., 1987). The ratio of peak areas for the two sequences during each sequencing cycle was consistent with the staining intensity of the three bands revealed by SDS-PAGE under reducing conditions. There was no evidence for contaminating mGK-22 in the sample, which differs from mGK-13 at four of the first eight mature N-terminal residues (Blaber et al., 1987; Drinkwater et al., 1987). This conclusion was further supported by evaluation of the electron density (Fig. 1), particularly in the kallikrein loop, where mGK-22 contains a two-residue deletion relative to mGK-13 (Drinkwater et al.,

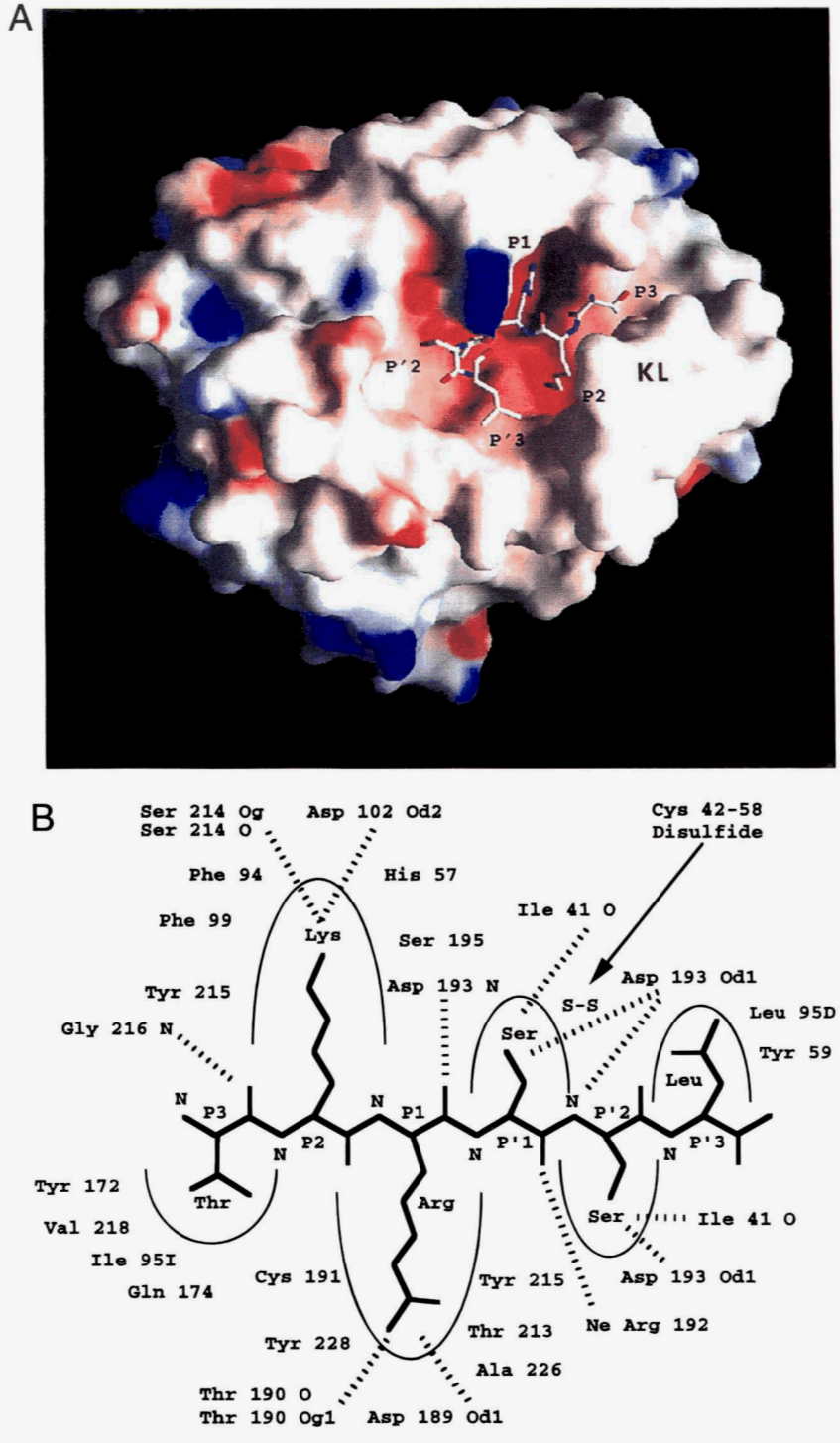


**Fig. 3.** Peptide modeled in the mGK-13 active site. A modeled hexameric peptide from Ren-2 prorenin is represented in (**A**) a space-filling electrostatic surface diagram calculated using GRASP (Nicholls et al., 1991) and (**B**) a schematic form. Positive and negative electrostatic surface potentials are displayed in blue and red, respectively. The mGK-13 substrate binding pocket, S3, could accommodate polar or hydrophobic residues; S2 accommodates the basic P2 lysine interacting with Asp 102 and Ser 214; S1 accommodates the basic P1 arginine interacting with Asp 189; P′1 and P′2 could accommodate small polar or hydrophobic residues; P′3 accommodates hydrophobic residues. The kallikrein loop is labeled KL.

1987). Therefore, approximately 30% of the isolated mGK-13 was not cleaved at the internal proteolytic cleavage site. A 1 mg/mL mGK-13 solution was found to be monomeric and monodisperse following characterization on a calibrated Superose 12 column (Pharmacia) and using a DynaPro 801 to measure the diffusion coefficient of mGK-13 by light scattering (data not shown).

The mGK-13 was concentrated to 180 mg/mL by ultrafiltration (Centricon) and stored at −20 °C. Equal volumes of precipitant

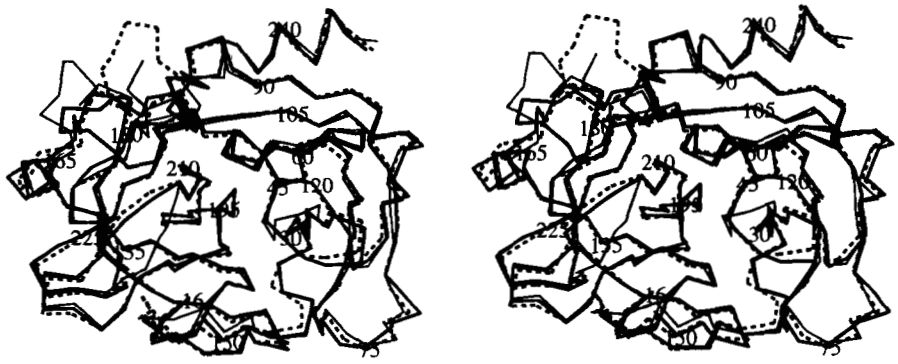

**Fig. 4.** An alpha carbon atom trace is shown for the superposed coordinates of mGK-13 (thick line), porcine pancreatic kallikrein (dashed line), and tonin (thin line). The mGK-13 chain has been numbered in steps of 15. The coordinates were superposed using O (Jones et al., 1991) and displayed using RASMOL (R. Sayle) and MOLSCRIPT (Kraulis, 1991).

solution were combined with 57 mg/mL mGK-13 in final drop volumes of 5–20 μL. The material crystallized at room temperature as tetragonal bipyramids up to 1 mm in length using both the hanging drop and sitting drop vapor diffusion methods and using a precipitant solution of 10–15% PEG 8000, 200 mM $Li_2SO_4$, 100 mM sodium cacodylate, pH 6.5. Under these conditions, crystals appeared within two weeks and reached full size in 4–6 weeks. The mGK-13 crystals revealed roughly the same SDS-PAGE band pattern as the starting material under reducing conditions.

### *Data collection and reduction*

The crystallographic data are summarized in Table 1. The X-ray diffraction data used to determine the mGK-13 structure were collected from a single crystal on a 30-cm MAR image plate at Daresbury Laboratory Station 9.5 at 4 °C using X-rays of wavelength 0.9 Å. The diffraction data was autoindexed and integrated using the MOSFLM program suite (Leslie, 1990). Programs from the CCP4 program suite (CCP4, 1979) were used to scale, merge, and reduce the data. A primitive tetragonal space group was consistent with the autoindexing results and 422 symmetry was apparent in reciprocal lattice sections. Systematic absences in the data along the reciprocal $c^*$ axis were indicative of a $4_1$ or $4_3$ screw axis.

### *Molecular replacement*

The mGK-13 structure was determined by molecular replacement using the porcine pancreatic kallikrein coordinates, PDB2KAI.PDB, as a search model (Chen & Bode, 1983). The program AMORE (CCP4, 1979; Navaza, 1992) was used to solve the Patterson rotation function search. The search model was centered at the origin of an orthorhombic P1 cell having axis lengths of $a = b = c = 75$ Å and $\alpha = \beta = \gamma = 90°$. The search radius was optimized as 20 Å and normalized structure factors ($E$'s) were used with optimized resolution limits of 12.0–3.75 Å. The top two rotation function solutions improved significantly compared to the background peak height values as the space group symmetry was increased from P4 to P422. The top two rotation function solutions calculated in P422 were then used to calculate translation functions for all of the possible primitive tetragonal space groups using the program TFFC (CCP4, 1979). The resultant molecular packing was evaluated graphically using the program O (Jones et al., 1991). The translation function calculated using the space group $P4_32_12$ resulted in the best correlation coefficients and $R$-factors and yielded a good packing geometry with two mGK-13 molecules in the asymmetric unit. The two molecules resulting from rotation and translation searches were then subjected to rigid-body refinement using the program XPLOR (Brünger, 1992), which yielded an initial $R$-factor of 45% using 8.0–3.0 Å data.

### *Model building/refinement*

A test set of 5% of the structure factors was flagged for monitoring the $R$-free (Brünger, 1992) during the course of refinement. The program O (Jones et al., 1991) was used for model building and the program XPLOR (Brünger, 1992) was used for refinement. The initial $2F_o - F_c$ and $F_o - F_c$ electron density maps were calculated with structure factors in the 8.0–3.0-Å range using the PKA coor-

**Table 3.** *Sequence identities and structural RMSDs*[a]

| | mGK-13 | Kallikrein | Tonin | Chymotrypsin | Trypsin |
|---|---|---|---|---|---|
| mGK-13 | — | 0.52 Å | 0.64 Å | 0.96 Å | 0.86 Å |
| Kallikrein | 55.2% | — | 0.54 Å | 0.98 Å | 0.924 Å |
| Tonin | 62.0% | 51.9% | — | 0.98 Å | 0.93 Å |
| Chymotrypsin | 31.4% | 32.1% | 31.0% | — | 0.86 Å |
| Trypsin | 38.0% | 40.9% | 38.8% | 44.2% | — |

[a]Upper right: RMSDs calculated by least-squares superposition of the indicated structures using MNYFIT (Sutcliffe et al., 1987). Lower left: amino acid sequence identities.

dinates resulting from rigid-body refinement. These coordinates differed most significantly from the initial electron density maps in the kallikrein loop region, where mGK-13 contains a seven-residue insertion relative to PKA, and near residue 148, where PKA contains a two-residue insertion relative to mGK-13. The final mGK-13 model has been refined as two chains with a break at the Arg 148 site of proteolysis. The Arg 148 has been omitted from the mGK-13 model, because it is disordered and missing from the electron density maps. Refinement using a single contiguous polypeptide model to account for the 30% intact material (see above) also resulted in a disordered Arg 148 in both mGK-13 molecules. The kallikrein loop residues were omitted from the first rebuilt model and were added sequentially at both the N- and C-terminal ends, as indicated by the $2F_o - F_c$ and $F_o - F_c$ maps, over three rounds of simulated annealing at 3,000 K, applying strict noncrystallographic symmetry constraints and overall temperature-factor refinement. Subsequent rounds of refinement used all data to 2.6 Å, applying a 2-sigma cutoff, a bulk solvent mask, noncrystallographic symmetry restraints, and group temperature factors refined for main-chain and side-chain atoms. Noncrystallographic symmetry restraints were removed in the final refinement cycle. Two N-linked NAG residues were also included for each subunit, where indicated by the electron density maps. Water molecules were added to peaks in the $F_o - F_c$ maps having heights greater than or equal to 3.0 RMS, and were included in the model if they were positioned to make potential hydrogen bonds to the protein and their refined temperature factors were less than 60 Å$^2$. The current mGK-13 model has one residue in a disallowed region of the Ramachandran plot, with 82% of the residues occurring in the most favored regions of the plot (data not shown). Other refinement statistics for the mGK-13 model are given in Table 1.

The prorenin hexapeptide model was constructed beginning with the coordinates for the complex between PKA and pancreatic trypsin inhibitor (Chen & Bode, 1983). The coordinates of the pancreatic trypsin inhibitor were first positioned by applying a matrix derived from the superposition of the mGK-13 and PKA coordinates. The side chains of inhibitor residues I12–I19 were then altered to the appropriate prorenin sequence using O (Jones et al., 1991). The P3–P′3 side chains were then positioned starting with preferred rotamer geometries and adjusted using minor rotations about their chi 1 bond angles to minimize steric collisions with groups from the mGK-13. The peptide was then subjected to energy minimization using the program QUANTA (Molecular Simulations Company) and the CHARRM force field with the mGK-13 atoms fixed to their crystallographically defined positions. The carbonyl carbon of the P1 Arg residue is positioned less than 2.8 Å from the nucleophilic Og of Ser 195 following energy minimization. Minimization of the protein structure has not been evaluated in these modeling studies. Based on the positional deviation between the initial and final minimized structures and the number of stabilizing contacts made during minimization, the P3–P′2 residues are judged to be modeled more accurately than the P′3 residue.

## Acknowledgments

I thank Ben Bax (Birkbeck College, London) and Tom Blundell (ICRF Unit, Birkbeck College, London) for technical advice and helpful comments on the manuscript, Nick Totti (Ludwig Institute for Cancer Research, London) for N-terminal amino acid sequencing of the mGK-13 sample, and the Imperial Cancer Research Fund, London for providing funding support. Coordinates will be deposited in the Protein Data Bank.

## References

Anundi H, Ronne H, Peterson PA, Rask L. 1982. Partial amino-acid sequence of the epidermal growth-factor-binding protein. *Eur J Biochem 129*:365–371.

Bartunik HD, Summers LJ, Bartsch HH. 1989. Crystal structure of bovine β-trypsin at 1.5 Å resolution in a crystal form with low molecular packing density. *J Mol Biol 210*:813–828.

Blaber M. 1994. Crystallographic analysis of mouse prorenin converting enzyme. *J Cell Biochem Suppl 18D*:151.

Blaber M, Isackson PJ, Bradshaw RA. 1987. A complete cDNA sequence for the major epidermal growth factor binding protein in the male mouse submandibular gland. *Biochemistry 26*:6742–6749.

Blaber M, Isackson PJ, Burnier JP, Marsters JC Jr, Bradshaw RA. 1993a. Synthetic chimeras of mouse growth factor associated glandular kallikreins. I. Kinetic properties. *Protein Sci 2*:1210–1219.

Blaber M, Isackson PJ, Holden HM, Bradshaw RA. 1993b. Synthetic chimeras of mouse growth factor associated glandular kallikreins. II. Growth factor binding properties. *Protein Sci 2*:1220–1228.

Blevins RA, Tulinski A. 1985. The refinement and the structure of the dimer of α-chymotrypsin at 1.67 Å resolution. *J Biol Chem 260*:4264–4275.

Bode W, Chen Z, Bartels K, Kutzbach C, Schmidt-Kastner G, Bartunik H. 1983. Refined 2 Å X-ray crystal structure of porcine pancreatic kallikrein A, a specific trypsin-like serine protease. *J Mol Biol 164*:237–282.

Bode W, Mayr I, Baumann U, Huber R, Stone SR, Hofsteenge J. 1989. The refined 1.9 Å crystal structure of human α-thrombin: Interaction with D-Phe-Pro-Arg chloromethylketone and significance of the Tyr-Pro-Pro-Trp insertion segment. *EMBO J 8*:3467–3475.

Bridon DP, Dowell BL. 1995. Structural comparison of prostate-specific antigen and human glandular kallikrein using molecular modeling. *Urology 45*:801–806

Brünger AT. 1992. *XPLOR version 3.1 manual*. New Haven, Connecticut: Yale University Press.

CCP4. 1979. Collaborative Computing Project, No 4: A suite of programs for protein crystallography. Warrington, England: SERC Daresbury Laboratory.

Chen Z, Bode W. 1983. Refined 2.5 Å X-ray crystal structure of the complex formed by porcine kallikrein A and the bovine pancreatic trypsin inhibitor. *J Mol Biol 164*:283–311.

Cohen GH, Silverton EW, Davies DR. 1981. Refined crystal structure of γ-chymotrypsin at 1.9 Å resolution. *J Mol Biol 148*:449–479.

Drinkwater CC, Evans BA, Richards RI. 1987. Mouse glandular kallikrein genes: Identification and characterization of the genes encoding the epidermal growth factor binding proteins. *Biochemistry 26*:6750–6756.

Frey P, Forand R, Maliag T, Shooter EM. 1979. The biosynthetic precursor of epidermal growth factor and the mechanism of its processing. *Proc Natl Acad Sci USA 76*:6294–6298.

Fujinaga M, James MNG. 1987. Rat submaxillary gland serine protease tonin: Structure and solution refinement at 1.8 Å resolution. *J Mol Biol 195*:373–396.

Hutchinson GE, Thornton JM. 1996. PROMOTIF—A program to identify and analyze structural motifs in proteins. *Protein Sci 5*:212–220.

Isackson PJ, Silverman RE, Blaber M, Server AC, Nichols RA, Shooter EM, Bradshaw RA. 1987. Epidermal growth factor binding protein: Identification of a different protein. *Biochemistry 26*:2082–2085.

Janin J, Chothia C. 1990. The structure of protein–protein recognition sites. *J Biol Chem 265*:16027–16030.

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A 47*:110–119.

Kim WS, Hatsuzawa K, Ishizuka Y, Hashiba K, Murakami K, Nakayama K. 1990. A processing enzyme for prorenin in mouse submandibular gland. *J Biol Chem 265*:5930–5933.

Kim WS, Nakayama K, Nakagawa T, Kawamur Y, Haraguchi K, Murakami K. 1991. Mouse submandibular gland prorenin-converting enzyme is a member of glandular kallikrein family. *J Biol Chem 266*:19283–19287.

Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.

Leslie AGW. 1990. *Crystallographic computing*. Oxford: Oxford University Press.

Lundgren S, Ronne H, Rask L, Peterson PA. 1984. Sequence of an epidermal growth factor-binding protein. *J Biol Chem 259*:7780–7784.

Marquart M, Walter J, Deisenhofer J, Bode W, Huber R. 1983. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr B 39*:480–490.

Meyer E, Cole G, Radhakrishnan R, Epp O. 1988. Structure of native procine pancreatic elastase at 1.65 Å resolution. *Acta Crystallogr B 44*:26–38.

Nakayama K, Kim WS, Nakagawa T, Nagahama M, Murakami H. 1990. Substrate specificity of prorenin converting enzyme of mouse submandibular gland. *J Biol Chem 26*:21027–21031.

Navaza J. 1992. AMORE: A new package for molecular replacement 1-87-90. Daresbury, UK: SERC.

Nicholls A, Sharp KA, Honig B. 1991. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct Funct Genet 11*:281–296.

Ronne H, Lundgren S, Severinsonn L, Rask L, Peterson PA. 1983. Growth factor-binding proteases in murine submaxillary gland. *EMBO J 2*:1561–1564.

Sutcliffe JJ, Haneef I, Carney D, Blundell TL. 1987. Knowledge-based modelling of homologous proteins. Part 1: Three dimensional frame work derived from simultaneous superposition of multiple structures. *Protein Eng 1*:377–384.

Taylor JM, Cohen S, Mitchell WM. 1970. Epidermal growth factors: High and low molecular weight forms. *Proc Natl Acad Sci USA 67*:164–171.

Taylor JM, Mitchell WM, Cohen S. 1974. Characterization of the high molecular weight form of epidermal growth factor. *J Biol Chem 249*:3198–3203.

Tsukada H, Blow DM. 1985. Structure of $\alpha$-chymotrypsin refined at 1.68 Å resolution. *J Mol Biol 184*:703–711.

Varon S, Nomura J, Shooter EM. 1967. Subunit structure of a high-molecular-weight form of the nerve growth factor from mouse submaxillary gland. *Proc Natl Acad Sci USA 57*:1782–1789.

Vihinen M. 1994. Modeling of prostate specific antigen and human glandular kallikrein structures. *Biochem Biophys Res Commun 204*:1251–1256.

Villoutreix BO, Getzoff ED, Griffin JH. 1994. A structural model for the prostate disease marker, human prostate-specific antigen. *Protein Sci 3*:2033–2044.

Villoutreix BO, Lilja J, Pettersson K, Lovgren T, Teleman O. 1996. Structural investigation of the alpha-1-antichymotrypsin:prostate specific antigen complex by comparative model building. *Protein Sci 5*:836–851.