*Acta Cryst.* (1997). D**53**, 169–178

# Structure of Human Plasminogen Kringle 4 at 1.68 Å and 277 K. A Possible Structural Role of Disordered Residues

BOGUSLAW STEC,[a] AKIHITO YAMANO,[a]† MARC WHITLOW[b] AND MARTHA M. TEETER[a]

[a]*Department of Chemistry, Boston College, Chestnut Hill, MA 02167, USA, and* [b]*Berlex Biosciences, 15049 San Pablo Avenue, Richmond, CA 94804-0099, USA. E-mail: stec@bcchem.bc.edu*

(*Received 15 April 1996; accepted 2 October 1996*)

## Abstract

Despite considerable effort to elucidate the functional role of the kringle domains, relatively little is known about interactions with other protein domains. Most of the crystal structures describe the interactions at the kringle active site. This study suggests a novel way to interpret structural results such as disorder located away from an active site. The crystal structure of human plasminogen kringle 4 (PGK4) has been refined against 10–1.68 Å resolution X-ray data ($R_{\rm merge}$ = 3.7%) to the standard crystallographic $R$ = 14.7% using the program *X-PLOR*. The crystals of PGK4 showed significant instability in cell dimensions (changes more than 1.5 Å) even at 277 K. The refinement revealed structural details not observed before [Mulichak, Tulinsky & Ravichandran (1991). *Biochemistry*, **30**, 10576–10588], such as clear density for additional side chains and more extensive disorder. Discrete disorder was detected for residues S73, S78, T80, S89, S91, S92, M112, S132, C138 and K142. Most of the disordered residues form two patches on the surface of the protein. This localized disorder suggests that these residues may play a role in quaternary interactions and possibly form an interface with the other domains of proteins that contain kringles, such as plasminogen. Although, an additional residue D65 was refined at the beginning of the sequence, still more residues near the peptide cleavage site must be disordered in the crystal.

## 1. Abbreviations

Both single-letter and three-letter codes are used for amino acids; PGK4, human plasminogen kringle 4; CAC, cacodylate; PEG, polyethylene glycol; BASA, benzylamine sulfonic acid; SDS, sodium dodecyl sulfate; r.m.s.d., root-mean-square deviation; PMSF, phenylmethyl sulfonyl fluoride.

## 2. Introduction

Kringles are homologous domains found in many proteins present throughout blood coagulation and fibrinolytic pathways (Edelberg & Pizzo, 1991; Matsumoto & Nakamura, 1992; Ponting, Marshall & Cederholm-Williams, 1992). A large number of kringles also are found in lipid-transporting proteins such as apolipoprotein a (McLean *et al.*, 1987). Kringles function as recognition domains and in enzymes they usually precede the catalytic domains [*e.g.* thrombin (Tulinsky, Park, Mao & Llinas, 1988) and plasminogen (Ponting *et al.*, 1992)]. They constitute approximately 80-amino-acid separate folding units (Radek & Castellino, 1988).

Specific functions of many kringle domains are not fully understood, but kringle 4 has been extensively studied by a variety of different biochemical methods. Kringle 4 domain is known to bind lysine, fibrin fragments and other zwitterionic ligands (De Marco, Petros, Laursen & Llinas, 1987). The initial model of the active site proposed by Tulinsky (Tulinsky, Park & Skrzypczak-Jakun, 1988) was later confirmed by the crystallographic structure refined at 2.2 Å (Wu, Padmanabhan, Tulinsky & Mulichak, 1991). Crystal structures of kringle 1 and kringle 2 domains as well as other crystal forms of kringle 4 also have been reported in the literature (de Vos, Ultsch, Kelly, Padmanabhan & Tulinsky, 1992; Mulichak *et al.*, 1991; Tulinsky, Park, Mao *et al.*, 1988).

Sequence alignments as well as spectroscopic studies suggest that kringles are structurally homologous (Atkinson & Williams, 1990). Because the tertiary interactions are conserved, knowledge of their three-dimensional quaternary interactions would shed additional light on their functional role in different proteins.

We initiated the project of determining the three-dimensional structure of kringle 4 to test our ability to solve its X-ray structure from a model constructed by homology modeling from prothrombin fragment 1. We had been successful previously in predicting and improving the structure of $\alpha_1$-purothionin constructed from crambin atomic data (Teeter, Ma, Rao & Whitlow, 1990). However, the coordinates for kringle 1 were not available when our diffraction data were collected. We built the original model based on the published figures. We obtained a stereochemically reasonable model which was optimized by molecular dynamics but we failed to solve the structure. The original model evaluated in hindsight had an r.m.s.d. of 1.48 Å from the final structure with the 65–81 loop and a few individual amino-

† Present address: Rigaku Corporation, X-ray Research Laboratory, 3-9-12 Matsubara-cho, Akishima-shi, Tokyo 196, Japan.




ISSN 0907-4449 

acid side chains excluded. These fragments showed more differences than the rest of the structure. We concluded that in a case when the search molecule has a low secondary-structure content, the probe needs to be very close to the target. Additionally, we observed that such a structure is very prone to accidental distortions introduced by prolonged molecular dynamics.

Later, when kringle 4 coordinates became available, we decided to pursue the crystal structure determination because of the higher resolution and lower data collection temperature for our structure. Moreover, it was known that the sequence of the isolated protein may vary because of different elastase cleavage sites (M. Teeter, data not shown).

Here, we present the structure solved by rigid-body refinement from published data (Bernstein *et al.*, 1977; Mulichak *et al.*, 1991). The structure of human plasminogen kringle 4 presented here was refined both at higher resolution (1.68 Å) than reported before (1.9 Å) (Mulichak *et al.*, 1991) against an almost two times larger data set and also at lower temperature, 277 K. This structure confirms previous findings concerning the folding and architecture of the active site. It provides further insight into the intrinsic details of intermolecular interactions of kringles with other domains.

We were able to describe the structure in much more detail because of the availability of better data. It was possible to locate side chains which were not detected in the previous structure determination including an additional Asp residue found before the initial Cys66. Disorder was modeled for several amino-acid side chains reported previously to have single conformations but high temperature factors. Below we try to describe and speculate on the possible role of this localized disorder as it may provide some information about the interactions of kringles with other domains in plasminogen and other multimeric proteins that have kringle domains.

## 3. Materials and methods

The protein was prepared by proteolytic cleavage of human plasminogen with elastase in our laboratory (according to procedures developed by Dr Richard Laursen from whom we obtained human plasminogen). The protein preparation usually started with 1.5 g of lys-plasminogen and 10 mg of soybean trypsin inhibitor dissolved in 100 ml 0.3 *M* $NH_4HCO_3$. 4.5 mg of elastase was added and the mixture stirred for 3.5 h at room temperature. 13 mg of PMSF was dissolved in 2 ml dimethyl formamide and added to the digestion. After stirring it and addition of solid $NH_4HCO_3$ to the final concentration of 0.5 *M*, the digest was stirred overnight. The digest was centrifuged for 30 min at 43 500*g* and supernatant loaded on a gel-filtration G-75 column.

Kringle 4 eluted from the column as the third peak. Subsequently the lyophilized peak II was dissolved in 40 ml of 0.1 *M* $NH_4HCO_3$ and loaded on the lysine sepharose column. The column was eluted with a gradient of 750 ml of 0.1 *M* $NH_4HCO_3$ and 0.01 *M* ε-aminocaproic acid in the same buffer. Peak IV was lyophilized and, after dissolving in water and dialysis against 0.01 *M* potassium phosphate at pH 6.9, applied on the CM-Sephadex column. Kringle 4 was eluted with a gradient of 0.01–0.25 *M* of potassium phosphate. The purity was checked by urea–SDS electophoresis as well as by NMR (De Marco *et al.*, 1985).

Crystals were grown by a vapor-diffusion sitting-drop technique in 20 µl droplets containing ~35 mg $ml^{-1}$ of protein (2.5 m*M*) in 50 m*M* CAC buffer at pH 6.0. PEG 4000 was used as a precipitant, where the droplets containing 10% concentration of PEG were equilibrated against 22% PEG in a reservoir. Despite the fact that the material used for crystallization showed a single band on the urea–SDS electrophoretic gel, it must have been inhomogenous because the crystals were very fragile and difficult to reproduce. They were stabilized when 10% of BASA was added. The protein crystallized in the orthorhombic $P2_12_12_1$ space group. The space group and cell dimensions were similar to those originally published (Mulichak *et al.*, 1991) despite the fact that original crystals were grown from ammonium sulfate solution.

The diffraction data were collected with Cu $K\alpha$ radiation on the Multiwire Area Detector Mark II system (Area Detector Systems, San Diego, CA, USA) at the Crystallographic Facility in the Chemistry Department of Boston College. The one area-detector system was controlled by a DEC Vaxstation 3500 computer. The system was mounted on a Rigaku RU-200 rotating-anode X-ray generator operated at 40 kV and 100 mA. The detector was positioned at 17 and 33° with a crystal-to-detector distance of 400 mm. This experimental setup allowed the collection of data to 1.67 Å resolution. Individual data collection lasted typically ~50 h. Each

```
                65
374 VQ DCYHG DGQSYRGTSS TTTTGKKCQS WSSMTPHRHQ KTPENYPNAG
                                              144
421 LTMNYCRNPD ADKGPWCFTT DPSVRWEYCN LKKC SGTEASV
```

Fig. 1. The sequence of the refined fragment which begins at residue Asp65 (which is equivalent to Asp376 in the amino-acid sequence of human plasminogen). The expected flanking regions not seen in the crystal structure are in bold. The plasminogen numbering scheme is indicated at the beginning of the line, the sequence numbers used in this paper are above the one letter amino-acid codes. It differs from the numbering scheme used by Mulichak *et al.* (1991) where the Cys377 of human plasminogen is called Cys1.

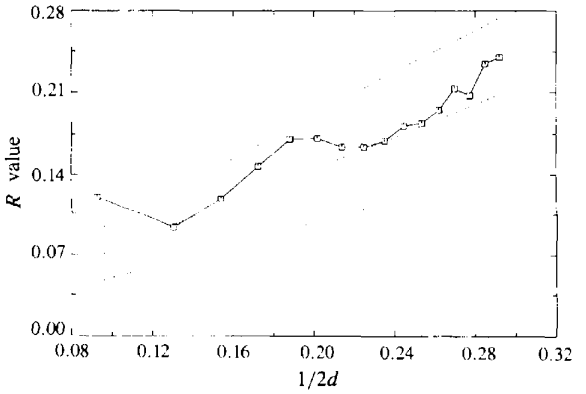


Fig. 2. The Luzzati plot showing that positional errors of the model are ~0.15 Å. Broken lines indicate the errors at 0.1, 0.15 and 0.2 Å, respectively.

data set was collected on a single crystal with overall decay of 9 and 6%, respectively. The data-collection procedures resulted in 86 and 88% complete data sets to 1.68 Å resolution (no $\sigma$ cut off).

We collected two data sets at 1.68 Å resolution on the same crystal. The first one collected on the freshly mounted crystal in a cold stream at 277 K showed smaller cell dimensions $a=49.22$, $b=49.87$, $c=32.27$ Å and higher $R_{merge}$ = 8.7%. We realized that the $R_{merge}$ was relatively high because the cell dimensions were continuously changing during data collection. They increased by ~1.5 Å (mostly in the $b$ direction) and subsequently decreased settling to the values $a=49.44$, $b=50.37$, $c=32.30$ Å. The second better data set with this unit cell was selected for refinement. It was collected after the crystal was in the cold stream for one week and had

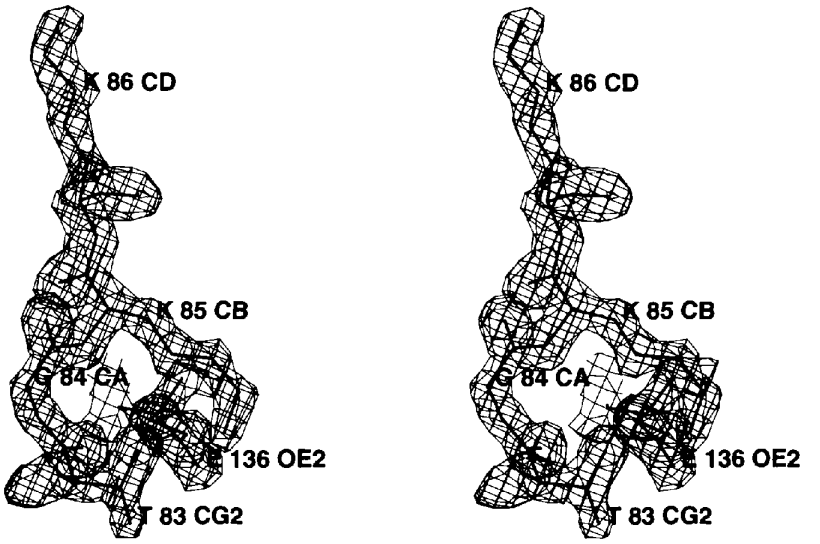


Fig. 3. An example of $2F_o - F_c$ electron density covering two exposed to solvent lysine residues (85, 86) contoured at the 1.5$\sigma$ level.

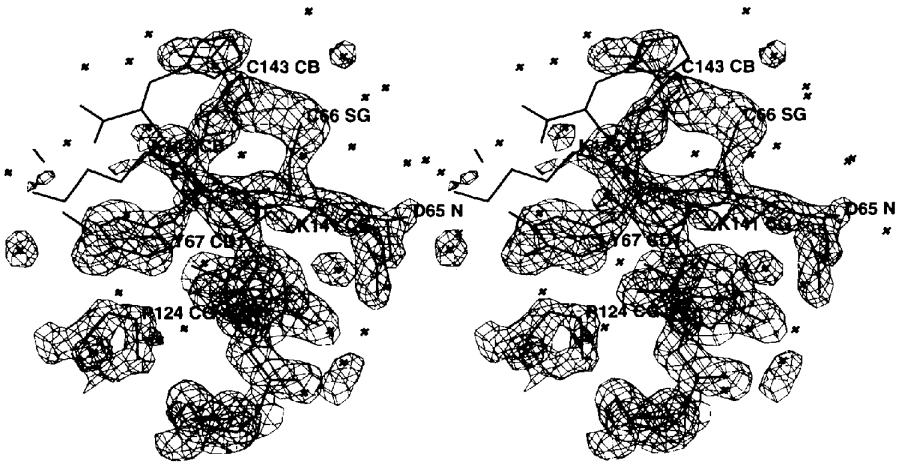


Fig. 4. The N–C termini fragment covered by the weakest electron density ($2F_o - F_c$) contoured at 1.5$\sigma$ level. Please note the electron density covering Asp65. The density is weak and disconnected but clearly interpretable at residues Lys141 and Lys142.

Table 1. *The statistics for the second data set and the R-factor distribution of the final model versus resolution*

| Resolution (Å) | $I/\sigma(I)$ | $R_{merge}$ (%) | Completeness Shell (%) | Completeness Sphere (%) | $R$ value Shell (%) | $R$ value Sphere (%) | $R_{free}$ Sphere (%) |
|---|---|---|---|---|---|---|---|
| 3.38–10.0 | 26.7 | 1.88 | 92.8 | 92.8 | 9.78 | 9.78 | 22.14 |
| 2.68–3.38 | 16.8 | 2.34 | 90.6 | 91.8 | 13.05 | 11.03 | 23.29 |
| 2.34–2.68 | 10.2 | 5.37 | 87.7 | 90.5 | 16.15 | 12.13 | 23.59 |
| 2.13–2.34 | 7.5 | 8.18 | 86.0 | 89.4 | 16.92 | 12.88 | 23.08 |
| 1.98–2.13 | 6.4 | 10.52 | 83.6 | 88.3 | 17.88 | 13.44 | 23.24 |
| 1.86–1.98 | 4.5 | 11.65 | 82.3 | 87.3 | 19.99 | 14.01 | 23.21 |
| 1.77–1.86 | 3.8 | 12.98 | 82.7 | 86.7 | 21.36 | 14.51 | 23.17 |
| 1.68–1.77 | 2.7 | 15.35 | 45.2 | 81.6 | 23.23 | 14.73 | 23.18 |

an $R_{merge} = 3.7\%$. The data set contained 7850 reflections and it was 81.6% complete [$>2\sigma(F)$] to 1.68 Å resolution (Table 1).

The atomic model for kringle 4 was retrieved from the Protein Data Bank (reference code 1pgk) (Bernstein *et al.*, 1977; Mulichak *et al.*, 1991). The same space group and similar cell dimensions ($a = 49.09$, $b = 49.39$, $c = 32.11$ Å) suggested that the position of the molecule was very similar to the original structure. Rigid-body refinement was performed on the initial model without water and the final position which gave $R = 41.0\%$ was chosen as a starting point for the full-atom refinement. The rotation proved to be smaller than 1° and translation below 1 Å.

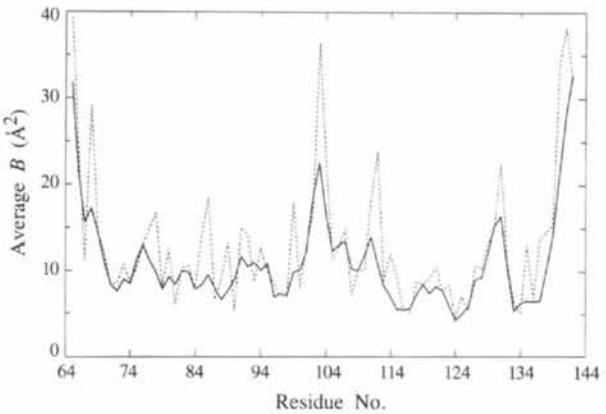


Fig. 5. Plot of temperature factors *versus* residue numbers shown as a solid line for main-chain atoms and as a dotted line for side chains.

The refinement was conducted using simulated-annealing method as implemented in the program *X-PLOR* (Brünger, 1992). Five heat–cool cycles of refinement were alternated with graphics modeling sessions. The heat–cool cycle consisted of 150 cycles of Powell minimization followed by 1 ps molecular dynamics at 1000–2000 K and slow cooling, beginning at the temperature applied at the molecular dynamics (MD) stage and ending at 100 K. Next, two rounds of Powell minimization and individual temperature-factor refinement completed the full heat–cool cycle. Convergence was very fast and after the first cycle the $R$ factor dropped to 28%. At this stage a sulfate ion was modeled along with 31 water molecules and all the side chains missing from the initial structure (Mulichak *et al.*, 1991). These include Thr77, Glu103, Lys141 and Lys142.

The numbering scheme used in this paper has been derived from the alignment of kringle domains presented by Tulinsky, Park, Mao *et al.* (1988). The sequence

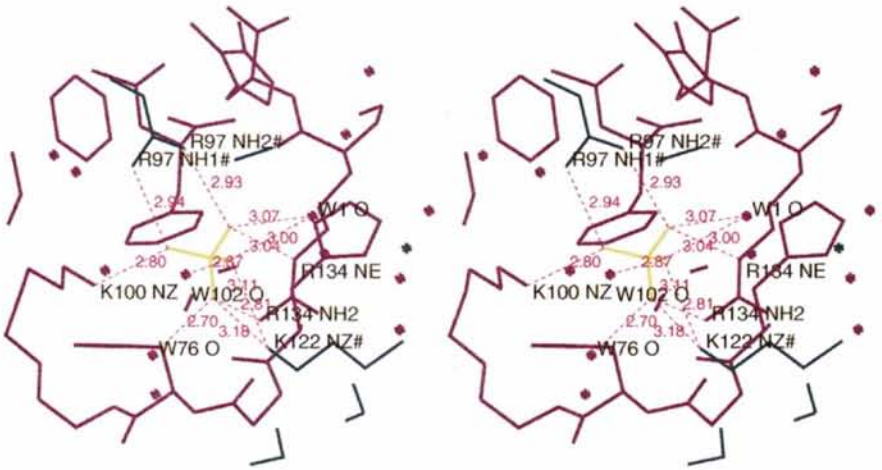


Fig. 6. Pattern of the sulfate (in yellow) hydrogen bonds (red). Note the pattern of hydration and bonds to residues from the symmetry-related molecules (in thinner light-blue lines and marked by #).

of the refined fragment begins at residue Asp65 which is equivalent to Asp376 in the amino-acid sequence of human plasminogen (see Fig. 1).

After an additional four heat–cool cycles with gradual modeling of water molecules and correcting small conformational problems, the refinement reached $R = 14.7\%$ as calculated with reflections from 10 to 1.68 Å resolution. The $R$ factor when calculated on all 7850 reflections (30–1.68 Å) was 16.5%. The model contains one sulfate ion in the lysine binding site and 256 water molecules. When this model without disorder was transferred to *PROLSQ* (Hendrickson & Konnert, 1980) and subjected to five cycles of least-squares refinement, the $R$ factor dropped to 13.7%. Despite the fact that the model refined by *PROLSQ* did not contain disorder, it produced a lower $R$ factor. This may indicate the inherent differences

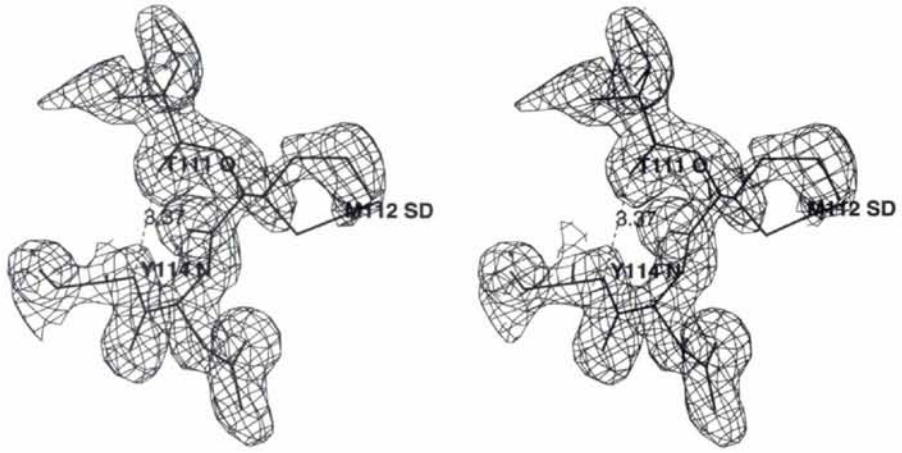


Fig. 7. The electron density contoured at $1.5\sigma$ covering residues 111–114. Note the unusual backbone conformation ($\varphi = 49.2$, $\psi = -132.7°$) and the discrete disorder of the side chain of Met112 (M112).

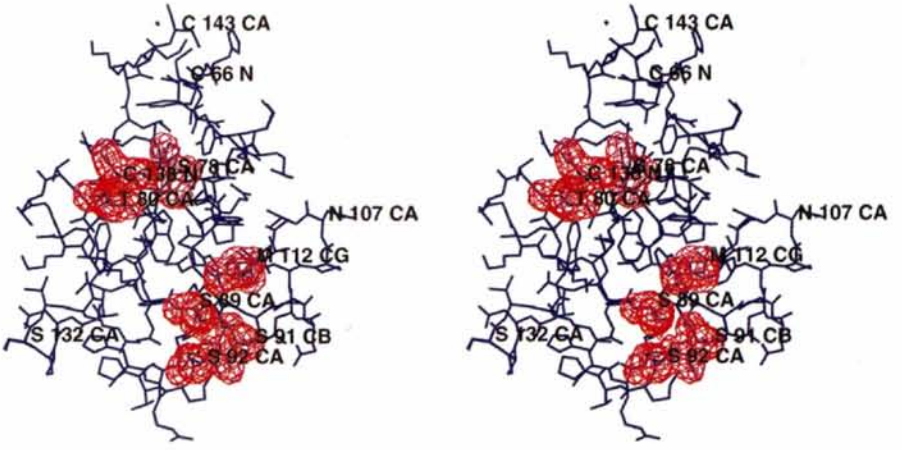


Fig. 8. The van der Waals surface (red) covering the patches of disordered residues on the surface of the kringle 4 domain of human plasminogen. Upper patch is comprised of Cys138, Thr78, Ser73 and the lower patch of Met112, Ser89, Ser91, Ser92.

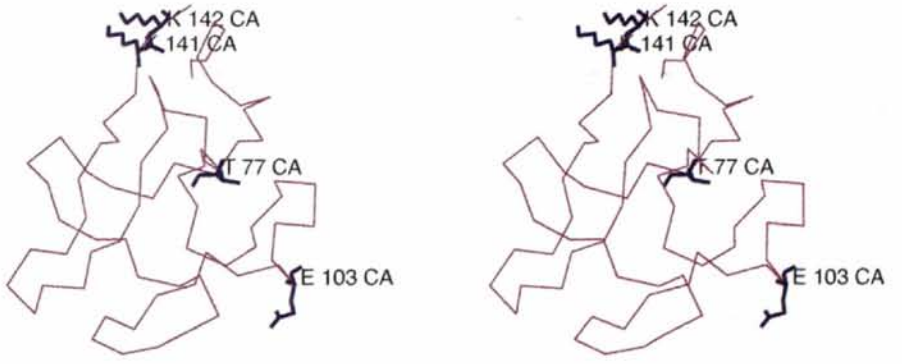


Fig. 9. Cα model of the refined kringle 4 domain (thin violet line) with newly located side chains (thick dark-blue lines).

between the two programs in handling resolution shells of data as well as weighting schemes.

In the last rounds of modeling ten residues were found to be discretely disordered. The disordered residues are: S73, S78, T80, S89, S91, S92, M112, S132, C138, K142. Most of these residues were reported before (Mulichak *et al.*, 1991) as having high temperature factors. We also observed the disorder at Cys138 as it was described before.

The pattern of disordered residues may indicate that those residues are usually in contact with the other domains in native proteins.

## 4. Results and discussion

### 4.1. *The model*

The final model exhibits good geometry with an r.m.s.d. on bonds of 0.009 Å and on angles of 1.89°. The geometry was checked by the program *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993). The results indicated a high-quality model. The Ramachandran plot (Ramachandran, Ramakrishnan & Sasisekharan, 1963) is very similar to that presented by Mulichak *et al.* (1991) with all the residues in the most preferred regions except for one. The refinement confirmed an unusual backbone conformation for Met112 ($\varphi = 49.2$, $\psi = -132.7°$). As it will be discussed later, this unusual conformation might be an artifact of exposing the methionine residue to the less preferred hydrophilic environment. As noted before, the structure has one *cis* peptide bond at Pro93 (Mulichak *et al.*, 1991).

The torsional angles for all side chains fall close to the expected rotamer values (Ponder & Richards, 1987). The positional errors estimated by the Luzzati method (Luzzati, 1952) showed that positional uncertainty of atoms is around 0.15 Å (Fig. 2). The hump at medium resolution was interpreted as being associated with the disorder fragments not seen in the crystal structure as discussed later. The $R_{free}$ calculated on 10% of the data was 23.2% indicating a well refined structure. The quality of the model is best demonstrated by the electron

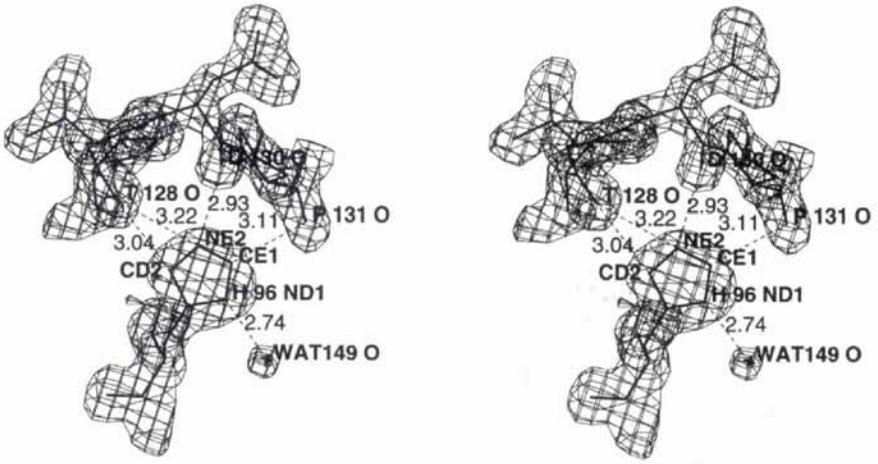


Fig. 10. Bifurcated hydrogen bond of His96 NE2. Close contacts are also provided by carbonyl O atoms to the carbon H atoms of the imidazole ring. The $2F_o - F_c$ electron density is contoured at the $1.5\sigma$ level.

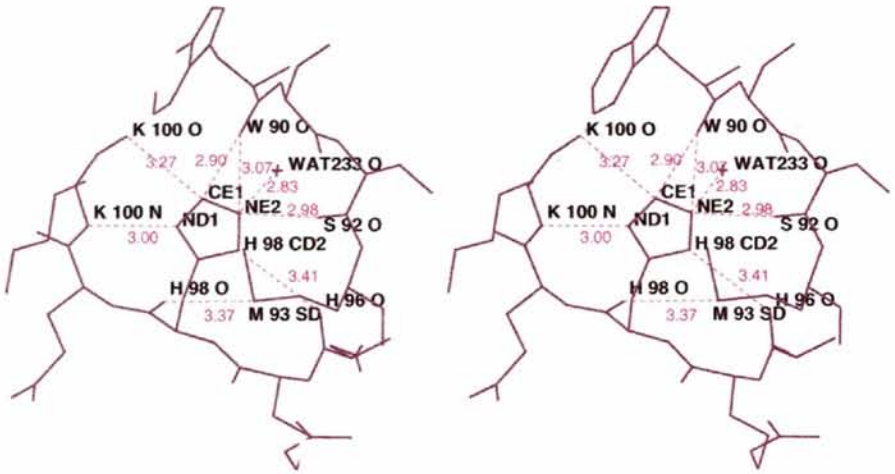


Fig. 11. Trifurcated hydrogen bond of His98 (H98) NE2. The hydrogen bond between Met93 (M93) SD and the carbonyl O atom is also depicted, as are short contacts to CE1.

density covering almost all the protein atoms at the $1.5\sigma$ level (*e.g.* Fig. 3).

The density is weaker at the region where the N-terminal and C-terminal residues come together and are connected by a disulfide bridge (Fig. 4). Temperature factors in this region reach 30 Å$^2$ while the rest of the structure has temperature factors significantly below 15 Å$^2$ (Fig. 5). Weaker densities and higher temperature factors correspond to less ordered fragments of the model. The N–C terminal fragments are flanked by the unrefined fragments of the cleavage peptide (in total nine amino acids) which explains higher mobility. This fact also sheds some light on the changes in unit cell. If the conformation of the unrefined fragments were changing during data collection, then packing of individual molecules would be perturbed leading to the observed changes in cell dimensions. Nevertheless, the density for the additional amino acid (Asp65) refined at the beginning of the sequence and not reported before is very clear (Fig. 4).

### 4.2. *Active site*

As described in the *Introduction*, kringle 4 has a zwitterion binding site. It can bind different zwitterionic compounds such as the $\varepsilon$-aminocaproic acid (Wu *et al.*, 1991), but its primary function is to recognize fibrin by binding at a surface lysine residue. The lysine binding-site residues are very well determined with low average temperature factors (~10 Å$^2$). The site defined by two aspartate residues (119, 121) and Arg134 and Lys100 is occupied by the sulfate ion (Fig. 6). Since sulfate was not used for crystallization it probably came from the hydrolysis of BASA. This fact could explain the stabilizing role of BASA in crystallization since no trace of BASA electron density could be found. The sulfate ion has a saturated pattern of hydrogen bonds with three O atoms having three hydrogen bonds and one O atom with two hydrogen bonds (Fig. 6). Six of these bonds come from protein donors and five from water molecules. Three out of six protein hydrogen-bond donors come from symmetry-related molecules (Arg97′ and Lys122′). It should be noted that two Arg and two Lys residues each from a different molecule are involved in binding to the sulfate ion. Without the sulfate ion a group of positively charged residues coming from different molecules would be without the counterion and would contribute to the instability of the crystal lattice. The fact that Arg97′ coming from the symmetry-related molecule occupies the site in the manner similar to binding of $\varepsilon$-aminocaproic acid (Wu *et al.*, 1991) gives additional importance to the intermolecular interactions at the active site.

### 4.3. *Disordered residues*

Ten residues (13%) in two alternative conformations (Table 2) were modeled (S73, S78, T80, S89, S91, S92, M112, S132, C138, K142). The commonly used criteria were used to identify the discrete disorder. If the temperature factors for a particular residue were significantly higher that neighboring atoms and it was accompanied by the elevated positive difference density we modeled the particular residue in an alternative conformation. The new side-chain conformation was always modeled with torsional angles of the allowed rotomer. Because temperature factors are highly correlated with occupancy factors, we refined only the temperature factors with the occupancy kept fixed. We changed the occupancy until a similar level of the temperature factors for both alternates was achieved (Smith, Hendrickson, Honzatko & Sheriff, 1976) which in most of the cases lead to a comparable level of temperature factors to those of neighboring atoms. Most of the disordered residues refined to a similar level occupancy (70–30%). One of the three disulfide bridges (Cys115—Cys138) is also disordered. The density indicated that the occupancy ratio of the disordered S atom of Cys138 is close to 2:1 and not 1:1 as observed previously (Mulichak *et al.*, 1991). In both side-chain conformations for Cys138 (*t* and +*g*) the disulfide bond is formed with distances 1.97 and 2.03 Å, respectively.

Table 2. *Conformation of disordered residues*

Major conformer comprises atoms refined with occupancy 70%, minor with 30%.

| Residue | Major conformer* | Minor conformer* |
|---|---|---|
| Ser73 | + | − |
| Thr78 | − | t |
| Ser80 | − | + |
| Ser89 | − | t |
| Ser91 | + | − |
| Ser92 | t | − |
| Met112 | t,t,−94.6° | −,− ,−78.9° |
| Ser132 | − | + |
| Cys138 | t | − |
| Lys142 | −,t,t,t | t,+,t,+ |

* Notation used: + ≃ +60°; − ≃ −60°; t ≃ 180°.

The disordered side-chain conformation of Met112 is illustrated in Fig. 7. The carbonyl O atom of the preceding residue is involved in the formation of the reverse turn. The carbonyl group of Met112 is involved in a strong water-mediated hydrogen bond to the amide N atom of Ser91. This forces the Met112 backbone into an unfavorable Ramachandran region and its side chain to be disordered. However, if this residue is in a different environment then the backbone can adopt a more regular conformation and as well as the side chain can be fully ordered.

The majority of disordered residues form two closely spaced patches on the surface of the molecule (Fig. 8). The first patch comprises the residues Cys138, Ser78 and Thr80. The second one is formed by Met112, Ser89, Ser91 and Ser92. It is not unusual to see disorder

in surface residues, especially serines, but it is a little surprising to see those residues forming well defined patches. Most of these residues come into direct van der Waals or hydrogen-bonded contact which would propagate the disorder.

There are two explanations for this which are not mutually exclusive. The first one would be related to diffuse-scattering experiments (Caspar, Clarage, Salunke & Clarage, 1988; Clarage, Clarage, Philips, Sweet & Caspar, 1992) in which one can see that the length of correlated motion is close to the size of individual patches (6–8 Å). Crystallographic experimental evidence for the existence of such correlated disorder has come from crambin crystals in which such patches of disordered residues were observed (Yamano & Teeter, 1994).

The second plausible explanation may come from the spatial relationships between the active-site and disordered-residue areas. Both patches are located at the opposite face to the active site of kringle 4. It seems reasonable to assume that when side chains that are usually buried are exposed to a non-native environment, they may undergo a conformational change detected as disorder. If the second possibility is correct it might indicate that those residues are involved in interdomain contacts in proteins of which the kringle is one domain.

Unfortunately, native kringle 4 crystal contacts do not shed much light on this subject because they consist primarily of the polar and charged interactions at the active-site surface as described before (Mulichak *et al.*, 1991). However, there are crystal structures of other multidomain proteins which contain the kringle domain. One such example is a meizothrombin (B. Edwards, personal communication). In this structure, the disordered residues in the most conserved patch (corresponding to our residues 78 and 80) come into contact with the thrombin domain.

### 4.4. *Newly located side chains*

In the previous crystal structure determination, the electron densities for several amino-acid residues were not visible. These included Thr77, Glu103 and two lysine residues at the C terminus, Lys141 and 142. As a result these side chains were not included in the refinement procedure. The higher resolution of the data and possibly lower temperature (277 K) allowed us to localize the missing side chains (Fig. 9). All of them have higher temperature factors than average (Fig. 5) and some of them are disordered (*e.g.* Lys142).

We have also detected additional density at the N terminus of the fragment and refined an additional residue (Asp65). The Asp65 is clearly visible in the $2F_o - F_c$ and omit electron-density maps. Kringle 4 fragment was prepared by proteolytic cleavage of human plasminogen with elastase. The consensus sequence for a cleavage site of this enzyme is small non-polar residues (*e.g.* alanine or valine). Nonetheless, the aspartate residue refined well at this site and had many side-chain hydrogen bonds. The fact that aspartate cannot be at the cleavage site indicates that the rest of the cleavage peptide must be disordered in the crystal. There is enough room between symmetry-related molecules to accommodate additional amino acids. A slightly elevated level of the difference electron density is also consistent with this conclusion, but no amino acids could be modeled.

### 4.5. *Hydrogen bonds*

Kringle 4 folding architecture is highly unusual. Besides three short tracks of $\beta$-ladder all other backbone hydrogen bonds are formed at turns or crisscross the molecule in an irregular pattern. Therefore, the side-chain to backbone hydrogen bonds have to lend additional stability to the overall structure. They contribute 36 hydrogen bonds *versus* 29 formed by only backbone atoms. In contrast side-chain to side-chain hydrogen bonds play a lesser role contributing only 11 hydrogen bonds. Patterns of hydrogen bonds in the model were described in some detail in the previous publication (Mulichak *et al.*, 1991). However, the higher resolution of this structure allowed us to notice some interesting hydrogen bonds not described before. For instance, three $\gamma$-turns with very good geometry were identified. Those are comprised of Gln99 O···Thr101 N (3.12 Å), Gly109 O···Thr111 N (2.93 Å) and of Cys138 O···Leu140 N (3.01 Å). Additionally, both internal histidine residues (His96 and His98) have bifurcated or even trifurcated hydrogen bonds (Figs. 10 and 11).

Both N atoms of the imidazole group of His98 side chain are hydrogen bonded to backbone atoms. The ND1 atom is bonded to an amide N atom of Lys100, and the NE2 N atom is bound to two backbone carbonyl groups (Ser92 and Trp90). The NE2 N atom is additionally hydrogen bonded to a water molecule. This may mean that this N atom is involved in an exceedingly rare [<1% (Taylor, Kennard & Versichel, 1984)] trifurcated hydrogen bond (Fig. 9). The occupancy of the water molecule is only 0.5, which would indicate a weak hydrogen bond present about 50% of the time. Nevertheless, this N atom would have at least a bifurcated hydrogen bond. Additionally, it is suggested by the pattern of hydrogen bonds that His98 is deprotonated which might bear on the recognition and binding of the zwitterionic compounds.

The side-chain conformation of the imidazole ring of His96 can be easily defined because the 'external' N atom (ND1) of histidine is hydrogen bonded to the adjacent water molecule (Wat149). NE2 is doubly hydrogen bonded to the carbonyl groups of Thr128 and Asp130. What is even more interesting, the conformations of both histidine residues are further stabilized by close contacts of carbon-bound H atoms to adjacent carbonyl groups of different residues (Figs. 10 and 11). These may represent

CH···O hydrogen bonds (Derewenda, Derewenda & Kobos, 1994; Steiner & Saenger, 1993).

Another interesting hydrogen bond detected in the same region involves the S atom of methionine 93 and the carbonyl O atom of His98. The distance is 3.37 Å, very similar to the distances observed in other proteins (Gregoret, Rader, Fletterick & Cohen, 1991). Furthermore, the S atom of this methionine is in close contact with the His98 amide N atom (M93 SD—3.68 Å—H98 N) and the indole N of Trp135 (M93 SD—3.66 Å—W135 NE1) of the symmetry-related molecule. This tryptophan indole N atom is also involved in a hydrogen bond with water 81.

The side chains which were not observed in the structure by Mulichak *et al.* (1991) contribute also to the stability of the structure through an intrinsic network of hydrogen bonds. The carboxyl group of Asp65 is involved in stabilizing the N terminus through hydrogen bonds to the backbone (D65 OD1···D65 N = 3.25 Å) and to the hydroxyl group of Tyr74 and Ser78 (D65 OD2···Y74 OH = 2.70 Å, D65 OD2···S78 OG = 3.17 Å). The Thr77 hydroxyl coordinates three water molecules. It is hydrogen bonded through one of them (Wat172) to the OD2 of Asp65 (OG—3.2 Å—O—3.2 Å—OD2).

Glu103 also is involved in stabilizing the turn where it is located by making a hydrogen bond to the backbone N atom (E103 OE1···E103 N = 3.0 Å). Lysine residues 141 and 142 make additional hydrogen bonds to symmetry-related molecules through the water molecule. They contribute to the stabilization of the lattice. The fact that both lysines are visible in the electron density (on low contouring levels) despite the expansion of the lattice in the *b* direction is directly associated with stabilizing contacts involving these residues.

## 5. Summary

In this paper we present the structure of kringle 4 at higher resolution than reported before, 1.68 Å, and 277 K. This high-resolution refinement confirmed all of the aspects of the previously reported structure and provided additional molecular details. The side chains missing from the previous structure were detected and described as well as the pattern of disordered residues. Most of these residues form a patch on the surface of the protein which was concluded to have some relevance for the quaternary assembly of the proteins containing kringle domains. One additional amino acid (Asp65) was detected and refined at the N terminus of the protein, which is inconsistent with the cleavage site for the proteolytic enzyme elastase. This would mean that additional fragments of the peptide where the cleavage did occur have to be disordered in the crystals. It suggests also that some parts of the electron density near the N and C termini, currently interpreted as coming from localized water molecules, might in fact reflect the mobile portion of the cleaved kringle beyond the disulfide.

We would like to thank Dr R. A. Laursen for the gift of partially purified plasminogen and helpful discussions. Contributions to the project by Drs B. Rasmussen and J. Petithory are gratefully acknowledged. Grants from the NIH (GM 38114 and GM 40601) have supported this research.*

* Atomic coordinates and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1KRN, R1KRNSF). Free copies may be obtained through The Managing Editor, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England (Reference: GR0638). At the request of the authors, the structure factors will remain privileged until 21 December 1998.

## References

Atkinson, R. A. & Williams, R. J. P. (1990). *J. Mol. Biol.* **212**, 541–552.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Brünger, A. T. (1992). *X-PLOR Version* 3.1, *A System for X-ray Crystallography and NMR*. New Haven and London: Yale University Press.

Caspar, D. L. D., Clarage, J., Salunke, D. M. & Clarage, M. S. (1988). *Nature (London)*, **322**, 659–662.

Clarage, J. B., Clarage, M. S., Philips, W. C., Sweet, R. M. & Caspar, D. L. D. (1992). *Proteins Struct. Funct. Genet.* **12**, 145–157.

De Marco, A., Petros, A., Laursen, R. & Llinas, M. (1987). *Eur. Biophys. J.* **24**, 359–368.

De Marco, A., Pluck, N. D., Banyai, L., Trexler, M., Laursen, R. A., Patthy, L., Llinas, M. & Williams, R. J. P. (1985). *Biochemistry*, **24**, 748–753.

Derewenda, Z. S., Derewenda, U. & Kobos, P. M. (1994). *J. Mol. Biol.* **241**, 83–93.

Edelberg, J. M. & Pizzo, S. V. (1991). *Blood Coagl. Fibrin.* **2**, 759–764.

Gregoret, L. M., Rader, S. D., Fletterick, R. J. & Cohen, F. (1991). *Proteins Struct. Funct. Genet.* **9**, 99–107.

Hendrickson, W. A. & Konnert, J. H. (1980). *Incorporation of Stereochemical Information into Crystallographic Refinement*, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Luzzati, V. (1952). *Acta Cryst.* **5**, 802–810.

McLean, J., W., Tomlinson, J. E., Wu-Jing, K., Eaton, D. D., Chen, E. Y., Fless, G. M., Scanu, A. M. & Lawn, R. (1987). *Nature (London)*, **330**, 132–137.

Matsumoto, K. & Nakamura, T. (1992). *Crit. Rev. Oncog.* **3**, 27–54.

Mulichak, A. M., Tulinsky, A. & Ravichandran, K. G. (1991). *Biochemistry*, **30**, 10576–10588.

Ponder, J. W. & Richards, F. M. (1987). *J. Mol. Biol.* **193**, 755–791.

Ponting, C. P., Marshall, J. M. & Cederholm-Williams, S. M. (1992). *Blood Coagul. Fibrin.* **3**, 605–614.

Radek, J. T. & Castellino, F. J. (1988). *Arch. Biochem. Biophys.* **267**, 776–786.

Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). *J. Mol. Biol.* **7**, 95–99.

Smith, J., Hendrickson, W. A., Honzatko, R. B. & Sheriff, S. (1976). *Biochemistry*, **25**, 5018–5027.

Steiner, T. & Saenger, W. (1993). *J. Am. Chem. Soc.* **115**, 4540–4547.

Taylor, R., Kennard, O. & Versichel, W. (1984). *J. Am. Chem. Soc.* **106**, 244–248.

Teeter, M. M., Ma, X. Q., Rao, U. & Whitlow, M. (1990). *Proteins Struct. Funct. Genet.* **8**, 118–132.

Tulinsky, A., Park, C. H., Mao, B. & Llinas, M. (1988). *Proteins Struct. Funct. Genet.* **3**, 85–96.

Tulinsky, A., Park, C. H. & Skrzypczak-Jakun, E. (1988). *J. Mol. Biol.* 202, 885–901.

de Vos, A. M., Ultsch, M. H., Kelly, R. F., Padmanabhan, K. & Tulinsky, A. (1992). *Biochemistry*, **31**, 270–279.

Wu, T., Padmanabhan, K., Tulinsky, A. & Mulichak, A. M. (1991). *Biochemistry*, **30**, 10589–10594.

Yamano, A. & Teeter, M. (1994). *J. Biol. Chem.* **269**, 13956–13965.