This is to certify that the

dissertation entitled

STRUCTURE OF MONOCLINIC
KRINGLE 4 FROM HUMAN PLASMINOGEN

presented by

K. G. Ravichandran

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Chemistry

A. Tulinsky
Major professor

Date Feb. 10, 1989.

STRUCTURE OF MONOCLINIC KRINGLE 4 FROM HUMAN PLASMINOGEN

By

K. G. Ravichandran

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

1989

ABSTRACT

STRUCTURE OF MONOCLINIC
KRINGLE 4 FROM HUMAN PLASMINOGEN

By

K. G. Ravichandran

Monoclinic crystals of kringle 4 from human plasminogen were grown from 35% PEG 8000, 0.1 M ammonium sulfate, 0.8% n-butanol, at pH 6.2 by the hanging drop method. The crystals contain two molecules of kringle 4 per asymmetric with a V_m = 1.86 $Å^3$/dalton. The crystals are twinned along the c^* direction. Three-dimensional intensity data were collected to 3.5 Å resolution for the crystal and 6.0 A resolution for the twin. The twinned reflections in the l=0,1,3,4, and 8 layers were untwinned. The structure was solved by molecular replacement methods. A self rotation search was carried out to find the orientation of the two independent molecules in the asymmetric unit. The search yielded a solution at ϕ=0.0°, ψ=80.0°, and χ=180.00° which corresponds to a non-crystallographic two-fold rotation approximately parallel to the a^* axis of the crystal unit cell. The orientation and position of the two independent molecules of kringle 4 were determined using the refined coordinates of the kringle of prothrombin fragment 1. The former were then refined using the non-crystallographic symmetry averaging procedure. The final rotation angles and the translation vectors for the two molecules are :

	a	b (deg.)	g	TX	TY (Å)	TZ
Molecule A	313.62	142.80	186.88	6.2	0.0	11.3
Molecule B	109.77	149.77	158.77	25.8	48.6	31.8

Phase angles for the structure factor amplitudes were calculated using the conserved structure of the kringle of fragment 1 and an electron density map was calculated with reflections to 2.8 Å resolution. In general, the kringle 4 structures of the independent molecules are very similar in overall conformation to the refined prothrombin fragment 1 kringle. The amino terminal (3 residues) and the carboxy terminal interkringle tail (5 residues) and most of non-conserved aromatic side chains were clearly visible in the electron density. There was even reasonable density in regions which have a deletion in kringle 4 compared to the fragment 1 kringle. From an examination of the molecular arrangement in the crystal lattice, each independent kringle 4 molecule makes close contacts with two other kringles, giving rise to a three dimensional network. This might be of relevance in kringle-kringle interactions in plasminogen which contains 5 kringles and in apolipoprotein(a) which has 37 copies of kringle 4 and one copy of kringle 5. The molecular packing also gives rise to two different "dimer" possibilities; it is difficult at this stage to reconcile these dimeric species to relevant kringle-kringle interactions in physiological systems.

The lysine binding site of kringle 4 is located on the surface of one of the oblate faces of the kringle as

reported previously from modelling studies. The binding site is characterized by an apparent dipolar surface, the polar parts of which are separated by a hydrophobic region of highly conserved aromatic residues. Zwitterionic ligands such as lysine and ε-aminocaproic acid can form ion pair interactions with Asp57 and Asp59 located on the dipolar surface. The cationic center of kringle 4 is formed by Arg72. Overall, the lysine binding site, with the exception of the indole ring of Trp73, which has a different orientation in the crystal structure, compares well with the predicted model of the lysine binding site of kringles.

To my parents

ACKNOWLEDGMENTS

It is my great pleasure to sincerely acknowledge Dr. Alexander Tulinsky for his guidance, support and inspiration throughout the course of this work.

I would like to express my deep gratitude to Dr. Miguel Llinas for providing us with an abundant supply of the material. I would like to thank Anne Mulichak for helping me with the crystals and the data collection. I would also like to thank Dr. Manuel Soriano, Dr. K. Padmanabhan, Trevor M. D'Souza and Tapankumar Nayak for their help in preparing the figures and tables. I am also grateful to Dr. Ewa Skrzypczak-Jankun for the many useful discussions regarding processing the data. My thanks also go to Dr. Pushpa for proof reading the manuscript. I would like to thank Dr. Chang H. Park, Dr. James P. Fillers and Timothy J. Rydel their help and support.

Finally, but not the least, to my family and friends for their love and support.

Support by the National Institutes of Health is gratefully acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

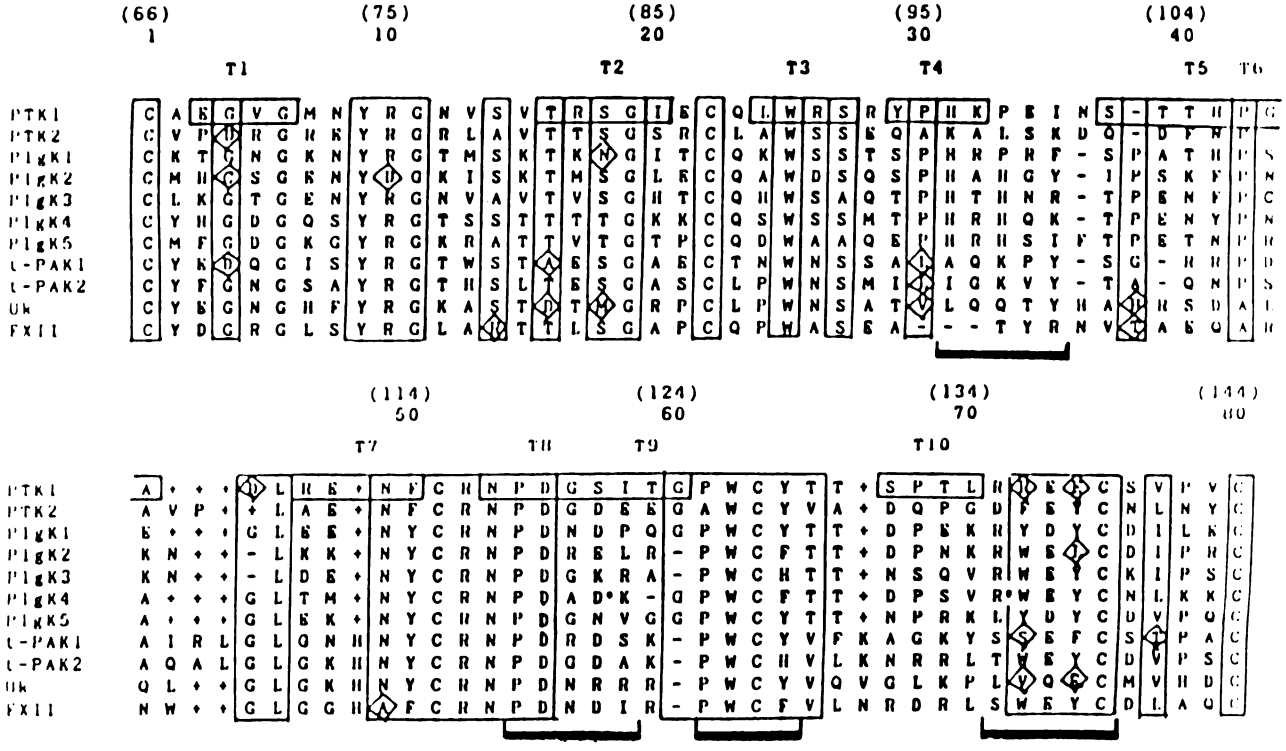
LIST OF FIGURES

I INTRODUCTION

Kringles are autonomous structural and functional folding domains of molecular weight approximately 10,000, that serve as protein binding modules of many proteins involved in blood coagulation and fibrinolysis. They possess characteristic three disulphide triple loop structures made up of approximately 80 residues. They are found singly in urokinase [1,2] and factor XII [3], as pairs in prothrombin [4] and tissue-plasminogen activator (t-PA) [5], as five copies in plasminogen [6] and remarkably as 38 copies in apolipoprotein(a) [7], 37 of which display 75-85% conservation with kringle 4 of plasminogen. It has been shown recently that type II finger domain structures of fibronectin also exhibit kringle homology [8,8a]. The sequences of the kringles in the different proteins mentioned above are listed in Table 1.

The biological roles of the kringles have been established in many cases. Fibrin binding sites are located in kringle 1 [9], kringle 4 [10], and miniplasminogen (kringle 5+light chain of plasminogen), while lysine binding sites have been localized to kringles 1 and 4 of plasminogen [11]. Some of the plasminogen kringles are also capable of binding low molecular weight compounds of the ω-aminocarboxylic acid class at the lysine binding sites [12] while benzamidine sites reside on kringle 5 and the light chain of plasminogen [13]. There is good evidence in the case of plasmin and plasminogen that kringles mediate

Table 1. Primary structure of homologous kringles aligned to maximize conservation.*



*The number of residues (1-80) is standardized to plasminogen kringle 5. Sites inserted or deleted relative to K5 sequence are indicated with + or -, respectively. Conserved residues are boxed, homologous conservation is dotted; obvious deviations from conservation are enclosed by diamond blocks. Segments involved in the fibrin/lysine binding region are underlined with heavy trace. Crucial Asp and Arg residues of plasminogen kringle 4 lysine binding site are designated with asterisks.
The abbreviations used are : PT, prothrombin; PlgK1, kringle 1 of plasminogen, etc.; t-PAK1, kringle 1 of tissue plasminogen activator, etc.; Uk, urokinase; FXII, factor XII (taken from [29]).

attachment of the protease or its zymogen to the target fibrin. Moreover, kringle 2 of prothrombin is known to bind to the heavy chain of membrane-bound factor Va in the prothrombinase complex [14,15] and kringle 2 of t-PA binds to fibrin [16]. Thus kringles appear to be involved in biological recognition and in imparting specificity to accompanying catalytic domains. On the other hand, there does not yet appear to be a clear function for the benzamidine binding site of plasminogen kringle 5, and so far, for the other kringles of plasminogen, prothrombin and t-PA and those of urokinase and factor XII.

Plasmin is a blood plasma serine protease primarily responsible for the dissolution of fibin clots deposited on the walls of blood vessels. Plasmin is produced by the enzymatic activation of its zymogen, plasminogen, a single chain plasma glycoprotein comprised of 791 amino acid residues [17]. Plasmin , on the other hand, is a glycoprotein composed of two chains, covalently linked by two cysteine bridges [6]. The light chain of molecular weight ~ 25000, carries the catalytic function [18], while the heavy chain, of MW ~57000, is formed by a tandem array of five kringles some of which are involved in fibrin binding [6,19]. A schematic of the heavy chain of the plasminogen is shown in Figure 1.

As already mentioned, fibrinolysis is the process of dissolution of the fibrin blood clot of coagulation by an enzyme system present in blood of all mammalian species. The fibrinolytic system consists of the plasma zymogen

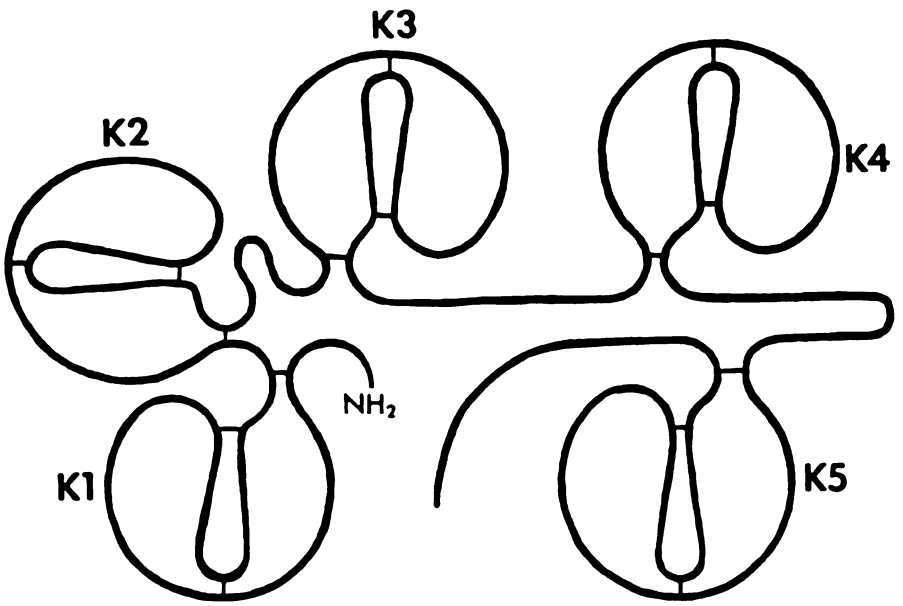


Figure 1. A schematic view of Plasminogen Kringles, K1-K5 [12]

plasminogen, its activated product, the proteolytic enzyme plasmin, activators of plasminogen, inhibitors of both plasmin and plasminogen activators, and fibrinogen and fibrin. On the basis of various chemical and biochemical studies, a molecular model for physiological fibrinolysis has been proposed [20], which explains the restricted action of plasmin **in vivo**. According to this model, the system seems to be regulated at two levels: localized plasminogen activation at the fibrin surface; and sequestration of the formed plasmin from circulating plasma. When fibrin is formed, a small amount of plasminogen is specifically bound to it by way of its lysine binding sites. t-PA present in the blood or released from the vascular endothelium is adsorbed on the fibrin surface and efficiently activates the adsorbed plasminogen (t-PA converts plasminogen to plasmin by cleaving two bonds, Lys76-Lys77 and Arg560-Val561). The formed plasmin has its lysine binding sites occupied in complex formation with fibrin and is also involved in fibrin degradation by way of its active site, and therefore reacts only very slowly with antiplasmin. In contrast, plasmin which is released from digested fibrin is rapidly and irreversibly neutralized by antiplasmin in the circulation.

Kringles 1 and 4 of plasminogen can be isolated **in vitro** by the controlled digestion of the zymogen [6]. The amino acid sequence and the disulfide crosslinks of kringle 4 from human plasminogen are illustrated in Figure 2 with the bridge between Cys4 and Cys81 bringing the N-terminus in the proximity of the C-terminus. It should be

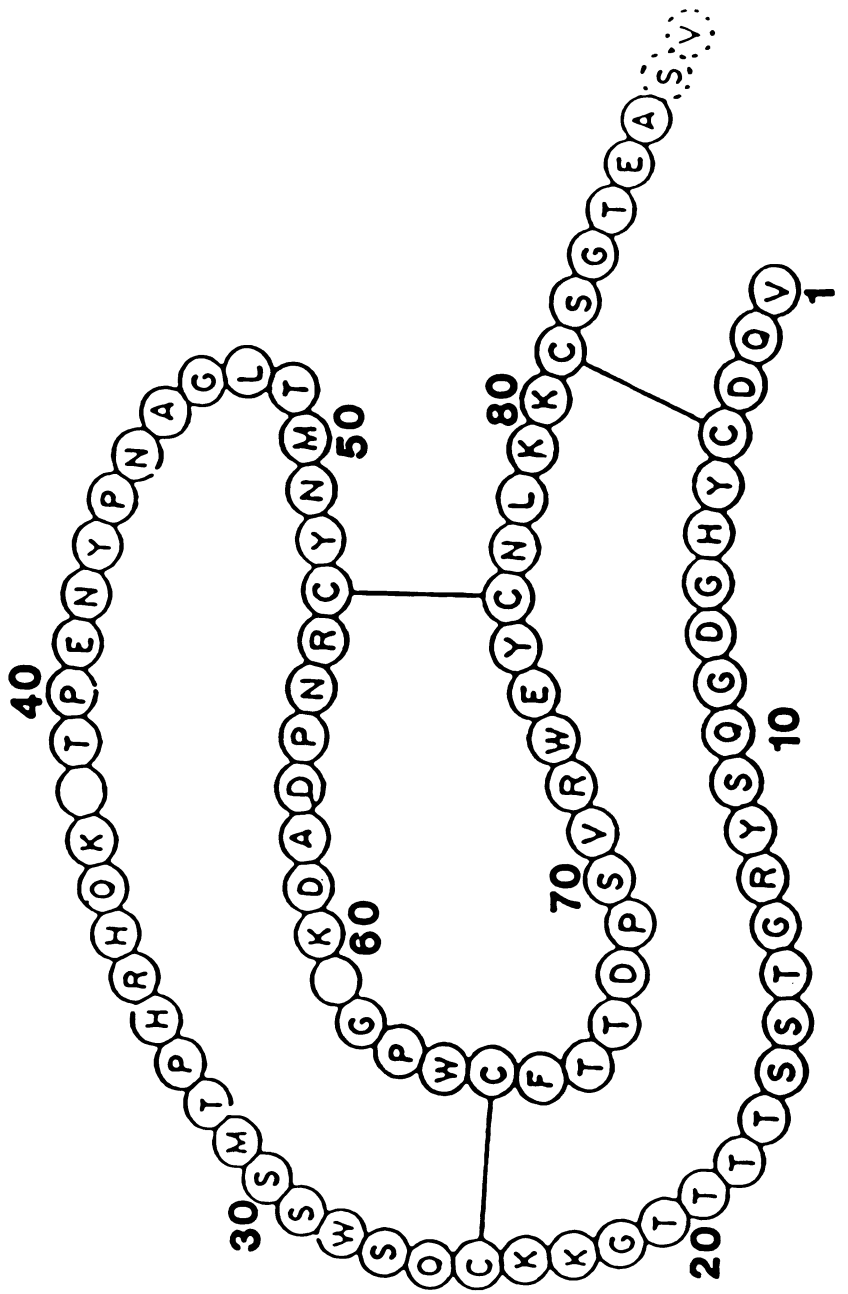
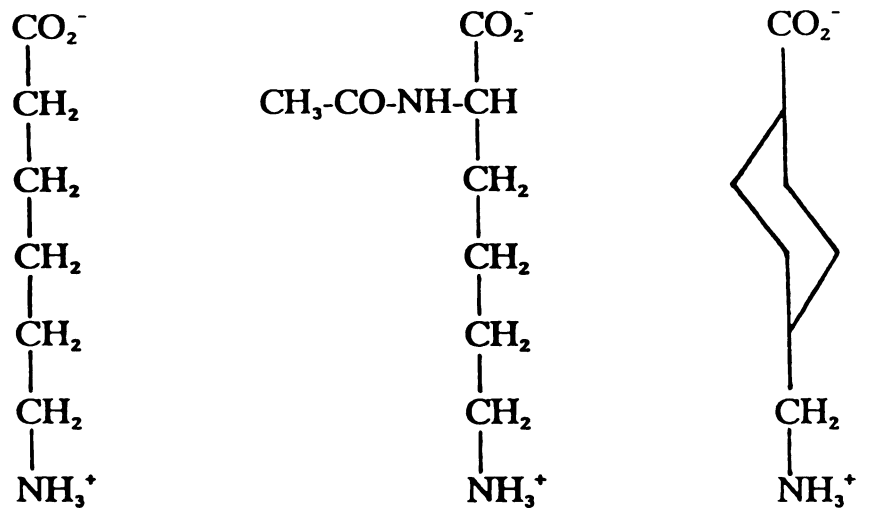


Figure 2. Kringle 4 from human plasminogen: outline of primary structure. Amino acid residues are labelled according to the conventional one-letter code. The predominant kringle 4 species terminates with Ala86 but, depending on the extent of elastase digestion of plasminogen, the C-terminal peptide extend two residues further to Val88. Blanks are deletions with respect to plasminogen kringle 5.

noted,however, that the end product of elastase digestion of plasminogen is a heterogeneous mixture of ~70% C-terminal Ala86 and 30% C-terminal Val88 [21].

Studies carried out on the lysine binding of kringle 4 of plasminogen and other kringles have been successful in characterising the various residues involved in ligand binding. The binding of various aminocarboxylic acids, of ~6.8 Å dipolar distance, to human plasminogen and its proteolytic fragments has been studied extensively, examples of which are the antifibrinolytic agent trans-4(aminomethyl) cyclohexane carboxylic acid, p-benzylamine sulfonic acid and linear aliphatic ligands like Nα-acetyl-L-Lysine and ε-aminocaproic acid (Figure 3).

Chemical modification of plasminogen with tetranitromethane has revealed that Tyr43 and Tyr52 are not essential for the maintenance of the ω-aminocarboxylic acid binding site of the protein [22]. Crosslinking of Lys 38 and Tyr43 shows that although these two residues are not directly involved in ligand binding, they are probably close to this site. It has also been shown that the core structure of the kringle-fold and the lysine binding site are unaltered when the disulfide bridge between Cys4 and Cys81 is selectively reduced and alkylated. Modification of kringle 4 with 1,2-cyclohexanedione or 1-ethyl-3-(3-dimethyl aminopropyl)-carbodiamide has shown that Arg72 and Asp59 are essential for ligand binding [9]; the complementary charges on these residues provide the dipolar ends necessary for the



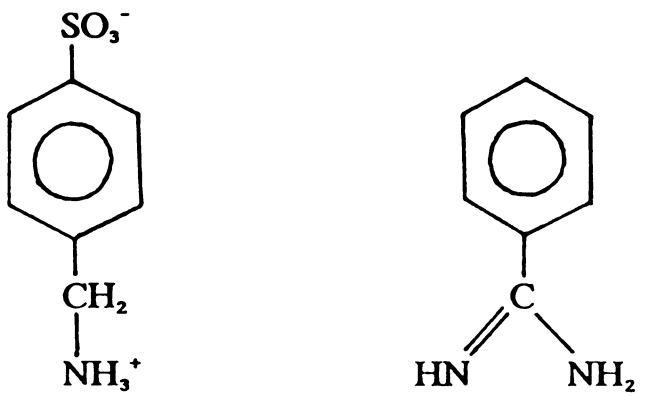


Figure 3. **Kringle ligands**
[Abbreviations used: εACA, ε-aminocaproic acid; AcLys, N^{α}-acetyl-L-lysine; AMCHA, trans-aminomethyl(cyclohexane)carboxylic acid; BASA, p-benzylamine sulfonic acid]

ligand binding [9]. Highfield NMR experiments carried out on kringles 1 and 4 have shown that aromatic residues are significant components of ligand binding probably due to the lipophilic interaction between the ligand and the aromatic side chains [21,23-27]. One of the aromatic residues which has been shown to be involved in the binding is Trp73 [28] with Trp63 and Phe65 probably close to this site.

Based on the three dimensional structure of the kringle of prothrombin fragment 1 [29,30] and NMR experiments, a model has been proposed for the lysine binding site of kringle 4 of plasminogen [31,32]. According to this model, the central loop Cys53 to Cys76 supports or carries the binding site. The binding site can be defined as a relatively open depression that extends from the Pro56-Lys60 peptide stretch which contains Asp57 and Asp59 to a short segment where Arg72 neighbors His31 (Figure 4). The depression is lined by Tyr75 and by Trp63 near the Asp57-Asp59 stretch. The side chain of Trp63 is also close to the Trp73 aromatic ring which interacts with another aromatic residue, Phe65. The three side chains, Trp63, Phe65, and Trp73, thus expose a lipophilic surface which separates the protruding anionic and cationic centers defined by residues Asp57/Asp59 and Arg71, respectively. The two ionic centers thus form complementary polar ends at a distance of about 6.8 Å. It is also apparent that the lysine binding site interacts with a number of other aromatic residues, such as His34, Tyr43, and Trp28, which neighbor the Leu48 side chain. The side chain of Leu48 has been shown earlier to

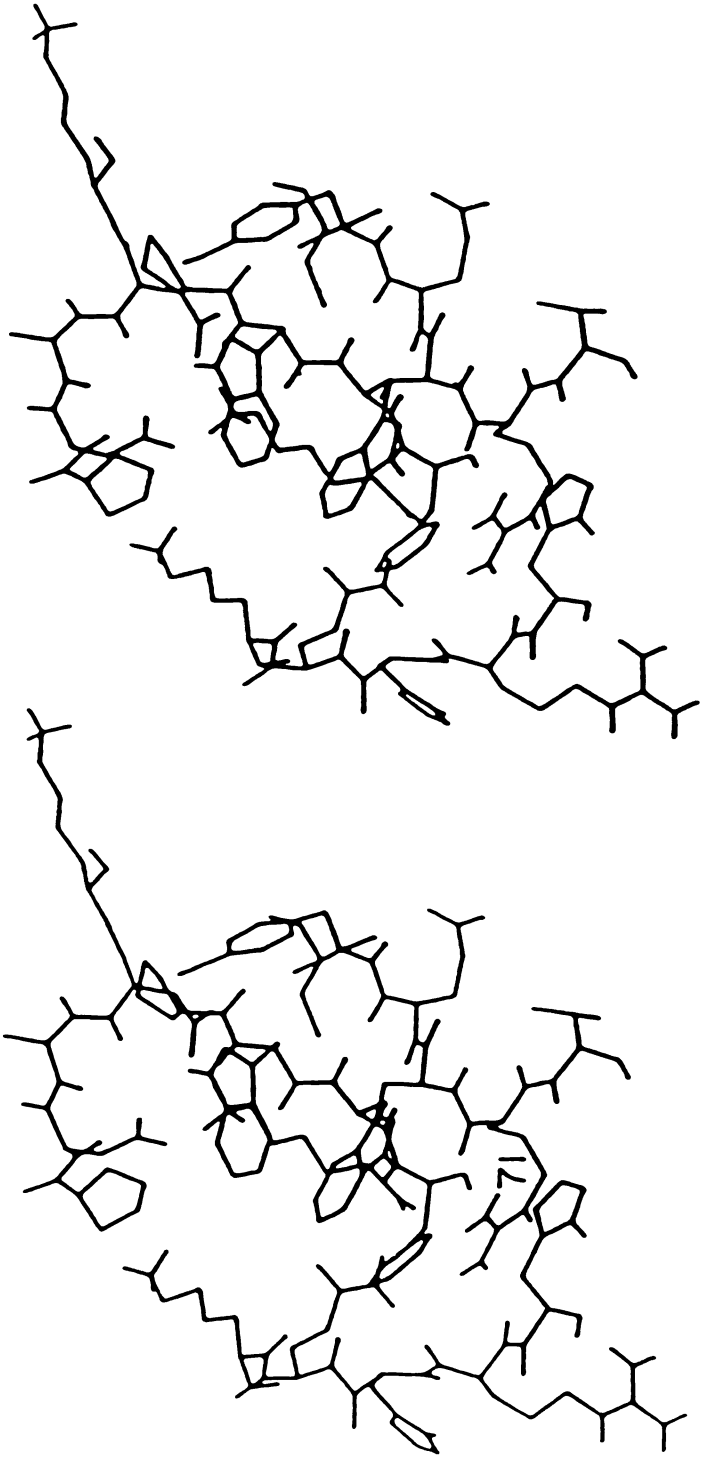


Figure 4. Stereoview of modeled energy-minimized lysine binding site of plasminogen kringle 4 (taken from [30]).

lie close to the binding site [33]. The model of the binding site given above is relatively open and exposed allowing for facile ligand access.

The lysine binding site on kringle 1 of plasminogen, on the other hand, may have a different arrangement of the aromatic residues. While in the case of kringle 4 all the residues implicated in lysine binding are present in the inner loop 51 to 75, there is evidence for the participation of Phe36 in kringle 1, which is extraneous to this loop [11,34]. Chemical modification of Arg34 with 1,2-cyclohexanedione abolishes the binding capability of kringle 1 to ω-aminocarboxylic acids (Arg34 immediately preceeds Phe36 in the primary sequence due to a deletion). On the other hand, glycosylation of Asn34 in kringle 4 of chicken plasminogen seems to have no effect on ligand binding [35]. A model has also been proposed for the lysine binding site of the kringle 1 based on the kringle structure of prothrombin fragment 1 , NMR experiments and energy minimization [31,34].

The initial aim of the present study was to determine the three dimensional structure of the kringle 4 of human plasminogen by the molecular replacement method, using as a model the known structure of the kringle of prothrombin fragment 1, to establish the exact nature of lysine binding site. The diffusion of small molecular substrates to the site in crystals will then provide a precise model for the interaction of plasminogen and fibrin and the role kringles, in general, play in biological recognition. Since

apolipoprotein(a) carries 37 copies of the kringle 4 of plasminogen and since the amino acid homology of some is close to 85%, it should be possible to confidently model such kringles based on the kringle 4 structure determined here. However, since the onset of the work a fortunate unexpected development has broadened the scope of the project considerably. This development was the crystallization of kringle 4 in two different crystal modifications, one of which contained two molecules or a "dimer" of kringle 4 in the asymmetric unit. This structure might also prove to be an excellent approximation and basis for discussing all-important kringle-kringle interactions of blood proteins.

II MOLECULAR REPLACEMENT METHOD

The large and continually growing number of three dimensional protein structures that have been determined (approximately 350 so far) increases the possibility that some new protein being investigated will have features in common with one or more of the known structures. In addition, there is growing interest among many biochemists and protein crystallographers in determining the changes in protein folding and/or side chain structures caused by site specific modifications of the amino acid sequence. In these cases, the application of the de novo multiple isomorphous replacement technique [36] to determine the phase angles of the diffraction pattern may not be the fastest or most economical way to achieve the ultimate goal. A more direct approach is to utilize information from a closely related protein of known structure by application of the molecular replacement technique [37]. If successful, this approach can dramatically decrease the time required to determine a protein structure and can render heavy-atom derivative preparation unnecessary.

The term molecular replacement method has been used to describe several techniques. These can be classified broadly into two categories. In the first, involving more than one copy per asymmetric unit, the orientation of non-crystallographic symmetry elements is searched for within an asymmetric unit of the crystallographic unit cell. This kind of molecular replacement has many applications. It can

be used in combination with multiple isomorphous replacement to improve the refinement of heavy atom parameters [38]. The method has also been used to improve the quality of an electron density map by averaging around the local symmetry elements. These techniques have been critical in the evaluation of virus structures, which generally contain many copies of a subunit and are initially poorly phased [39-46].

The second category of molecular replacement methods involves an attempt to model an unknown structure by orienting and positioning a closely related molecule of known structure in the unit cell of the unknown. These are in general six-dimensional searches comprising the determination of three orientation angles and three positional parameters. In practice, the six dimensional problem is evaluated in two three dimensional parts: angles are determined using a 'rotation function' and positional parameters with a 'translation function' or some other positioning technique. Most of the methods mentioned below apply to macromolecular structures but they are equally valid for small molecules as well.

The occurrence of many identical structures not related by the crystal symmetry has given rise to the concept of "non-crystallographic symmetry". Crystallographic and non-crystallographic symmetries are differentiated by the property that an operator applies throughout the whole infinite crystal for the former, whereas a non-crystallographic symmetry element relates only to a localized volume within the crystal. The difference between

the two classes of the symmetry elements is shown in Figure 5.

The molecular replacement method essentially consists of three stages. These are :

a). Determination of the relative orientation of the independent molecules or subunits of molecules within one crystal or subunits of molecules within one crystal lattice or between different crystal forms. This may be accomplished by analyzing the convoluted Patterson function of the unknown structure and the closely related known model protein. In case of more than one molecule in the asymmetric unit of the crystal, the relative orientation between the crystallographically independent molecules can be determined by an analysis of the Patterson function of the protein convoluted with itself. In practice, this is achieved by calculating a "rotation function" [47], to be described below.

b). Once the relative orientation of the protein molecule is known with respect to a closely related model, the next stage is to position the molecule in the unit cell in the correct orientation. This is accomplished by calculating a special function of the Patterson function called "translation function" [48,49]. At the completion of stage a and b, the exact equivalence between point X′ in the unknown structure and the point X in the model will be known and can be expressed by

$$X' = [C]X + d \quad \ldots.(1)$$

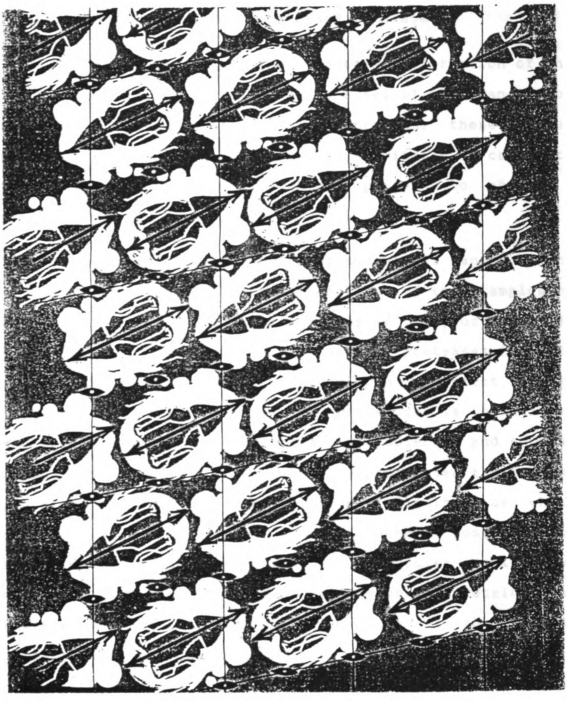


Figure 5. Crystallographic (perpendicular to the page) and non-crystallographic (in the plane of the page) symmetry operations (taken from [37]).

where [C] is the rotation matrix determined in stage a and d is the translation vector determined in stage b.

c). In the third stage, the structural information of the model protein is used to calculate a set of phase angles of structural amplitudes of the unknown protein; these phases are then combined with the measured amplitudes to calculate the electron density of the unknown protein to obtain a trial structure.

There are numerous examples of successful applications of the molecular replacement methods some of the examples of which are various virus structures [39-46], lysozyme [50], insulin [51], myoglobin [52], serine proteases [53-55], phospholipase A2 [56], immunoglobin fragment [57], phycocyanin [58], etc. The procedure is relatively straightforward in the case of a rigid molecule and success depends on how well the model approximates the unknown protein. Proteins that are relatively flexible or which deviate considerably from spherical shape pose special difficulties in the application of molecular replacement method. Even in the case of rigid globular proteins, the success depends on the careful choice of the various parameters in the calculation.

A. PATTERSON FUNCTION

Since the concept of both the rotation function and the translation function involves the Patterson function, a brief description of the latter follows. The Patterson

function is a Fourier series for which only the indices of diffraction and the intensity of the diffracted beam are required as coefficients, quantities that are directly obtained from the experiment [59]. It is given by

$$P(uvw) = (1/V) \Sigma |F|^2 \mathrm{Cos}(2\pi(hu+kv+lw)) \quad ....(2)$$

where V is the volume of the unit cell, $|F|^2$ is proportional to the intensity of the diffracted beam corrected for absorption and other factors, and u, v, w are the coordinates of a point within the unit cell. No phase information is required for this function as no origin for the unit cell is implied, only the relative positions of the atoms.

The Patterson map is the sum of the appearances of the structure when viewed from each atom. A peak in the map, at a position related to the origin of the map by a certain vector, implies that at least one position of that particular vector in the corresponding structure has atomic positions at both ends. If there are many pairs of atomic positions related to a particular peak in the function or if there are only a few but the atoms involved have many electrons, then the function will have a relatively large peak height at that position. The heights of the peaks in the Patterson function are approximately proportional to the product $Z_i Z_j$, where Z_i is the atomic number of an atom at one end of the vector and Z_j that of the atom at the other end. A difference Patterson between a native and an isomorphous heavy atom derivative of a protein crystal

yields valuable information in the form of the positions of the heavy atom which may then be used to determine the phase angles of all the diffracted beams, as in the multiple isomorphous replacement method.

Patterson space can be thought of as made up of two types of vectors: intramolecular vectors and intermolecular vectors. The former are vectors that are characteristic of the distance between atoms within one molecule. These are the vectors that are important in rotation function calculation. The latter are vectors that reflect the distance between atoms in different molecules, whether symmetry related or independent. The translation function is calculated based on these vectors.

B. ROTATION FUNCTION

The rotation function can be defined as an integral over a spherical region of the product of the given Patterson function P_1 with rotated version of either itself or the Patterson P_2 of a closely related model protein [60]:

$$R(C) = \int P_1(x).U(x).P_2(x)\, dv \quad ....(3)$$

where C is a variable rotation operator and U is a shape function that is used to limit the evaluation of the function to a smaller spherical region centered on the origin of the Patterson space. In general, such a region will be rich in intramolecular Patterson vectors, but will contain relatively few intermolecular Patterson vectors.

The rotation function thus measures the degree of coincidence between the two Patterson functions at various values of the rotation matrix C. At certain value of the matrix, this rotation function will have a large value; this rotation will correspond to a transformation where the vector set relating electrons within the model coincides with the corresponding vector set from the unknown molecule.

C. ROSSMANN AND BLOW ROTATION FUNCTION

A number of methods have been proposed for the evaluation of the rotation function, either in reciprocal space or in direct space. Rossmann and Blow expanded the rotation function in the form of a Fourier series [47]:

$$R(C) = \Sigma\Sigma\, F_2^{\,2}(p).F_1^{\,2}(h).G(h+h') \quad ....(4)$$

where $F_1^{\,2}(h)$ and $F_2^{\,2}(p)$ represent the intensities of the crystal and the model, h' is a non-integral reciprocal lattice point created by the operation of rotation matrix C' on h, C' is the transpose of the matrix C, and G is an interference function which is the transform of U. The quantities h' and p are related by the expression :

$$h' = Cp \quad ....(5)$$

This method, though historically very important as it was the first attempt to evaluate the rotation function, has very little practical application. It is extremely cumbersome and lengthy to evaluate. The length and expense of the computation usually forces the use of only the most

intense reflections in the calculation. Furthermore, the interpolation function G is usually truncated to speed up the process; this introduces serious errors in the rotation function results.

D. CROWTHER FAST ROTATION FUNCTION

Since rotation functions deal with the rotation of spherical volume of Patterson density, a natural form for the Patterson function would be an expansion in terms of spherical harmonics and spherical Bessel functions, as suggested by Crowther and Blow [49]. By using expansion functions appropriate to the rotation group, the rotation function can be broken down into two parts:

$$R(C) = C_{lmn'} D_{m'm}(C) \quad ....(6)$$

where the coefficients $C_{lmn'}$ refer to a particular pair of Patterson densities and are independent of the rotation part. The coefficients, $D_{m'm}(C)$, which contain the whole rotational part of the problem, refer to rotations of spherical harmonics and are independent of the particular Patterson densities.

This formulation of the rotation function has been coded into a computer program called the Fast Rotation Function and is one of the most widely used rotation programs today. It is much more accurate than the rotation program of Rossmann and Blow and at least 100 times faster since the summation is reduced to a series of two

dimensional Fourier series that can be evaluated with fast Fourier transform algorithms. Moreover, any amount of data can generally be used. However, the program has some limitations. There is a limit to the maximum radius of integration that be used for a given data set resolution. Sampling of rotation space is also restricted by the symmetry of the Patterson of the model and the unknown. Another potential problem is with respect to monoclinic space groups, where the program reindexes the data to align the crystal system to the reference coordinate system.

E. LATTMAN ROTATION FUNCTION

Lattman used the property that the Patterson function of a model placed in a sufficiently large artificial cell is self bounded to eliminate the shape function U from the rotation function which then takes the form [60]:

$$R(C) = \int I_1(hkl).C.I_2(h'k'l') \quad ....(7)$$

where $I_1(hkl)$ is the intensity of the unknown structure, $I_2(h'k'l')$ is the squared transform of the model and C is a rotation matrix. When R is large, the matrix C will rotate the model structure into alignment with some element of the unknown structure. In general, the rotated reciprocal point h'k'l' will not lie near the unrotated reciprocal point hkl. This problem is resolved by using a very finely sampled transform of the known structure. No hkl will then be far from a rotated h'k'l' and the value of I_2 at hkl can be

obtained by interpolation from the eight nearest points of rotated h'k'l'.

The greatest difference between Crowther's and Lattman's formulation of the rotation function is that there is no radius of integration limit in the latter, a limit that is present in the former due to mathematics of the formulation [61]. Lattman's version is also more flexible, especially in regard to the sampling of rotation space. The sampling can be as fine as the user desires. Crowther's program, on the other hand, samples the angles alpha and gamma in a specific way depending on the symmetry of the unit cell of the unknown and the model ; this limits the precision with which the positions of the peak maxima can be determined. The disadvantage of Lattman's method is that the program is slow. The strategy that one usually follows is to determine the approximate positions of the peak maxima using Crowther's program and use Lattman's program, sampling on a fine grid around those maxima, to determine the precise location of the maxima.

F. ROTATION FUNCTIONS IN DIRECT SPACE

Another variation of the rotation function is using the vectors in the model Patterson limited to the relatively large peak heights. This requires carrying out the calculation in direct space [62,63]. There are many other methods which use the same principles and evaluate the rotation function in direct space. The advantages are that

one can modify the search vector set in physically reasonable ways to enhance the search procedure.

In the case of macromolecular structures, the number of vectors required depends on the number of atoms in the model and the unknown. For a typical protein, with 150 residues, choosing the 3000 highest peaks from the Patterson has been shown to be more than adequate [55]. The Patterson vectors are then rotated and interpolated to the nearest grid in the Patterson of the unknown molecule and a product function calculated. This process is repeated for each set of the angles. The angles or the rotation matrix corresponding to the highest value of the product function corresponds to the transformation needed to bring the model and unknown to same orientation. The SEARCH routine in the program "PROTEIN" [64], makes use of this principle. Other major advantages of the SEARCH program is that very fine sampling in rotation space is possible and peaks close to the origin can be removed.

G. PRACTICAL ASPECTS OF ROTATION FUNCTION CALCULATIONS

The matrix C is usually expressed in one of two systems of angular variables. One way is in terms of the three Eulerian angles $\theta_1, \theta_2, \theta_3$, as described by Goldstein [65]. This is illustrated in Figure 6. The rotation operation consists of (a) a rotation of Cartesian axes by θ_1 about the Z axis; (b) a rotation of θ_2 about the new position of the X axis; (c) a rotation θ_3 about the new position of the Z

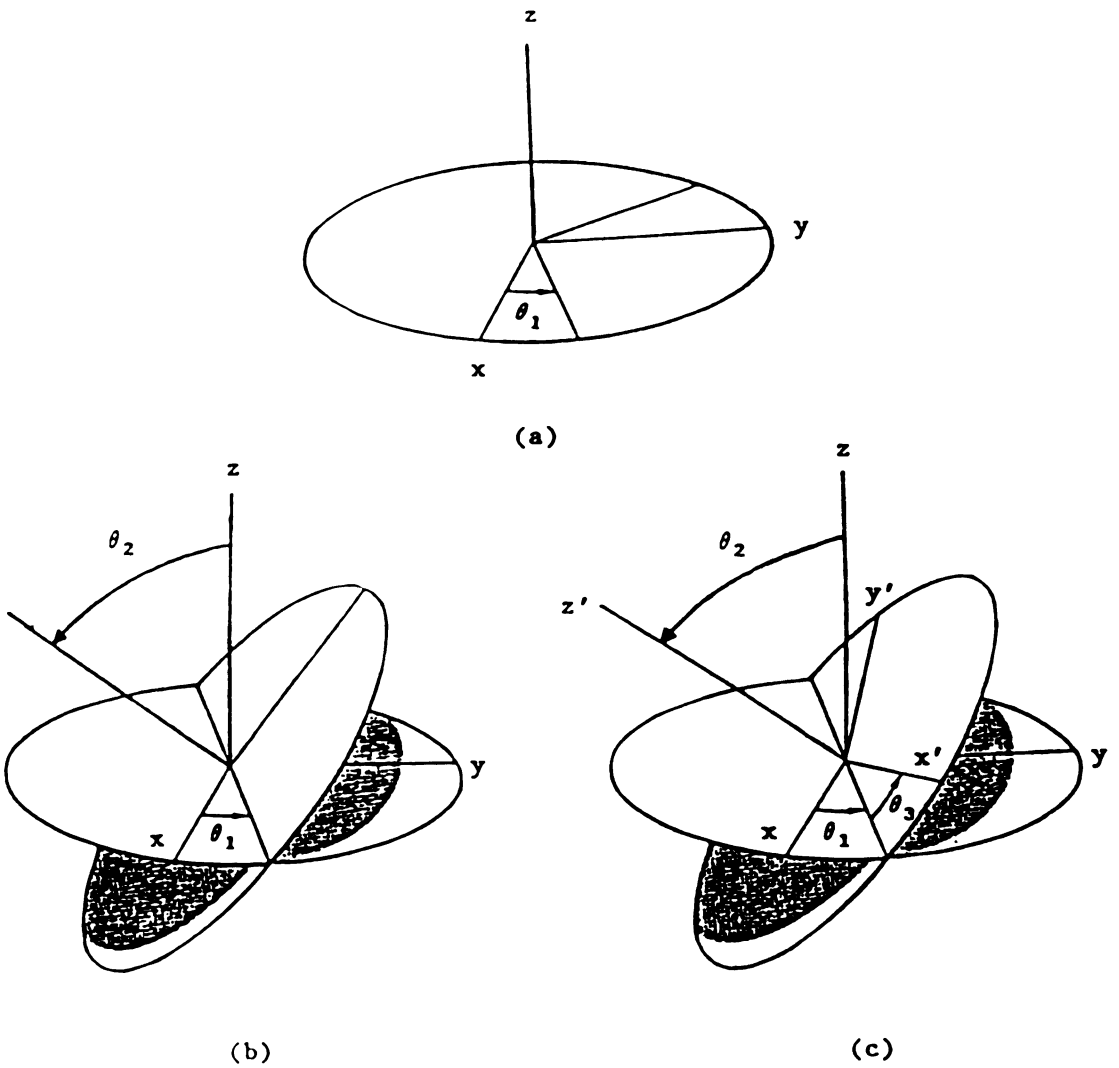


Figure 6. The rotations defining the Eulerian angles θ_1, θ_2, and θ_3 (taken from [65]).

axis. These angles are positive if they are clockwise when viewed along the relevant axes towards the origin, as in the usual right-handed convention. This system of angle definition is difficult to visualize; however, it reveals the symmetry of the rotation function in a convenient way [47]. The rotation matrix in terms of the Eulerian angles is given in Table 2.

The second angular system makes use of the theorem that an arbitrary rotation can be accomplished by an appropriate spin about a properly chosen axis [66]. The spherical polar angles ϕ and ψ specify the longitude and co-latitude of this axis, and the azimuthal angle χ, the spin about it. This is illustrated in Figure 7. In this representation, it is easy to visualize the direction and order of rotation axis with respect to the crystallographic unit cell. This is the preferred system when the rotation matrix relates two molecules in one asymmetric unit of the unit cell.

In eqn. (3), if P_1 and P_2 refer to the Patterson of the same crystal, then the resulting rotation function is called the self-rotation function and if P_2 represents the Patterson of the model, the rotation function is called the cross-rotation function. Thus the self-rotation function describes the relative orientation of two or more molecules within one asymmetric unit of a crystallographic unit cell. Cross-rotation,on the other hand, relates the orientation of the model and the unknown or one crystal structure to another.

Table 2. Rotation matrix in terms of Eulerian angles $\theta_1, \theta_2, \theta_3$.

$-\sin\theta_1 \cos\theta_2 \sin\theta_3 + \cos\theta_1 \cos\theta_3$	$\cos\theta_1 \cos\theta_2 \sin\theta_3 + \sin\theta_1 \cos\theta_3$	$\sin\theta_2 \sin\theta_3$
$-\sin\theta_1 \cos\theta_2 \cos\theta_3 - \cos\theta_1 \sin\theta_3$	$\cos\theta_1 \cos\theta_2 \cos\theta_3 - \sin\theta_1 \sin\theta_3$	$\sin\theta_2 \cos\theta_3$
$\sin\theta_1 \sin\theta_2$	$-\cos\theta_1 \sin\theta_2$	$\cos\theta_2$

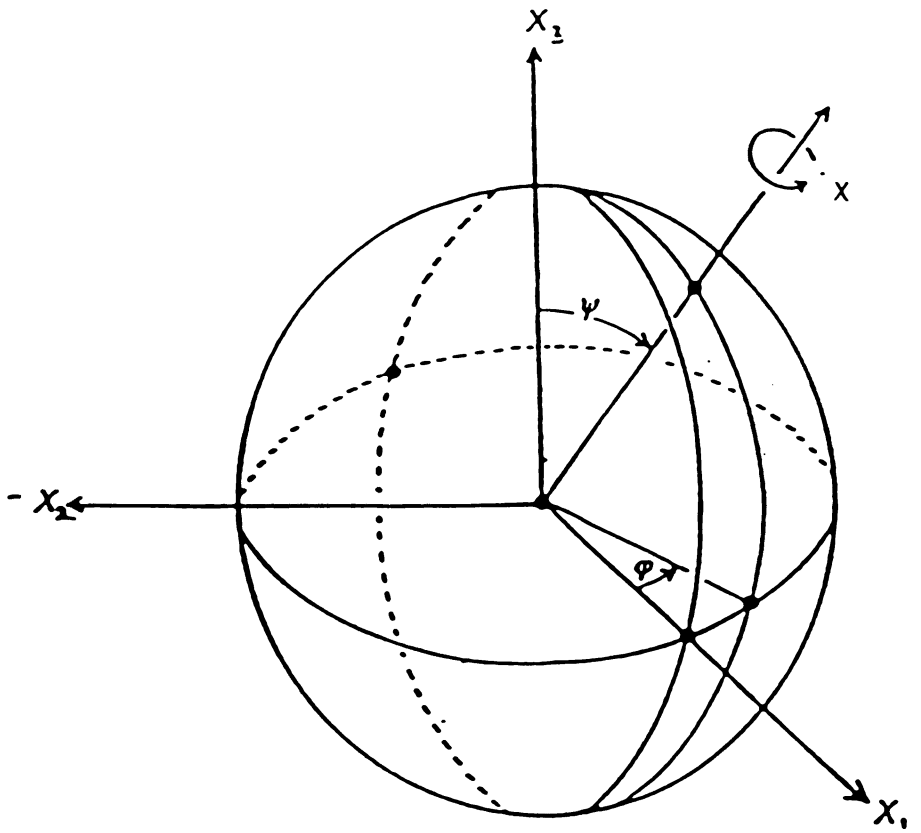


Figure 7. Spherical polar angles ϕ, ψ, and χ with respect to crystallographic axes.

Since the origin of a Patterson function is always an enormous number being the sum of all vectors and since it does not contain much useful information as far as the atomic arrangement is concerned, many of the rotation programs have a provision to remove the origin. Removing the origin of the Patterson function helps to increase the signal to noise ratio of the true solutions. The volume over which the integration is carried out depends on the molecular size of the model or the size of the unknown protein, predicted by some experiments like radius of gyration experiments. Usually, the upperlimit on Patterson vectors is chosen so as to include as many of the intramolecular vectors as possible and to cut off the unwanted intermolecular vectors. This is required by the fact that in the crystal lattice there are thousands of molecules and the measured intensity corresponds to all these molecules. The rotation function, on the other hand, assumes that the vectors in the Patterson function represent the atomic arrangement within a single molecule. For a typical protein of dimension 25 x 25 x 25 Å, the best results are obtained if only Patterson vectors between 4 and 20 Å are used.

Another parameter in rotation function calculation is the resolution of the data used to calculate the function. The low order reflections usually have a large contribution from the solvent which may be either disordered or even if ordered may be different in the model and the unknown protein. Very high angle reflections , on the other hand,

contain fine details of the structure which again may be different in the model and the unknown. It has been found that using reflections between 10.0 Å and 3.5 Å gives the best result and good signal to noise ratio for many protein molecules. Rejecting reflections based on individual statistics and intensities is also found to enhance the overall success of the method.

The first step in a cross-rotation function calculation is to calculate a set of structure factors, to a suitable resolution, of the model in a triclinic unit cell. The size of the cell chosen is two to three times the size of the model molecule. The choice of an extra large triclinic unit cell produces a calculated Patterson function of the model that corresponds to a single molecule. The number of atoms of the model that are used to calculate the structure factors depends on the extent of homology of the primary sequences between the model and the unknown protein and how well the side chains of the model are defined in the electron density. The use of only the main chain protein atoms and the conserved side chains gives good results. The thermal parameters of the model are adjusted so as to correspond to the average fall off in intensity of the reflections of the unknown protein.

A model Patterson is then computed with the calculated structure factors. It is this Patterson which is rotated onto the Patterson of the unknown. There are a number of parameters which have to be optimized in order to get a reasonably good solution. These parameters may be different

for different proteins and different crystal systems. Initially, one proceeds by sampling on a coarse grid in rotation space. Rotation peaks are generally quite broad and it has been found that sampling at five degree increments is more than adequate. This is followed by a fine sampling around the approximate solution obtained in the first step.

The cross-rotation results are best expressed in terms of the Eulerian angles $\theta_1,\theta_2,\theta_3$ for this allows one to see the symmetry in the rotation function clearly. Part of the symmetry in the rotation function arises from the definition of the Eulerian angles themselves; the crystal symmetry of the unknown also contributes to it [67,68]. The symmetry expected in the rotation function when a triclinic Patterson is rotated onto various crystal systems and the size of the asymmetric unit for the rotation function calculation is listed in Table 3. Instead of expressing the results in $\theta_1,\theta_2,\theta_3$, frequently another set of angles is used with a simple relationship between the two sets of angles:

$$\theta_1 = \alpha + 90 \; ; \; \theta_2 = \beta \; ; \; \theta_3 = \gamma - 90 \quad ....(8)$$

H. TRANSLATION FUNCTION

The second stage in molecular replacement method is the determination of the position of molecule with respect to the crystallographic symmetry elements. Several different methods have been proposed for this stage of the molecular replacement method, which is usually referred to as the

Table 3. Symmetry of rotation function when a model triclinic Patterson is rotated onto the Patterson of a test molecule with various symmetries.

CRYSTAL LATTICE	PATTERSON SYMMETRY	SYMMETRY OF ROTATION FUNCTION	ASYMMETRIC UNIT FOR ROT. FUN.
Triclinic	1	$\theta_1, \theta_2, \theta_3$ $\pi+\theta_1, -\theta_2, \pi+\theta_3$	θ_1 : 0 - 360 θ_2 : 0 - 360 θ_3 : 0 - 180
Monoclinic (b axis unique)	2/m	$\theta_1, \theta_2, \theta_3$ $\pi+\theta_1, -\theta_2, \pi+\theta_3$ $\pi-\theta_1, \pi+\theta_2, \theta_3$ $-\theta_1, \pi-\theta_2, \pi+\theta_3$	θ_1 : 0 - 360 θ_2 : 0 - 180 θ_3 : 0 - 180
Monoclinic (c axis unique)	2/m	$\theta_1, \theta_2, \theta_3$ $\pi+\theta_1, -\theta_2, \pi+\theta_3$ $\pi+\theta_1, \theta_2, \theta_3$ $\theta_1, -\theta_2, \pi+\theta_3$	θ_1 : 0 - 180 θ_2 : 0 - 360 θ_3 : 0 - 180
Orthorhombic	mmm	$\theta_1, \theta_2, \theta_3$ $\pi+\theta_1, \theta_2, \theta_3$ $\pi-\theta_1, \pi+\theta_2, \theta_3$ $-\theta_1, \pi+\theta_2, \theta_3$ $\pi-\theta_1, \pi-\theta_2, \pi+\theta_3$ $-\theta_1, \pi-\theta_2, \pi+\theta_3$ $\theta_1, -\theta_2, \pi+\theta_3$ $\pi+\theta_1, -\theta_2, \pi+\theta_3$	θ_1 : 0 - 180 θ_2 : 0 - 180 θ_3 : 0 - 180

translation function. In some of the earlier methods, the position of the molecule is independently determined relative to each of the space-group symmetry elements either in reciprocal space or in direct space [49,62,69-71]. This affords the possibility of reducing the dimensionality of the problem from three to one or two dimensions. A more direct but computationally more demanding approach is a search for a minimum of an agreement index or correlation coefficient between the observed structure factor amplitudes and those calculated from the model [72-78].

I. CROWTHER AND BLOW TRANSLATION FUNCTION

One of the earliest methods, which is widely used even now, is a formulation due to Crowther and Blow [49]. Consider the problem of positioning a known molecule relative to a particular crystallographic symmetry element.The term 'known molecule' here implies that not only do we know the atomic coordinates relative to some local region, O, fixed in the molecule, but also that the molecule is in the same orientation as one of the molecules in the unknown crystal structure. We wish to determine the position of the local origin with respect to the chosen crystallographic symmetry element. If the known molecule is placed at an arbitrary position in the unit cell, the position of symmetry related molecule is fixed and it is possible to calculate the set of Patterson vectors from the known molecule to the symmetry related molecule. As the position of the known molecule varies, the relative

configuration of these Patterson vector sets is unchanged but the set moves bodily to a position characterized by the vector joining the local origins of the known molecule and the symmetry related molecule. This set of vectors are referred to as the cross-Patterson of the model structure and is a function of two variables, namely position in Patterson space and also position of the local origin of the known molecule in the model structure.

A well known method of finding the position of the local origin is to move this set of cross-Patterson vectors over the observed Patterson function of the crystal, using a mimimum function or similar measure of fit to determine the correct solution. In the case of molecules with thousands of atoms as in proteins, this can be achieved by the convolution of the observed Patterson function with the cross-Patterson function set of the model structure. This convolution may be carried out by multiplying the Fourier coefficients of the two functions and performing a Fourier summation with the coefficients so obtained. The resulting function should have a prominent peak corresponding to the vector set between the two origins. Mathematically the translation function can be defined as :

$$T(t) = \Sigma |Fobs(h)|^2 . F_M(h) . F_M^*(hA) . \exp(-2\pi i . h . t) \quad ....(9)$$

where $|Fobs\ (h)|$ is the observed amplitude, $F_M(h)$ is the structure factor of the known molecule calculated relative to the local origin, $F_M^*(hA)$, the complex conjucate of the strucutre factor amplitude of a symmetry related molecule,

related by the symmetry operator A, and t is the translation vector. In crystals having n crystallographic symmetry elements, there will be n such expressions for translation function, each solution representing the position of a molecule with respect to one symmetry element; thus the position of the known molecule can be overdetermined with respect to the symmetry elements.

The observed Patterson function of the crystal contains both intramolecular and intermolecular Patterson verctors. In translation function determination, unlike in the rotation function, we are interested in fitting the intermolecular vectors ; the intramolecular vectors in fact only serve to increase the background noise in the summation. However, since the molecular structure is known, it is possible to remove the intramolecular vectors from the observed Patterson function, provided the observed intensities can be put on an appropriate scale. Suppose that there are n molecules in the unit cell and that the ith molecule is related to the known molecule by a symmetry operator, whose rotational part is represented by A , where A , the matrix relating the known molecule to itself, is the identity matrix. Then the modified translation function can be expressed as :

$$T_M(t) = \Sigma(|F_{obs}(h)|^2 - \Sigma |F_M(hAi)|^2) * F_M(h).F_M^*(hA).\exp(-2\pi i.h.t) \quad ....(10)$$

where F_{obs} and F_M are assumed to be on an absolute scale.

The translation function results invariably appear on a Harker section or on an axis. This can be illustrated by an example. Let us take the space group $P2_12_12_1$ with one molecule in the asymmetric unit. There are 3 non-intersecting 2_1 screw axes in this space group and the corresponding equivalent positions are

Molecule 1 :	x	y	z
Molecule 2 :	1/2-x	-y	1/2+z
Molecule 3 :	1/2+x	1/2-y	-z
Molecule 4 :	-x	1/2+y	1/2-z

The independent intermolecular translation vectors are then

Mol. 2 to Mol. 1 :	2x-1/2	2y	1/2
Mol. 3 to Mol. 1 :	1/2	2y-1/2	2z
Mol. 4 to Mol. 1 :	2x	1/2	2z-1/2

.....(11)

Each of the vectors is required to lie on a Harker section. Calculating all the three Harker sections will give multiple determinations of each value and these should all be consistent. In $P2_12_12_1$, there is an eightfold ambiquity in the choice of the origin, with each value being equally well expressed as T or T + 0.5. However, this problem arises only if we have more than one molecule in the asymmetric unit of the unit cell.

One of the modifications that have been suggested to the translation function of eqn. (1) is due to Langs [70]. In his formulation, the terms $F_M(h)$ and $F_M^*(h)$ which contain both amplitude and phase components of the model structure

factor, are replaced by terms containing only the phase component. The magnitude of spurious peaks produced by this modified translation function is almost half of that produced by original translation function.

J. R VALUE MAPPING

Another commonly used procedure in place of the translation function is to calculate the R factor :

$$R = \Sigma (||Fo| - |Fc||) / \Sigma |Fo| \quad ....(12)$$

where |Fo| and |Fc| are the observed and calculated structure factor amplitudes respectively as the model is translated in the unit cell. One simply calculates R values on a grid covering all potential molecular positions, and investigates those positions that give minimum R values. The computation is not very efficient since the fast Fourier transform (FFT) cannot be used to compute the whole translation function. Another major drawback of this approach is that the method is very sensitive to errors in the relative scale of |Fo| and |Fc|. This is very important when high resolution data are not available and especially if only part of the known structure is used for the translation search ; in such cases an accurate scale factor cannot be determined. Neverthless, many structures have been solved by the method [79].

K. PACKING FUNCTIONS

There is another method for determining the translation function which makes use of the fact that the correct translation solution should result in the fewest unacceptable intermolecular contacts in the crystal lattice or the molecules should not interpenetrate each other in the lattice [80,81]. In this method, one samples the translation space and for each possible position calculates the number of "bad" intermolecular contacts. This is a relatively powerful method for macromolecules, where crystals contain large solvent space and limited regions of intermolecular contact. Bott and Sharma [80] have proposed a modification which speeds up the process considerably. One of the modifications that has been suggested is to abandon a particular solution as untenable when a certain number of "bad" contacts is encountered [81]. Even with the simplification, the packing program is an extremely slow process and the method cannot be used in cases having more than one molecule in the asymmetric unit.

Many of the methods mentioned above have been compiled in the form of a program package called "MERLOT" by Paula Fitzgerald [61]. The Merlot package in addition has many small utilities programs also, such as a program for calculation of the structure factor of a model, refinement of the rotation and translation solutions in Patterson space, etc.

L. BRUTE TRANSLATION FUNCTION

Another of the methods that been suggested recently and have been widely used is a translation function that uses a linear correlation coefficient to determine the correct position of the oriented molecule in the unit cell. The method has been implemented in a computer program called "BRUTE" [77]. In this method a linear correlation coefficient is calculated between two sets of origin removed Patterson functions or intensities :

$$C = \int P_o' . P_c' du \left[\int P_o^2 . du \int P_c^2 du \right]^{-\frac{1}{2}} \quad ....(13)$$

where P_o and P_c represent the origin removed Patterson functions of the unknown and the model. The expression given above can be represented in a form which is a general expression of a correlation coefficient

$$C = \Sigma(|Fo|^2 - |\overline{Fo}|^2).(|Fc|^2 - |\overline{Fc}|^2) * \left[\Sigma(|Fo|^2 - |\overline{Fo}|^2)^2 . \Sigma(|Fc|^2 - |\overline{Fc}|^2)^2 \right]^{-\frac{1}{2}} \quad ....(14)$$

Like the R factor, the correlation coefficient is basically a measure of the agreement between observed ,|Fo|, and calculated quantities, |Fc|. But unlike the R factor, the correlation coefficient is scale insensitive, as replacement of |Fo| by k|Fo|, where k is an arbitrary scale constant, gives the same value for C. The only drawback of the method is it is a real number cruncher calculation and the

correlation coefficient for each solution must be evaluated separately.

In many cases the translation function fails to give a consistent solution. This is often a stumbling block in molecular replacement calculations. Success in this stage usually depends on the accuracy of the rotation results. In case none of the translation function programs mentioned above gives a consistent solution, one goes back to the rotation function calculation, in the hope that a better starting rotation matrix can be obtained.

It is important to refine the orientations and positions obtained from molecular replacement before proceeding with the structure analysis. One technique proposed for doing this involves local minimization of the R value [82]. One varies all independent parameters (angles and translations) in small increments around their current values. R values are calculated for each parameter value. The parameter values are shifted to those that give the lowest R value and used as input to another cycle until convergence is obtained.

Several criteria of the validity of a potential solution have been used. These include: (a) low residual (b) non-interpenetration of molecules (c) consistency between replacement solution and any non-crystallographic symmetry elements that are present (d) phasing power of the solution. The last criterion is by far the most powerful. "Omit maps" are calculated by removing a part of the

structure and analyzed to see whether it returns in subsequent calculations.

III EXPERIMENTAL

The kringle 4 sample from human plasminogen was kindly provided to us by Prof. M. Llinas; it was generated from plasminogen by elastase digestion and affinity chromatography on lysine-Sepharose [6] as described by Winn et. al. [83]. The product is a heterogeneous mixture of approximately 70% C-terminal Ala86 and about 30% C-terminal Ala86-Ser87-Val88. The kringle 4 used for crystallization was in the form of a lyophilized solid stored at freezer temperatures in a dry enviornment.

Kringle 4 crystallizes in two different crystal systems and in some cases, from very similar conditions. One form is monoclinic $P2_1$, with two molecules per asymmetric unit while the other is orthorhombic, spacegroup $P2_12_12_1$, with one molecule per asymmetric unit. The structure of monoclinic kringle 4 with two molecules might be relevant to intact cassette domain structure molecules such as plasminogen. The crystals diffract x-rays well and are relatively stable with respect to degradation upon radiation exposure.

A. CRYSTALLIZATION

Monoclinic kringle 4 crystals were first obtained by Dr. Chang Park in our laboratory [84]. The kringle 4 was dissolved in distilled water at a protein concentration of approximately 10 mg/ml. Tris-saline buffer was added to

adjust the pH to 7.4. Crystals were grown by the hanging drop method using 45% saturated ammonium sulfate (pH 7.4) as the precipitant. Numerous small hair-like rosettes appeared within a week. These small crystals were then used in a repetitive seeding process to obtain large crystals in a sitting drop set up. Crystals with dimensions of 1.0 x 0.2 x 0.2 mm were obtained in this fashion.

Under slightly different conditions, monoclinic crystals were also obtained by Anne Mulichak of our laboratory [84]. Hanging drop experiments were set up with 10 μl drops consisting of 1:1 ratio of well solution and 10 mg/ml of kringle 4, where the well solution was 0.1 M ammonium sulfate, 35% PEG 8000, 0.8% n-butanol at pH 6.2. Needle-like crystals appeared within a few days, sometimes along with thin plates in small clusters, and reached their maximum dimensions in two weeks. The small crystals were later used as seed crystals in a sitting drop set up to produce crystals of dimensions 1.0 x 0.25 x 0.20 mm. These crystals exhibit identical space group symmetry and lattice parameters and similar gross morphology (some crystals display striations) as the crystals obtained by Dr. Park and by a different method (Figure 8). The unit cell parameters and the relevant crystal data are shown in Table 4.

Crystals of kringle 4 belong to the space group $P2_1$. They contain four molecules in the unit cell, with two molecules per asymmetric unit. Assuming a protein specific volume of 0.74 cm^3/gram, a protein fraction of 66% , with

Figure 8. Crystals of Monoclinic kringle 4.

Table 4. Summary of crystal data of Monoclinic plasminogen kringle 4.

Space group	$P2_1$
Num. of molecules per unit cell	4
Molecular weight (daltons)	9819
Crystal size (mm)	1.0 x 0.25 x 0.20
a (Å)	32.75
b (Å)	49.19
c (Å)	46.20
β (deg.)	100.6
Solvent fraction	34%
Matthews number (Vm; $Å^3$/dalton)	1.86
Unique reflections	3785
Num. of observed reflections	2801

V_m = 1.86, was calculated. The very large protein fraction is consistent with the excellent extent of diffraction from relatively small crystals. The small V_m value compares well with other low molecular weight proteins [85].

A characteristic feature of the monoclinic kringle 4 crystals is the tendency to form twinned crystals. The crystals are twinned along the c^* direction in a manner similar to α-chymotrypsin [86]. The relative orientation of the crystal and the twin are shown in Figure 9.

B. DATA COLLECTION

The crystal used for data collection was of dimensions 1.0 x 0.25 x 0.20 mm. The crystal was mounted with the b^* axis along the long axis of the capillary which becomes coincident with the phi axis of the diffractometer when the crystal is aligned. Crystals were mounted in capillary tubes of common glass and diameter 1.5 mm. After a crystal had been oriented properly with drops of mother liquor above and below it, the tube was sealed in a low flame. The crystal is held to the wall of the capillary by the surface tension of a drop of mother liquor.

Preliminary x-ray examination of the crystal included the recording of the diffraction pattern along the three principal axes of the crystal. The principal axial intensity distributions are shown in Figure 10. The axial runouts show that the crystals scatter x-rays well beyond 2.0 Å resolution (2θ=45.3°).

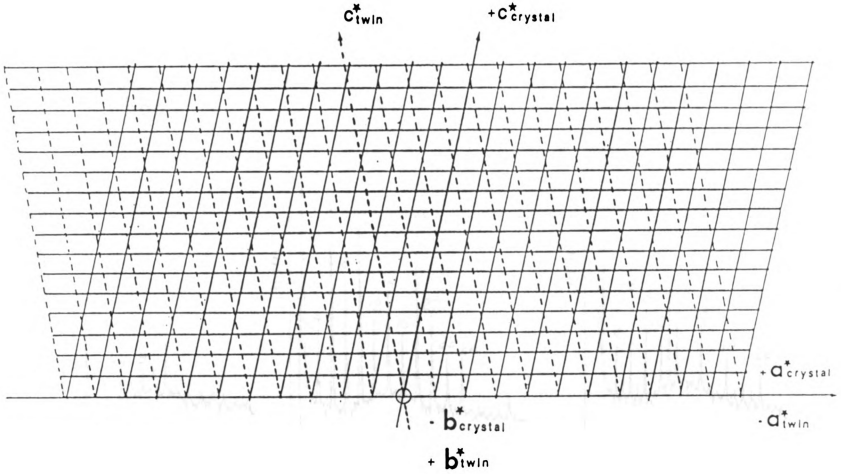


Figure 9. The reciprocal lattices of the crystal and the twin. The crystal c*-axis is shown by the solid line, that of the twin by the dotted line.

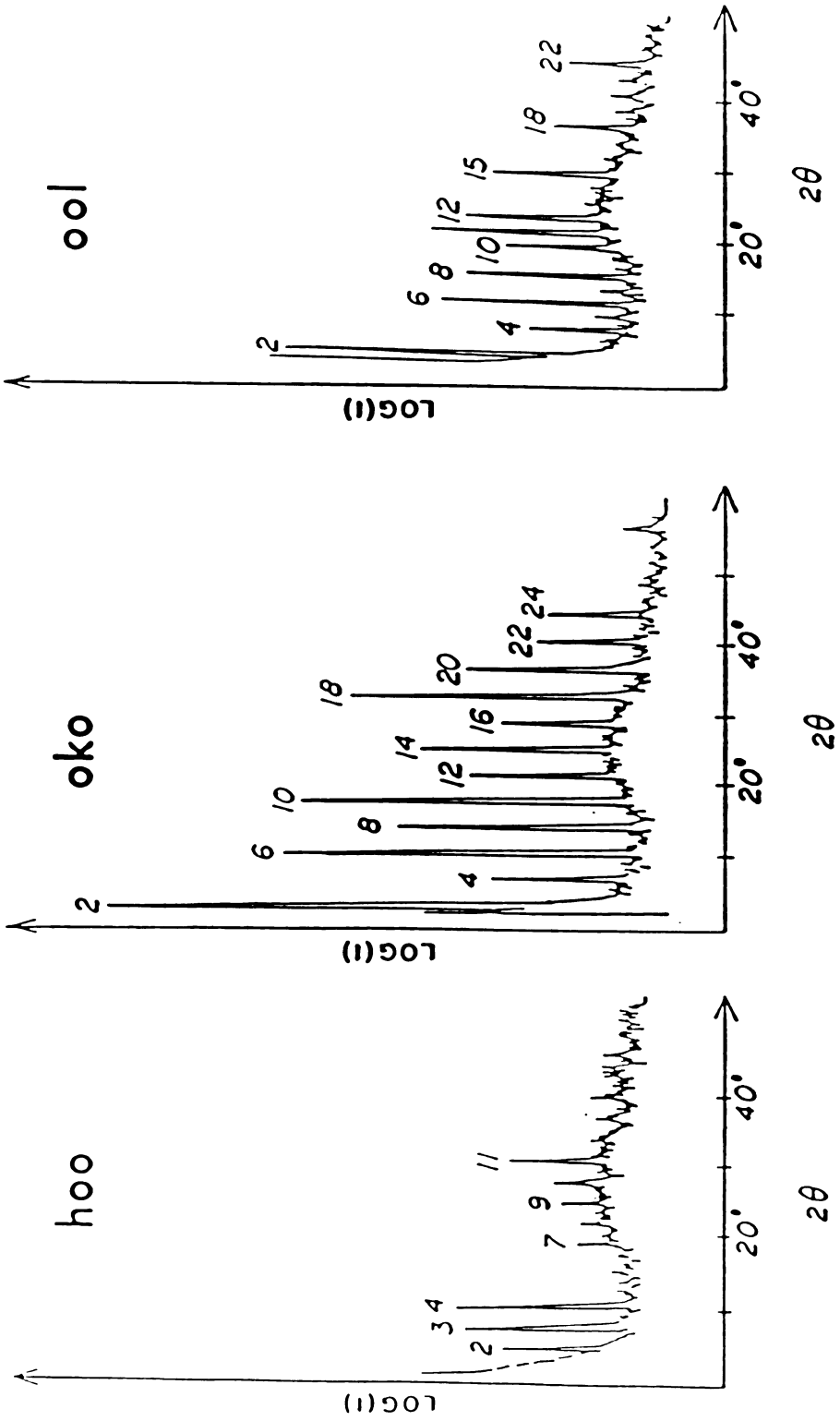


Figure 10. The principal axial intensity distributions of kringle 4 crystal.

Reflection intensities were measured by using Cu Kα radiation (1.5418 Å) with a graphite monochromator on a Nicolet P3/F four-circle diffractometer controlled by a Data General Nova/4 minicomputer. A schematic drawing of the goniostat of this diffractometer is shown in Figure 11. The goniostat orients the crystal and detector with respect to the stationary x-ray beam and brings Bragg planes sequentially into reflection position. The four circles of the goniostat refer to the three circles swept out by rotation of the crystal about the ω, ϕ, and χ axes, and the circle swept out by the arm of the detector through rotation about the phi axis(which is coincident with the axis). To measure the intensity of a diffracted beam, the diffractometer is driven to the appropriate angular positions for each of the diffracted beam to satisfy the Bragg condition :

$$2d \sin \theta = \lambda \qquad ....(15)$$

where d is the interplanar spacing of the reflection and λ is the wavelength of the x-ray radiation.

The $2\theta, \omega, \phi$, and χ angles which the goniostat drove to in the course of making an intensity measurement resulted from the unit cell parameters and the orientation matrix of the crystal. The information about both the orientation and the unit cell parameters of the crystal is contained in the orientation matrix A, which describes the reciprocal vectors a^*, b^*, and c^* in terms of their components in the instrument coordinate system.

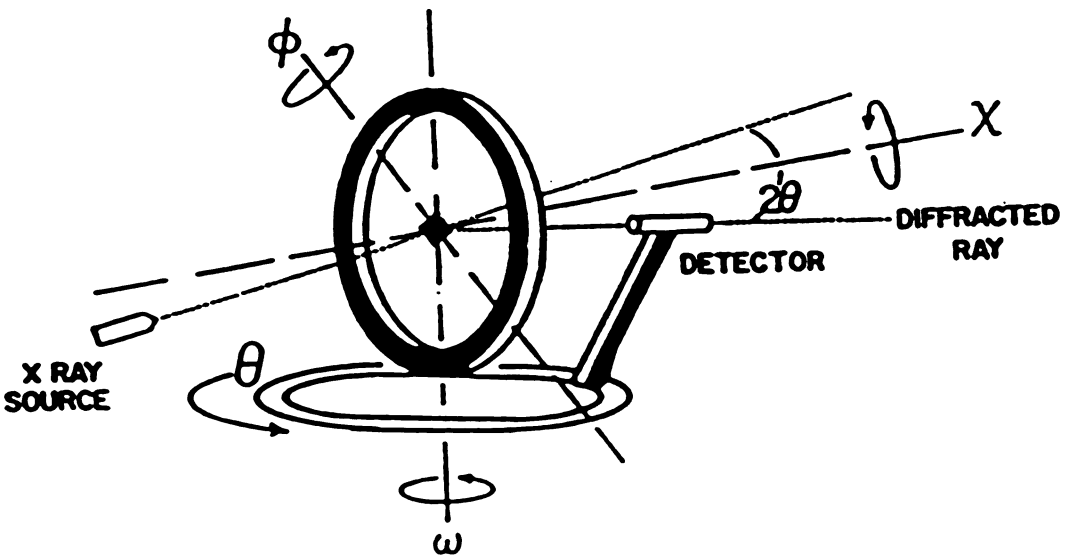


Figure 11. A schematic drawing of the goniostat of a four-circle diffractometer.

$$A = \begin{matrix} a^*x & b^*x & c^*x \\ a^*y & b^*y & c^*y \\ a^*z & b^*z & c^*z \end{matrix} \qquad .....(16)$$

The orientation matrix is first calculated from the indices and angular positions of three strong reflections. Later, a better orientation matrix is obtained by finding another 8 or 9 strong reflections of high 2θ, distributed uniformly in the part of the reciprocal space to be used for intensity measurements. These reflections constitute the 'reflection array'. The unit cell parameters are obtained by performing a least squares calculation between the calculated and observed angles of the reflection in the reflection array. Usually two or three reflections with relatively high χ value are included in the reflection array to insure reliability of the calculated orientation angles.

In this data collection, there were eleven reflections in the reflection array, the 2θ of which ranged between 17.0° and 25.0°. Since the χ range of data collection was from 0-90°, three reflections had χ between 0 and 40°, five reflections had $30<\chi<65°$, and three reflections had $65<\chi<90°$. Each time the least squares calculation was performed, the cell dimensions of the crystal were updated.

The alignment and deterioration of the crystal was monitored by measuring the intensity of three standard reflections periodically every 60 reflections (about 70 minutes). Because the crystals are mounted in glass capillaries with only a drop of moisture adhering the

crystal to the walls of the capillary, most crystals tend to undergo some small orientational motions. If the intensity of these monitor reflections decreased by more than 20%, the crystal was considered misaligned. This triggered the re-measurement of the positions of the eleven reflections in the reflection array, followed by the calculation of a new orientation matrix and cell parameters by the least squares method. The three check reflections used for this purpose and their corresponding 2θ values were : (0,-10,0) (2θ =18.02°); (-1,-7,9) (2θ=21.48°); and (1,0,8) (2θ=16.36°). The intensities of these reflections were plotted versus exposure time of the crystal and the resulting slopes were used to obtain decay corrections for the data (Figure 12).

The reflection intensities were measured by the Wyckoff ω-step procedure [87]. In this method, data were collected by holding the detector arm (2θ) fixed and scanning ω. Usually this method of data collection is used when the reflections are spaced closely as in a macromolecular crystal or when it is desirable to collect data quickly to avoid the danger of x-ray damage fo the crystal. In this method, the top of the peak is sampled instead of the profile from background to background. If the peak maximum is too close to one edge of the scan range, the computer program is made to collect additional steps. Thus, slight errors in crystal orientation can be compensated.

Using the Wyckoff method, each peak was scanned 0.09° on either side of the calculated value for the reflection in 7 steps, performed in 0.03° increments. Each step was

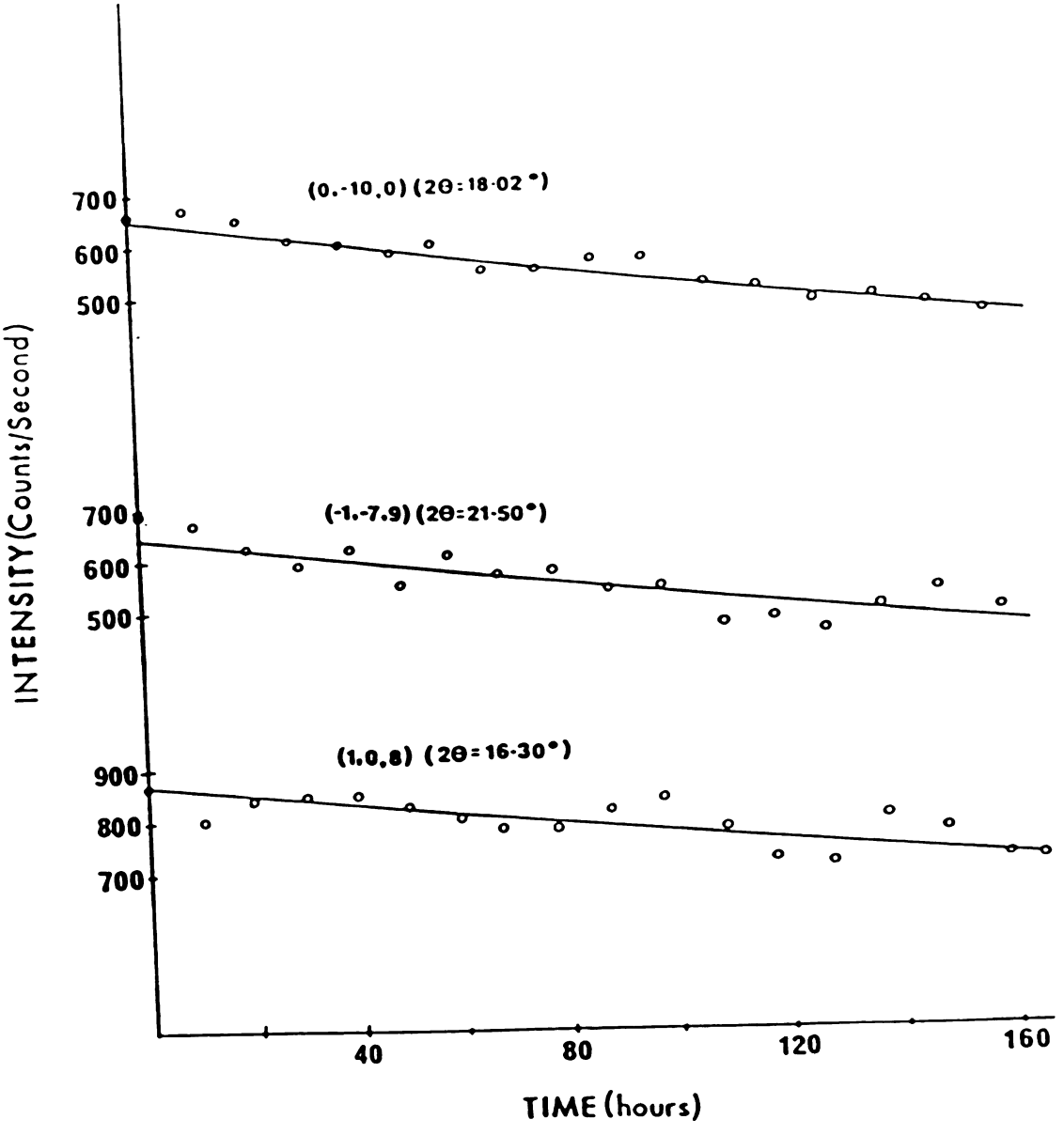


Figure 12. The intensity of the kringle 4 monitor reflections Vs. time of exposure.

measured for a duration of 8 seconds and the five largest measurements were summed to calculate the integrated intensity of the reflection. Background measurements were made , on either side of the peak, for each reflection for a total duration of 11 seconds while the time spent in scanning the peak was 54 seconds. The maximum number of additional steps allowed was 3.

Several measurement procedures preceded and followed the actual intensity data collection. These included: the ω-profile of a reflection (-5,5,5), the intensity profile versus angle (absorption) for the reflections (0,-2,0) and (0,-10,0), and the (h01) zone of reflections to 6.0 Å resolution. The ω-profile of a reflection assesses the quality of a crystal for 3-D data collection from which information is used to set the Wyckoff step scanning parameters. The profile of (-5,5,5) was obtained by measuring the intensity of the reflection at ω values ± 0.65° from its calculated value in 0.05° intervals of ω (Figure 13). The width of the peak at half the maximum height defines the region of the peak to be scanned. From this plot, an offset value can also be obtained to measure the background. For kringle 4, a scan range of 0.18° and an offset of 0.2° was used.

The intensity of absorption versus ϕ-angle was measured for reflections (0,-2,0) and (0,-10,0) to find the region in ϕ of low x-ray absorption so the data collection could be carried out over this ϕ range, and so that an absorption correction could ultimately be applied to the measured

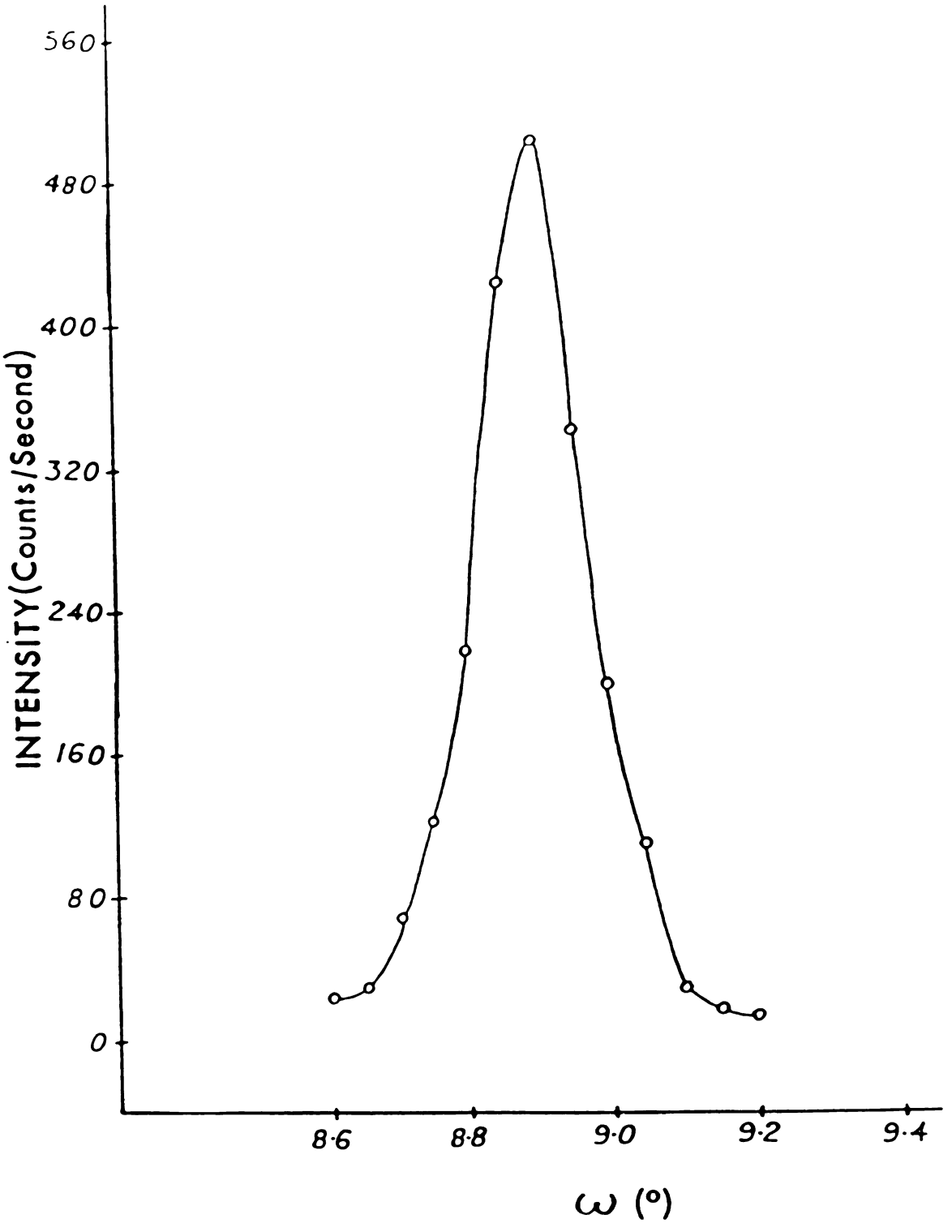


Figure 13. ω-profile of (-5,5,5) taken before the 3-D data collection. Half-width ~ 0.2°*.

intensities. The b^* axis reflections were used because in kringle 4, the b^* direction is coincident with the ϕ axis so that the (0,1,0) Bragg planes are always in the reflecting position for all values of ϕ at χ=90°. Intensity measurements were made every 10 degrees in ϕ over a 200 degree range for both reflections. A plot of Imax/I, where Imax is the maximum intensity of the measured reflection and I the intensity of the same reflection at any other ϕ angle, is shown in Figure 14.

The intensities of the (h0l) reflections to 6.0 Å resolution ($2\theta < 15°$) were measured before and after the data collection to assess the decay due to radiation damage to the reflections in the $2.5<\theta<15.0°$ range. As mentioned earlier, the monoclinic crystals are twinned along the c^* direction. By virtue of the unit cell parameters , certain reflections from the crystal and the twin with l=0,1,3,4, and 8 overlap so closely that they cannot be resolved with intensity data collection diffractometer conditions(4° x-ray take-off angle, detector slit size, etc.). Therefore a 6.0 Å resolution set of twin intensity data also was collected and used to determine a crystal/twin ratio (~1.60) which was then used to correct the twinned measurements to the corresponding individual reflection intensities from the crystal and the twin.

Approximately 3800 reflections were measured to 2.8 ∇ resolution during the data collection for the crystal and 450 reflections for the twin. The 3-D data to 2.8 ∇ resolution was collected in two shells of 2θ in the

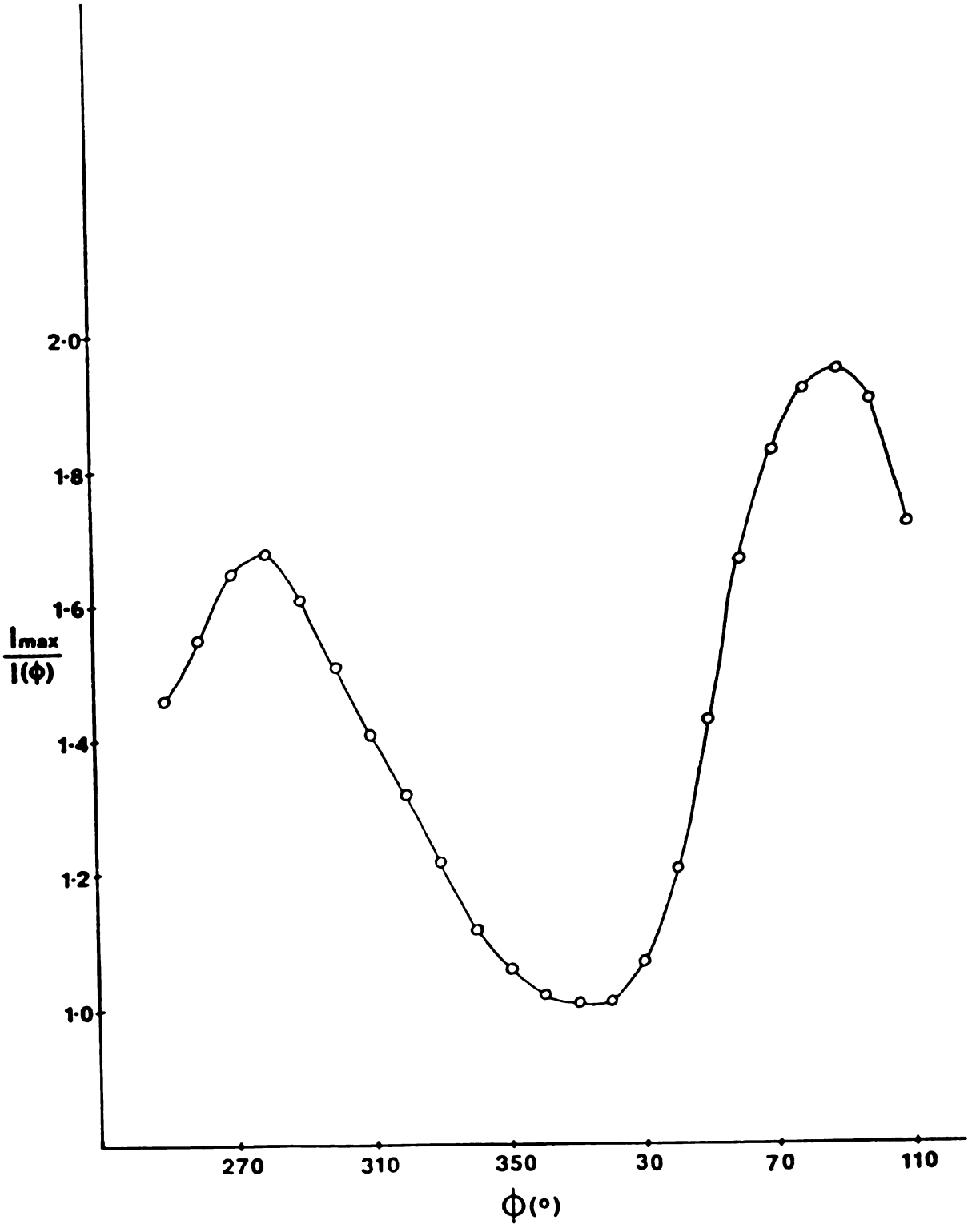


Figure 14. Absorption curve of the (0,-10,0) reflection.

following order ; 25.0°- 32.0° and 2.0° - 25.0°. It took approximately 110 hours of x-ray exposure to measure all of the intensities. Of these 3800 crystal reflections, about 2900 were considered observed (~76%). The observation limit was defined from the average negative intensity of the accidentally unobserved reflections.

IV DATA REDUCTION

The first step in processing protein intensity data is to reduce the measurement for each reflection (hkl) to its corresponding structure factor modulus, |F(hkl)|. The data was reduced using a program called P-DATA, written by C.D.Buck , of our laboratory. The program calculates the structure factor amplitude from the diffraction intensities by means of the equation :

$$|F(hkl)|^2 = CONST \times ABS \times DECAY \times LORPOL \times I(HKL), \quad ....(17)$$

where CONST is a scale constant, ABS is the correction factor due to absorption, DEC is the decay correction to compensate for the loss of intensity due to crystal deterioration from x-ray exposure, LORPOL is the Lorentz-polarization factor which is a geometrical factor, and I(hkl) is the background-corrected intensity of a reflection (hkl).

The absorption correction applied in the P-DATA program was based on the method proposed by North, Phillips, and Mathews [88]. The program uses the absorption curve of Imax/I(ϕ) versus ϕ to apply the absorption correction (Figure 14). The decay correction factor DECAY corrects the intensity deterioration of a reflection as a function of exposure time and it is usually 2 dependent. DECAY factor has the form :

$$I^{\circ}(t) = I(t) * \{1/(1-St)\} \quad ....(18)$$

where $S=-s/I^{\circ}_{CR}$ where s is the slope of the intensity versus

time of exposure curve of the monitor reflections (Figure 12) and I°_{CR} is the intensity of the monitor reflections at zero time. From the three monitor reflections, an average value of 0.00248 was obtained for S, which corresponds to 40% decay at the end of 162 hours of x-ray exposure. A decay correction can also be calculated from a comparision of the (h0l) projection Patterson before and after the data collection. This value (36% at the end of 162 hours of exposure) was in good agreement with the decay factor from the monitor reflections.

Before applying the P-DATA program, the background of the intensity data was averaged in shells of 2θ. Background measurements made at the left and right side of a reflection peak and the total intensity measured are used to calculate the integrated intensity from the equation :

$$I = [\text{total scan} - \frac{\text{sum of background counts}}{\text{background to scan time ratio}}] * \text{scan rate} \quad \ldots.(19)$$

Averaging of background measurements is usually carried out in shells of 2θ and in shells of ϕ if there is a ϕ dependence. For kringle 4, there was no variation of background measurements with ϕ-angle.

A. PROCESSING OF TWINNED REFLECTIONS

By virtue of the unit cell parameters and the relative orientation of the crystal and the twin, certain reflections of the crystal with l=0,4,7,8,11, and 15 overlap very

closely with the corresponding reflections from the twin. The reflections with l=0 from the crystal and the twin overlap exactly. For the other overlapping reflections, the background was systematically higher. Hence they were excluded from the background averaging procedure. Their background was corrected based on the average background of the non-overlapping reflections in the appropriate 2θ range.

To untwin the overlapping reflections from the crystal and the twin, the crystal to twin ratio must be known. As mentioned in the previous Chapter, a 6.0 Å data set of the twin was collected as part of the data collection. The twin data set was corrected with the appropriate background measurements from the crystal data set and using the program P-DATA, the corrected intensities were reduced to structure factor moduli, |F(hkl)|. A systematic comparision of |F(hkl)| of the crystal and the twin was made to obtain a crystal to twin ratio of 1.24. This corresponds to a crystal to twin ratio of 1.55 based on intensity, which agrees very well with the previous estimate of 1.6. Approximately 200 observed reflections were used for this comparision.

When a crystal lattice possess a rotational symmetry axis which is not a symmetry element of the space group of the crystal, crystal specimens may grow as twins [89]. In such cases, the reciprocal lattices of the different crystal twin domains of the specimen overlap. The resulting diffraction intensities are given by linear combinations of the true untwinned intensities of the reflections which are

related by the twinning operation. In order to extract the true intensities from the observed intensities, one must be able to determine the crystal to twin ratio or twinning fraction. Several methods have been proposed for untwinning or resolving the measured intensities into their individual components [89-93].

In the case of kringle 4, the overlap of the reciprocal lattices of the crystal and twin are not exact; this implies that only a fraction of the measured diffraction intensities need to be untwinned. This is very clear from Figure 9 , which shows the reciprocal lattices of the crystal and the twin. In the Figure, the c^* axis of the crystal is shown by the solid line, that of the twin by the dotted line. The a^* and b^* axis of the crystal and the twin are almost coincident. From the drawing it appears as if all the reflections of the crystal with l=0,4,7,8,11, and 15 overlap with reflections from the twin. Since the crystal is mounted with b^* axis along the goniometer axis, the ϕ value of a reflection is related to the ω value directly. Thus, a comparision of the ϕ values of the crystal and the twin reflections will establish whether any reflection is twinned. Since the scan range used in the Wyckoff ω-scan was 0.18°, from the ω-profile (Figure 13) any two reflections with a ϕ separation of greater than 0.4° will be clearly resolved.

Before untwinning the measured intensities, the reflections that are twinned must be identified. From the orientation matrices of the crystal and the twin (Table 5),

Table 5. Orientation matrices of the crystal (a) and the twin (b).

a. crystal

0.03014	0.00001	0.00500
-0.00136	-0.00001	0.02145
0.00004	-0.02033	0.00006

b. twin

-0.03116	-0.00004	-0.00312
0.00136	0.00000	0.02182
-0.00003	0.02033	0.00000

the angles $2\theta, \omega, \phi$, and χ were calculated for all the reflections for the crystal and the twin. Those reflections of the twinning layers with a ϕ seperation of less than 0.4° were considered twinned. A systematic examination of all the reciprocal points to 2.8 Å revealed that out of about 2800 observed reflections, 500 were twinned (~18%). The fraction of the twinned reflections and the average ϕ separation between the crystal and the twin reflection in each layer of l are listed in Table 6.

Aspects of twinning in Figure 9 can be somewhat misleading. For example, it appears as if all the reflections in l=15 layer are twinned. But none of the reflections in l=15 layer are twinned (Table 6). The reason for this is that the Figure was drawn on the assumption that the b^* axis of the crystal and the twin overlap exactly. But the orientation matrices of the crystal and the twin reveal a small angular separation between the two. The separation increases as the distance from the origin of the reciprocal lattice increases.

As mentioned earlier, for the twinned reflections the measured intensity is a linear combination of the intensities which would be observed if there were no twinning. Let I1 and I2 be the measured intensities of reflections which are related by the twinning operation and J1 and J2 represent the true intensities of the reflections (in the absense of twinning). Also, let the crystal to twin ratio be K_t and its inverse by K_{inv}. Then the following relations hold :

Table 6. Fraction of twinned reflections and the average ϕ separation of the crystal and the twin reflections in twinning l-layers.

l	Number of reflections	Number of twinned reflections	% of twinned reflections	Average ϕ seperation (deg.)
0	324	324	100	0.05
1	323	139	43	1.48
2	316	0	0	3.47
3	314	61	19	1.78
4	308	231	75	0.27
5	295	0	0	2.45
6	279	0	0	2.93
7	264	0	0	1.15
8	252	71	28	0.54
9	225	0	0	2.32
10	202	0	0	2.38
11	182	0	0	0.73
12	156	0	0	0.80
13	126	0	0	2.20
14	88	0	0	1.88
15	58	0	0	0.44
16	16	0	0	0.90
total	3785	826	22	1.57

$$I1 = J1 + K_{inv} \times J2 \qquad ....(20)$$
$$I2 = J2 + K_{inv} \times J1$$

These are two simultanious equations in two unknowns, J1 and J2 which solve to :

$$J1 = (I1 - K_{inv} \times I2)/(1-(K_{inv})^2) \qquad ....(21)$$
$$J2 = (I2 - K_{inv} \times I1)/(1-(K_{inv})^2)$$

If both I1 and I2 are unobserved reflections, J1 and J2 are considered to be unobserved. If I1 is observed and I2 is unobserved, then J1=I1 and J2=0 ; on the otherhand if I1 is unobserved and I2 is observed, then J1=0 and J2=I2. In the case of reflections with l=0, where the overlapping reflections of the crystal and the twin are Friedel mates with the same magnitude for h (with opposite signs), there is a special relatioship :

$$J1 = J2 = (I1+I2)/2(1+K_{inv}) \qquad ....(22)$$

The set of simultanious equations (Equation 20) were solved for approximately 700 reflections to resolve the measured diffraction intensities to the individual reflections. Of these, close to 200 reflections were unoberved after untwinning. Thus only 18% of the observed reflections out of a total data set of 2800 are twin-corrected.

V PATTERSON SEARCH

A. SELF ROTATION FUNCTION

As mentioned in Chapter III, monoclinic kringle 4 crystals contain two molecules in the asymmetric unit. To find the non-crystallographic symmetry element or operation which relates the independent molecules with respect to one another, a self rotation search was carried out by Patterson search methods [64]. The Patterson function was calculated with a 1.0 Å grid using reflections from 9.0 Å to 3.5 Å resolution. For the rotation search, the grid points representing the 1500 highest Patterson peak heights within a radial shell between 6.0 and 23.0 Å were selected to represent the Patterson. The self rotation search was carried out, using the SEARCH routine in Steigeman's PROTEIN pacakge [64], in the angular range θ_1= 0°to 360°, θ_2= 0° to 180°, and θ_3 = 0 to 90° for the three Eulerian angles in steps of 5°. The lower limit of 6.0 Å was chosen to eliminate the Patterson origin ripple effects; the higher limit of 23.0 Å, on the other hand, was chosen with a view to eliminate the intermolecular vectors from the rotation calculation. The molecular dimension of the kringle as measured from the prothrombin kringle is 20x28x30 Å [29]. The product correlation function was calculated as a function of the orientation angles θ_1,θ_2, and θ_3.

The rotation search yielded a correlation peak of 14.3 standard deviations (12.3 σ above the mean) at θ_1=90.0°,

θ_2=160.0°, and θ_3=90.0°. The next highest peak was only 5.2 standard deviations. Then a finer grid scan was carried out in steps of 0.3Å around the peak of highest correlation. The fine scan increased the value of the highest peak to 14.5 standard deviations, yielding an optimal fit at θ_1=90.0°, θ_2=161.1°, and θ_3=90.0°.

The self rotation calculation was repeated with a different set of Patterson vectors. This time the Patterson was calculated with reflections from 12.0 Å to 3.5 Å resolution. The grid points representing 1000 highest function values within a radial shell between 5.0 and 15.0 Å were selected to represent the Patterson function. The highest solution appeared at the same set of Eulerian angles but with a peak height of 15.3 σ (9.0 σ above the mean value). The next highest peak had a correlation of 9.5 σ. The subsequent refinement of the rotation angles in steps of 0.3° around the highest peak yielded basically the same angles obtained previously with the different set of vectors and other parameters and the peak height increased to 15.5 σ.

The rotation angles can also be expressed in a set of spherical polar angles where it is easier to visualize the orientation of the rotation axis with respect to the crystallographic symmetry elements (Figure 7). In terms of the polar angles, the self rotation is at ϕ =0.0°, ψ =80.0°, and χ =180.0° (Table 7). This corresponds to a two-fold rotation approximately parallel to the crystallographic a^* axis. Thus the two independent molecules in the asymmetric

Table 7. Self rotation results, both in terms of Eulerian angles $\theta_1, \theta_2, \theta_3$ and spherical polar angles ϕ, ψ, χ. All the symmetry related solutions are listed in degrees.

	θ_1	θ_2	θ_3	ϕ	ψ	χ
1	90.0	60.0	90.0	0.0	80.0	180.0
2	270.0	200.0	270.0	0.0	80.0	180.0
3	90.0	340.0	90.0	180.0	10.0	180.0
4	270.0	20.0	270.0	180.0	10.0	180.0

unit are related by a non-crystallographic two-fold rotation approximately parallel to a^* axis, similar to the non-crystallographic symmetry observed in α-chymotrypsin [86].

When the Patterson search is carried out over the entire angular range, the rotation function shows additional solutions which are symmetry related to the first solution. This arises from the symmetry of the Patterson and the definition of the Eulerian angles, as described in Chapter II. The symmetry related solutions are given in Table 7, both in terms of the Eulerian angles and the polar angles. The solution at ϕ=180.0°, ψ=10.0°, and χ=180.0° corresponds to a non-crystallographic two-fold rotation approximately parallel to the c axis. Thus, depending on how one chooses the crystallographic asymmetric unit, the non-crystallographic symmetry can be considered as a two-fold rotation parallel to either the a^* or c^* axis. This can be seen from Figure 15. The molecules marked A and B are the independent molecules and those marked A′ and B′ refer to the symmetry mates related by the crystallographic 21 screw parallel to the b axis. The relationship between A and B is a two-fold rotation parallel to a^* axis whereas B′ and A are related by a non-crystallographic two-fold rotation parallel to the c^* axis. Thus there are two posssible "dimer" solutions for kringle 4. These will be elaborated upon later in Chapter IX.

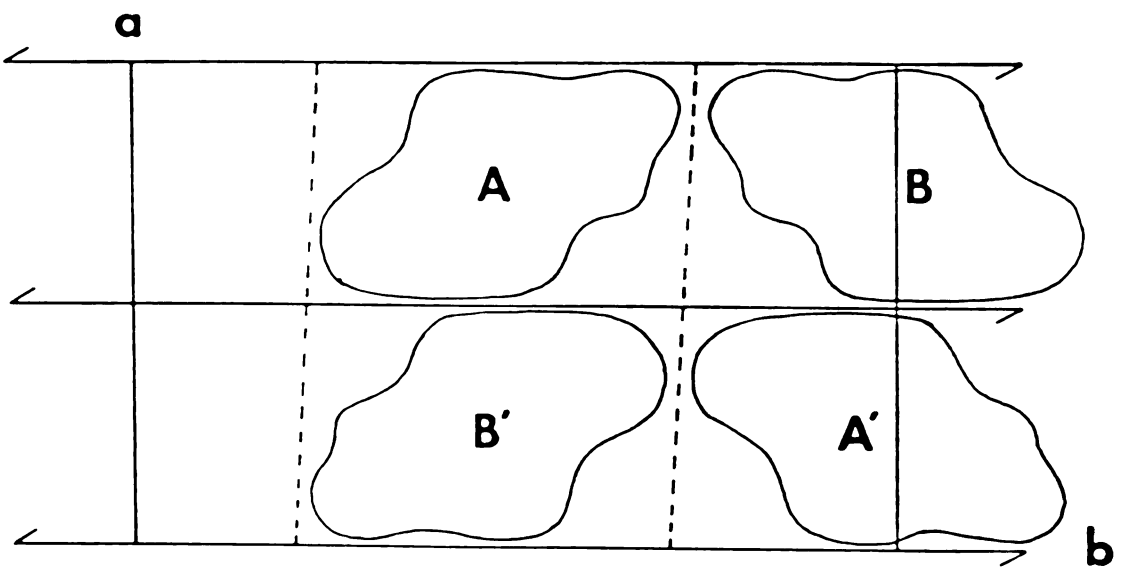


Figure 15. Schematic of the four molecules in the unit cell. Non-crystallographic symmetry elements are indicated by the dotted lines.

B. CROSS ROTATION FUNCTION

The next stage of the molecular replacement analysis was to determine the relative orientation of the kringle 4 molecules with respect to the crystallographic coordinate system using a model based on a known kringle structure. The only kringle structure that has been solved so far is that of kringle 1 (K1) of prothrombin fragment 1, carried out in our laboratory [29,30]. The primary sequences of kringle 4 of plasminogen (K4) and K1 are shown in Figure 16. Thus, from Figure 16, of a total of 80 residues in the kringle, 36 are identical and 5 are very similar. This amounts to a homology of 51% between the two primary sequences. It is also clear from Figure 16 that the highest degree of conservation in the kringle sequences occurs in the region of the inner loop residues (C and D) near the disulfide bridges Cys53-Cys76 and Cys25-Cys64.

A cross rotation search was carried out with the coordinates of K1 with many different sets of cross rotation caluculations being performed with different model structures. Initially, a simple polyalanine backbone of K1 was used as the model. This contains 387 atoms and corresponds to a little more than 50% of the K4 structure. Later, the conserved side chain atoms were also included in the calculation, increasing the total number of atoms to 489 (~72%). Finally all the atoms in K1 were used in the calculation.

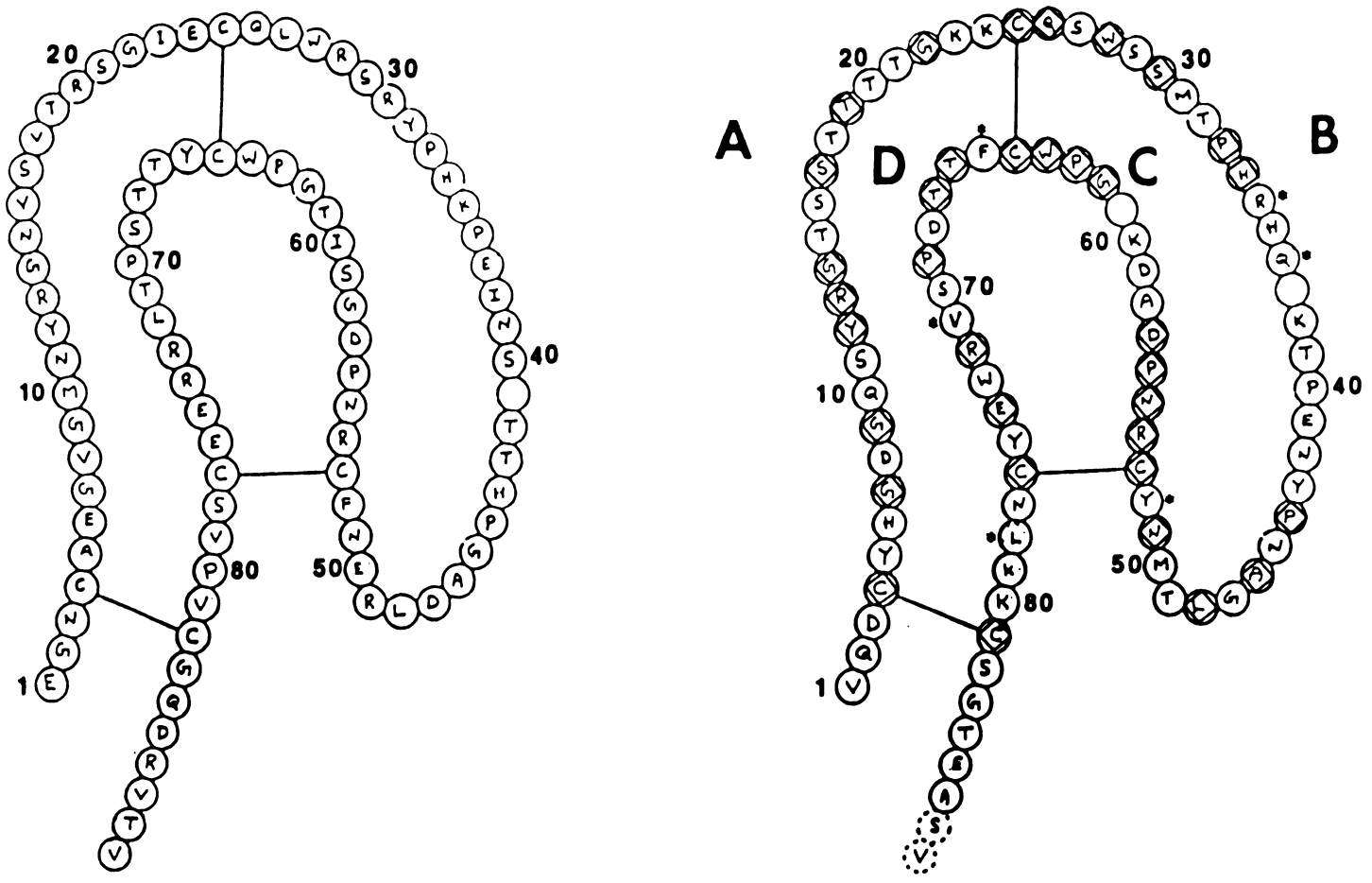


Figure 16. Sequences of bovine prothrombin fragment 1 kringle (a) and human plasminogen kringle 4 (b): residues identical in both sequences are enclosed by diamond blocks; similar residues are indicated by asterisks; loop segments designated A, B, C, and D.

The model structure was based on a 2.8 Å refined structure of K1 , with 178 water molecules, refined to a R value of 18% (T. P. Seshadri, unpublished results of this laboratory). The average thermal parameter (B) at this stage was 39.8 $Å^2$. On the other hand, K4 scatters x-rays well and does not decay very much on long x-ray exposure. Hence, a B value of 20 $Å^2$ was applied instead for the structure factor calculations.

Each model of K4 was placed in a triclinic cell with orthogonal axes of length 75 Å. Triclinic structure factors were calculated from 9.0 Å to 3.5 Å with an overall temperature factor of 20 $Å^2$. A 3.5 Å model Patterson with grid spacings of 1.0 Å was then calculated. The Patterson synthesis of the crystal was calculated similarly using all observed refelctions from 9.0 Å to 3.5 Å resolution with 1.0 Å grid spacings. The product correlation of the crystal and the model Patterson maps was calculated as function of the Eulerian angles in steps of 5° in the angular range θ_1=0 to 360°, θ_2=0 to 180°, θ_3=0 to 180°, and subsequently in smaller steps of 0.3° around high correlation peaks. All the rotation searches were performed using the SEARCH routine in PROTEIN package [64].

The various K1 models were used separately in the rotational searches to determine the orientation of the independent molecules A and B within the crystal. Each of these rotational searches yielded two practically indistinguishable highest correlation peaks in general

positions with their height corresponding to about 7 σ (Table 8).

The non-crystallographic two-fold rotation found in the self rotation calculation could now be compared with the solutions of the cross rotational search [56]. Let K4A and K4B represent the two independent kringle 4 molecules in the asymmetric unit of the monoclinic crystal and [S], the self rotation matrix relating these two molecules. If [CR1] and [CR2] are the rotation matrix solutions 1 and 2 of the cross rotation search, then the following relation hold good :

$$\begin{aligned} K4B &= [S] \times K4A \\ K4A &= [CR1] \times K1 \\ K4B &= [CR2] \times K1 \\ [CR2] &= [S] \times [CR1] \end{aligned} \qquad \ldots\ldots(23)$$

Using [CR1] and [CR2] from the cross rotation, [S] was calculated. The various matrices used are listed in Table 9 from which it is very clear that there is excellent agreement between the self rotation matrix calculated from the cross rotation results and the observed self rotation matrix.

As in the case of self rotation, the cross rotation calculation also revealed the symmetry of the Patterson function, when the search was extended to the whole range of the Eulerian angles. Four solutions were obtained which correspond to the four molecules in the unit cell. The relationship among the four solutions can be calculated as before and are listed in Table 10 from which the agreement between the self rotation and the cross rotation is clearly evident.

Table 8. Cross rotation results. (a): Polyalanine backbone kringle model (387 atoms; 57%). (b): Kringle with conserved side chains (489 atoms; 72%). (c): All the kringle atoms (621 atoms; 92%). All models based on the coordinates of K1 of prothrombin fragment 1.

Model	Two highest correlation peaks (deg.)			Peak height (σ)	Height above the mean value (σ)
	θ_1	θ_2	θ_3		
a	202	152	68	7.0	5.1
	46	142	100	6.5	4.6
b	200	149	71	6.2	4.6
	44	142	99	6.8	5.2
c	199	149	66	5.9	4.1
	45	145	98	5.6	3.9

Table 9. Correlation between the cross rotation and the self rotation results.

Cross rotation matrix 1 (α=313.62°, β=142.80°, γ=186.88°)			Cross rotation matrix 2 (α=70.23°, β=30.30°, γ=338.77°)		
0.45883	-0.78454	0.41710	0.61299	-0.77144	0.17065
-0.65512	-0.61583	-0.43769	0.63488	-0.60951	0.47479
0.60025	-0.07243	-0.79653	-0.47029	-0.18270	0.86340

Predicted self rotation matrix (ϕ=0.15°, ψ=81.63°, χ=182.43°)			Observed self rotation matrix (ϕ=0.00°, ψ=80.55°, χ=180.00°)		
0.95766	-0.00120	0.28789	0.94609	0.00000	0.32392
0.01115	-0.99909	-0.04124	0.00000	-1.00000	0.00000
0.28768	0.04271	-0.95678	0.32392	0.00000	-0.94609

Table 10. Correlation between cross rotation and self rotation results.

a. Symmetry related cross rotation angles. (deg.)

	α	β	γ
1	313.62	142.80	186.88
2	109.77	149.70	158.77
3	70.23	30.30	338.77
4	46.38	322.80	186.88

b. Self rotation angles and the related cross rotation angles.

Related cross rotations	Predicted self rotation (deg.)			Observed self rotation (deg.)		
	φ	ψ	X	φ	ψ	X
1 - 2	188.30	8.45	180.29	180.0	9.45	180.00
1 - 3	0.15	81.63	182.43	0.0	80.55	180.00
1 - 4	90.00	90.00	180.00	Crystallographic 2_1 screw axis		

C. TRANSLATION FUNCTION

The location of the two correctly oriented molecules in the unit cell was determined using the program BRUTE written by Fujinaga and Reed [77] to calculate the translation function. In this phase, only the K4 model with the conserved side chains was used for the search. The unit cell of K4 contains two crystallographically symmetrical molecules of type A (A and A') and two crystallographically symmetrical molecules of type B (B and B'). First, molecules A and B were correctly oriented corresponding to the solutions of the cross rotation search. BRUTE translation function were then calculated to position A and B separately. Later, one of them (molecule A) was fixed at the correct position and a search was carried out for the other molecule (molecule B). This was followed by another calculation where molecule B was fixed and a search performed for the position of A.

The results of BRUTE translation function calculation are expressed in terms of a correlation coefficient between the observed structure factor |Fo| and the calculated structure |Fc|, as given by equation 14 of Chapter II. The calculation was repeated using reflections in various resolution ranges. In all cases, the calculation intially was performed in steps of 0.5 Å translation and in steps of 0.1 Å subsequently around the solutions of highest correlation. For the independent search of the molecules A and B, the search could be restricted to two dimensions,

since in monoclinic space group $P2_1$, the origin along the y axis is arbitrary.

The results of the translation function calculation for the independent searches of molecule A and B are listed in Table 11 which shows very good agreement between the calculations performed at various resolution ranges. The best correlation coefficient is obtained at low resoltion (9.0 Å to 5.0 Å). The heighest peak has a correlation coefficient of 8.9 σ (about 3.8 σ above the mean) for molecule A while it is 8.0 σ and 4.1 σ respectively for molecule B (with 7.0 - 3.5 Å data set). The rotation and translation parameters for molecules A and B after refinement in steps of 0.1 Å are listed in Table 12.

To determine the relative positions of molecules A and B, molecule A was fixed in the unit cell at the position given in Table 12 and a translation search was carried out for molecule B. This time the search was extended to three dimension since the origin along the y axis has already been fixed by specifying the position of molecule A. Only reflections in the resolution range 5.0 Å to 3.5 Å were used for this search. The maximum correlation coefficient was obtained at TXB=0.80, TYB=1.00, TZB=0.70 and has a correlation coefficient of 0.41. The correlation coefficient is about 11 σ(~6.0 σ above the mean). The search was repeated by fixing molecule B and searching for molecule A. Essentially the same results were obtained, confirming the relative positions of molecules A and B. The

Table 11. "BRUTE" translation search solutions for molecules A and and B at various resolution ranges.

Molecule A.

TX = 6.0 Å; TZ = 11.0 Å

Resolution range (Å)	Correlation coefficient	Height (σ)	Height above the mean value (σ)
9.0 - 5.0	0.32	6.8	2.6
5.0 - 4.0	0.25	6.2	2.8
7.0 - 3.5	0.22	8.9	3.8
9.0 - 3.5	0.21	8.6	3.7

Molecule B.

TX = 9.8 Å; TZ = 9.3 Å

Resolution range (Å)	Correlation coefficient	Height (σ)	Height above the mean value (σ)
9.0 - 5.0	0.31	6.6	2.9
7.0 - 3.5	0.22	8.0	4.1
5.0 - 4.0	0.21	5.8	3.8
9.0 - 3.5	0.20	7.8	4.0

Table 12. "BRUTE" translation search solutions for the two molecules A and B. (a): Search for molecule B while molecule A is stationary. (b): Search for molecule A with stationary B in the correct position. (c): Final refined rotation and translation parameters for molecules A and B.

a.

Highest correlation peaks (Å)			Correlation coefficient	Peak height (σ)	Height above the mean value (σ)
TX	TY*	TZ			
25.8	48.6	31.8	0.43	11.5	6.5
20.3	28.5	26.0	0.36	9.6	4.7
1.1	47.3	38.2	0.33	8.9	3.9

b.

6.2	49.0	11.3	0.42	11.0	6.0
21.2	15.2	9.0	0.34	8.9	4.0
12.0	25.1	20.3	0.33	8.7	3.8

c.

	Rotation angles (deg.)			Translation vectors (Å)		
	α	β	γ	TX	TY*	TZ
Molecule A	313.62	142.80	186.88	6.2	0.0	11.3
Molecule B	109.77	149.77	158.77	25.8	48.6	31.8

* Only relative separation important due to arbitrary nature of origin along the 2_1 screw axis in $P2_1$.

final refined rotation and translation parameters for molecules A and B are listed in Table 12.

To verify the correctness of the rotation and translation solutions, the two molecules were positioned properly in the unit cell and displayed on an Evans and Sutherland PS390 interactive computer graphics system equipped with a stereographic screen using the program FRODO [94]. All the symmetry related molecules within a radius of 40.0 Å of the center of the unit cell were generated to examine whether there was any unacceptable contact or interaction between the moleucles. In practice, this is equivalent to the packing function program used frequently to determine the translation solution [80-81]. There were very few short van der Waals contacts between neighboring molecules, thus confirming the correctnes of the molecular replacement solutions. The packing arrangement in the crystal is described in Chapter IX. The molecules pack very tightly which agrees with the excellent x-ray scattering power of the crystal and its V_m value. Based on crystal packing criterion alone, there would be exceptionally few allowed molecular replacement solutions because of the limited space availabe for the solvent (34%).

VI SYMMETRY AVERAGING

Using the refined orientational and positional parameters, molecule A and B of kringle 4 were placed in the monoclinic crystal. The non-conserved side chain atoms were removed from the model of kringle 1 of prothrombin fragment 1 so as not to bias the starting model of K4. A set of phases were calculated using the model. For the structure factor calculation, 978 atoms were used of a total of 1352 atoms that are present in two molecules of kringle 4 (~72%) with an overall temparature factor of 20^2. The discrepancy index (R factor), between the observed structure factor modulus, |Fo|, and the calculated structure factor modulus, |Fc|, at this stage was 45%.

A modified Wilson distribution was plotted to obtain the overall scale factor (K) and the difference in temparature factor (DB) between the observed and calculated structure factor moduli. A. J. C. Wilson [95] had derived the following equation :

$$\ln(\Sigma|Fc|/\Sigma|Fo|) = \ln K - \Delta B\ (\mathrm{Sin}^2\theta)/\lambda^2 \quad ....(25)$$

where λ is the wavelength of the x-ray radiation. If the left hand side of the equation is evaluated for shells of 2θ and the values plotted against $\sin^2\theta/\lambda^2$, the result should be a straight line in which the extrapolated intercept at $\sin^2\theta = 0.0$ is ln K and the slope is $-\Delta B$. Since protein crystals contain large amounts of mother liquor, and since

the low 2θ reflections are influenced by the solvent, the low angle reflections were excluded from the Wilson plot. From the Wilson plot, the scale factor was found to be 77.4 and the ΔB value = -1.4 $Å^2$, thus confirming the overall B value that was used for the structure factor calculation.

The crystallographic asymmetric unit of kringle 4 contains two molecules in intimate association with one another. The transformation between the two molceulces can be expressed as

$$X2 = [C]X1 + d \qquad ....(26)$$

where each point X1 in molecule A is related to the point X2 in molecule B by a rotation matrix [C] and a translation vector d. The rotation matrix [C] specifies the direction of the rotation axis and translation vector d contains information about the position of the rotation axis in the unit cell.

In non-crystallographic symmetry averaging, which has been used extensively in the structure analyses of virus particles, the value of C and d are sought which minimize the quantity

$$\Sigma \{\rho_1(X1) - \rho_2(X2)\}^2 \qquad ....(27)$$

where ρ_1 and ρ_2 are the electron density values at the points X1 and X2, respectively. The rotation matrix and the translation vector were determined using the rotation function and the translation function, as described in

Chapter V. These values were refined by the least squares minimization of the equation 27.

Refinement of the transformation parameters between the two molecules of kringle 4 in the asymmetric unit was performed using a computer program written by Wayne Hendrickson and Janet Smith. The 2.8 Å resolution electron density map calculated using the trial phases was used as the starting map for refinement. An approximate molecular boundary was defined so that during the refinement, only those portions of the electron density which lie within a radius of 15 Å centered on the center of mass of molecule A were used. The intermolecular solvent regions and portions of the electron density associated with neighboring molecules were thus partially excluded during the refinement process.

The measure of agreement between the electron densities of the two molecules was assessed in terms of a correlation coefficient which can be expressed as :

$$CC = \Sigma(\rho_1-\bar{\rho}_1).(\rho_2-\bar{\rho}_2)/[\Sigma(\rho_1-\bar{\rho}_1)^2.\ \Sigma(\rho_2-\bar{\rho}_2)^2]^{-\frac{1}{2}} \quad ....(28)$$

where ρ_1 and ρ_2 are the electron densities of molecules A and B, $\bar{\rho}_1$ and $\bar{\rho}_2$ are the average electron densities , with the summations being taken over all points within the molecular boundary defined above. For the case where the two electron densities are exactly the same, CC would have a value of 1.0. The correlation coefficient between the two molecules using the initial transformation parameters was

0.77. The coefficient increased to 0.80 after five cycles of refinement. The initial and final parameters are listed in Table 13.

The position of the local two-fold axis can be derived from these parameters and is shown in Figure 17. The rotation axis intersects the Y axis at 0.74 and Z axis at 0.23, given in fractional units of the crystallographic unit cell.

Table 13. Results of non-crystallographic symmetry averaging. The spherical polar angles ϕ, ψ, and χ are defined according to Rossmann and Blow [47]. Xoffset, Yoffset, and Zoffset refer to the intersection of the rotation axis with the crystallographic a, b, and c axes respectively.

	Initial parameters	Final parameters
ϕ' (deg.)	-9.45	-8.49
ψ' (deg.)	90.00	89.53
χ (deg.)	180.00	180.00
Xoffset (Å)	0.00	-0.12
Yoffset (Å)	36.82	36.78
Zoffset (Å)	11.55	11.62
Correlation coefficient	0.767	0.797

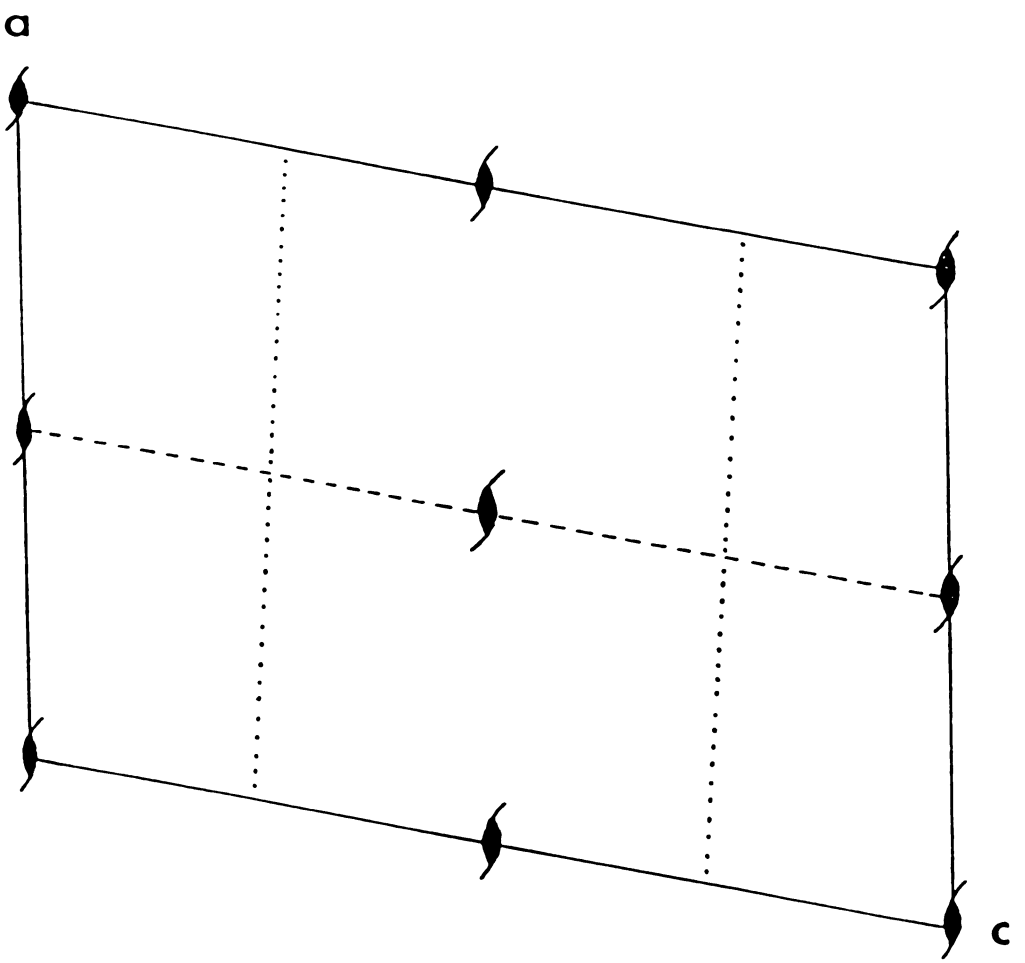


Figure 17. A schematic diagram of the non-crystallographic symmetry elements. The local two-fold axis approximately parallel to the a* axis is shown by dotted lines (·····) and the local two-fold axis approximately parallel to the c* axis is shown by broken lines (------). The crystallographic 2_1 screw axis is indicated by the conventional symbol.

VII RESULTS

The building of a model of K4 structure started with the symmetry averaged 2.8 Å electron density map. The unaveraged 2.8 Å and 3.5 Å maps were examined closely where the averaged map had weak density and in the interkringle tail regions of the molecule which are involved in crystal packing interactions with neighboring molecules. The modelling of the electron density maps was performed on an Evans and Sutherland PS390 computer graphics using the program FRODO. All the maps were contoured at 1 σ level and the amino acid residues were adjusted to fit the electron density as well as possible.

Most of the residues have well defined side chains. This was especially true for the larger side chains like Tyr5, His6, Trp73, Met50, Tyr43, Lys23, Lys24, Tyr75 etc., (Figure 16) which are different from the smaller side chains of the fragment 1 kringle. The electron density of the side chains of Tyr5 and Trp73 are shown in Figures 18 and 19. The presence of very good electron density for most of the non-conserved side chains was another good indication of the correctness of molecular replacement solutions. There was reasonable density even in regions which have a deletion in kringle 4 compared to fragment 1 kringle. The amino terminal (3 residues) and the carboxy terminal (5 residues) also have strong density though it is difficult to fit all the side chains in these regions (Figures 20 and 21). The carboxy terminal corresponds to density for only five residues. Thus, it appears that only the fraction with the

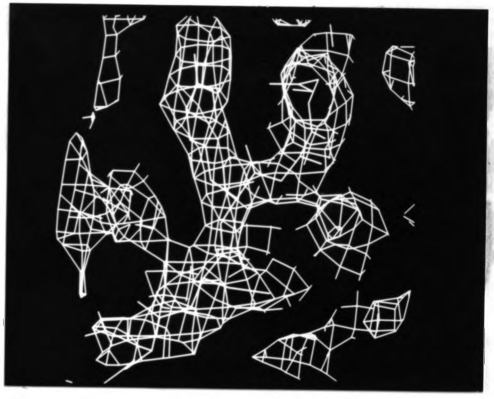


Figure 18. Electron density of the side chain of Tyr5 of molecule A (contours at 1 σ).

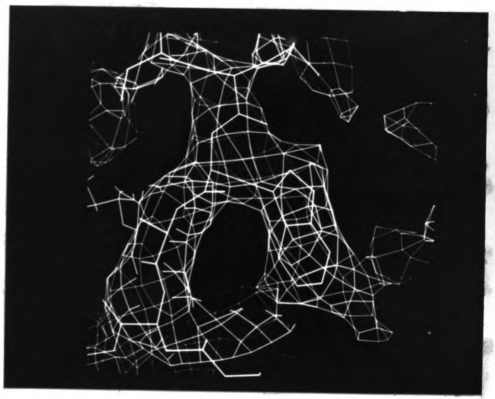


Figure 19. Electron density of the side chain of Trp73 of molecule B (contours at 1 σ).

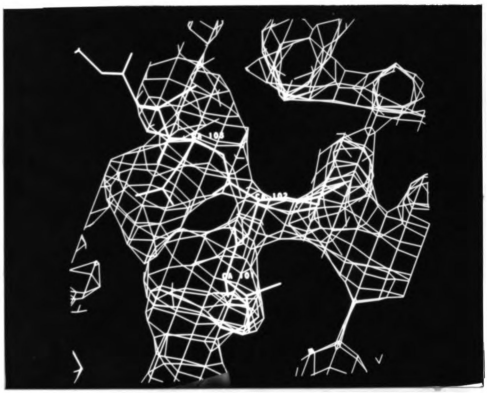


Figure 20. Electron density of the amino terminal of molecule A (contours at 1 σ).

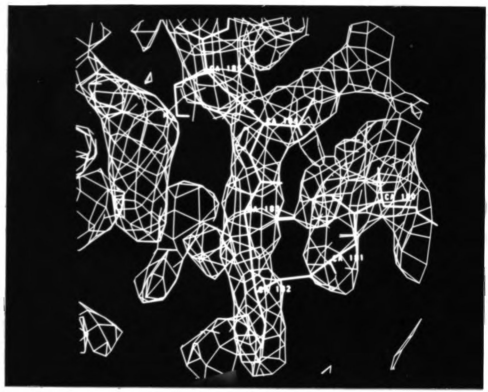


Figure 21. Electron density of the carboxy terminal of molecule B (contours at 1 σ).

shorter interkringle tail which ends with Ala86 crystallizes; this fraction constitutes ~ 70% of the kringle 4 mixture used in crystallization.

The Ramachandran plot of kringle 4 molecule A is shown in Figure 22 from which it is very clear that the structure conforms satisfactorily to the allowed conformational regions. The Ramachandran plot of molecule B is very similar to that of molecule A. A graphic illustration of the structure emphasizing the secondary structural elements of kringle 4 is shown in Figure 23. The structure is in good agreement with the Chou-Fasman calculations of secondary structure [96] which predict no helix, 37% β-structure, and a high percentage of β-turns(49%). However, the implications of the ultraviolet circular dichroism spectra of kringle 4 [96] does not agree with the Chou-Fasman predictions and the structure observed crystallographically. A comparision of the secondary features as predicted by Chou-Fasman calculations, CD spectra and the crystal structure is summarized in Table 14.

A. GENERAL CONFORMATION OF KRINGLE 4

The overall conformation of the polypeptide chain of kringle 4 is very similar to that of prothrombin fragment 1 kringle [29,30]. The secondary structural features of kringle 4 are listed in Table 15. The closed interkringle loops (C and D, Figure 16) form two distinct antiparallel chain segments due to the close contact of two disulfide

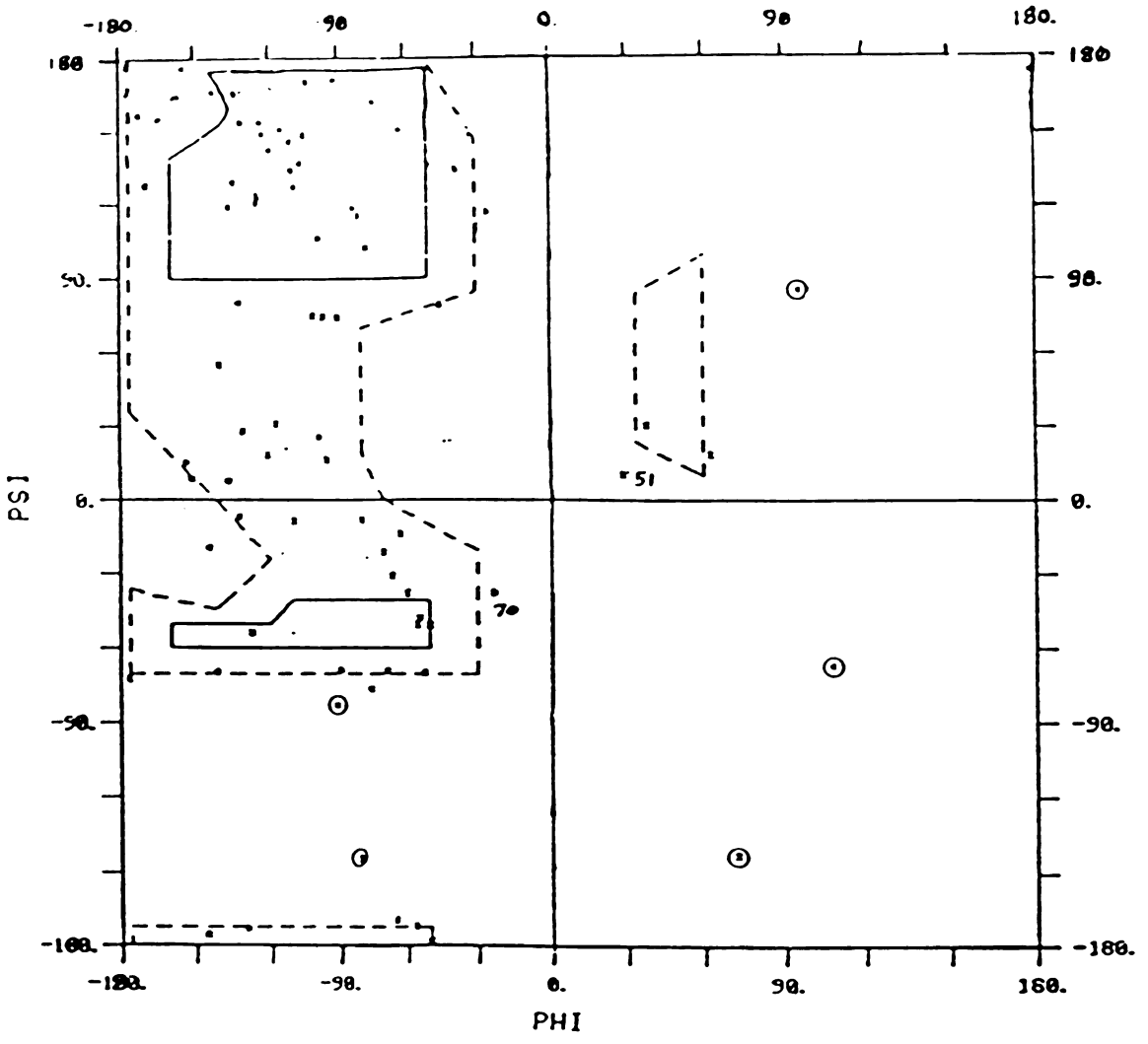


Figure 22. Ramachandran angles of kringle 4 (molecule A). Allowed regions enclosed by dotted lines. Circled residues Gly; outlier residues numbered.

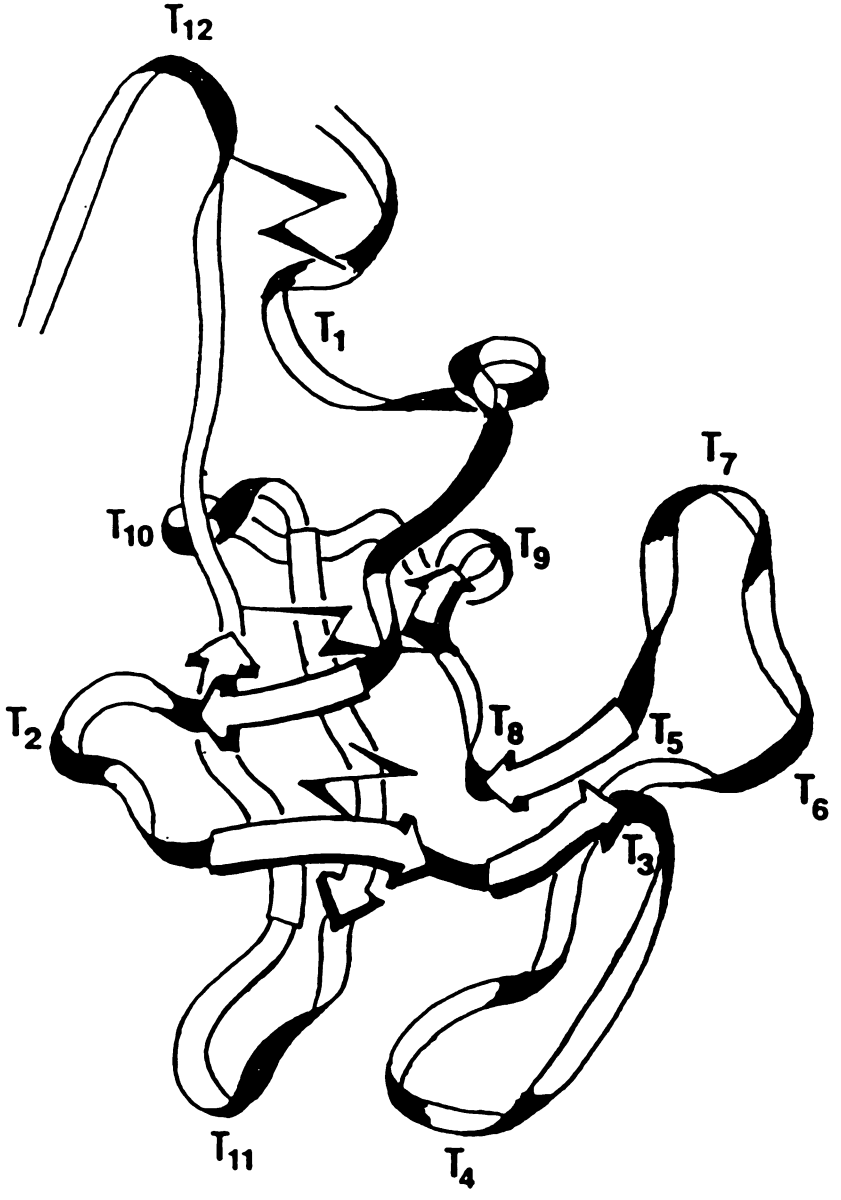


Figure 23. Graphic representation of the folding of kringle 4 emphasizing secondary structure. The reverse turns labelled T1-T12; disulfide bridges indicated appropriately.

Table 14. Comparison of secondary structural features of kringle 4.

Secondary structure	Chou-Fasman prediction (%)	CD spectra (%)	Crystal structure (%)
α-helix	0	0	0
β-sheat	37	64	35
β-turns	49	30	50
random coil	14	6	15

Table 15. Secondary structural elements of kringle 4.

A. β-structure

	Ser17	Cys25		Gln36	Asn51
β1	Thr18	Lys24	β2	Ser27	Met50
	Thr19	Lys23		Trp28	Thr49
				Trp63	Tyr75
β3	Arg54	Trp63	β4	Cys64	Glu74
	Asn55	Pro62		Phe65	Trp73
				Thr66	Arg72

B. Kringle β-turns

T1	His6-Gly7-Asp8-Gly9
T2	Thr19-Thr20-Thr21-Gly22-Lys23
T3	Ser27-Trp28-Ser29-Ser30
T4	Thr32-Pro33-His34-Arg35
T5	His36-Gln37-Lys38-Thr39
T6	Thr39-Pro40-Glu41-Asn42
T7	Tyr43-Pro44-Asn45-Ala46
T8	Thr49-Met50-Asn51-Tyr52
T9	Asn55-Pro56-Asp57-Ala58-Asp59
T10	Asp59-Lys60-Gly61-Pro62
T11	Asp68-Pro69-Ser70-Val71

C. Other β-turns

T12	Ser82-Gly83-Thr84-Glu85

bridges in the folded state; one produces a short stretch of two-strand antiparallel β-structure in the D loop (β4, Table 15) linked to a strand of the C loop (β3, Table 15). A more liberal assignment of β-structure could expand the regions to include : Tyr52-Asn55, Pro62-Thr66, Trp73-Asn77 (Figure 23). These surround Cys25-Cys64 and Cys53-Cys76 and display practically absolute residue conservation among 11 kringle sequences (Table 1). The sulfur atoms of disulfide bridges Cys25-Cys64 and Cys53-Cys76 make unusually close van der Waals contacts, which produce a sulfur cluster not far removed from the center of gravity of the kringle structure (Figure 23) [30]. The two loop structures (C and D) appear to provide the nucleus of the kringle folded structure [30] which is also consistent with the aerobic restoration of these disulfide bridges in reduced kringle 4 with concomitant refolding [97].

The segments A and B of the outer kringle sequence, which differ in length by seven peptide units, display a similar three-dimensional conformation. The A and B segments each fold into two antiparallel open loops. The initial residues of segment A (near Cys4-Cys81) appear to mimic the folding of the antiparallel inner loop C and stack adjacent to it while those of B (near Cys25-Cys64) copy D and stack with it (Figure 23). This gives rise to a pair of stacked duplex loop structures that are approximately related to each other by two independent 90 rotations and a translation. Thus the overall three-dimensional structure of the kringle appears to be the result of an intricately

duplicated folding pattern [30]. The overall shape of the kringle structure resembles that of a highly eccentric oblate ellipsoid of approximate dimensions 20 x 28 x 30 Å [29].

The kringle structure appears to contain about 11 reverse turns (Table 15) of varying degrees of sharpness. Of these, four make up two larger, slower turns (T6-T7, and T9-T10). In addition, the carboxy terminal has a sharp turn (T12, Table 15). Thus it appears that about half of the kringle residues concern themselves with turns, of which one-third are conserved among 11 kringle sequences (Table 1). The relatively small fraction of the β-structural features of the kringle fold may be due to the large number of turns [29].

The path of the main chain in the kringle structure starts with a turn from His6 to Gly9 (T1), similar to the prothrombin fragment 1 kringle structure [29]. A 15 Å stretch of more or less extended chain follows from Ser11-Ser17 and runs approximately parallel to Asn45-Asn51 at a seperation of about 6.0 Å. The structure then forms β1 from Ser17 to Cys24 (Table 15), which was interpreted as two antiparallel strands with T2 at the turn. A short antiparallel contact is made at b2 (Table 15) immediately followed by T3 which leads to a large loop. The latter extends from Trp28-Gln37 with T4 effecting the turn. The main chain then enters into a complicated curve region (T5, T6, and T7). Of these the curves T6 and T7 are sharp and similar to the curves T5 and T6 of fragment 1 kringle [29].

This part of the structure embodies most of the B segment of the kringle sequence, which is the longest of the four segments. A short piece of chain (Ala46-Thr49) runs antiparallel to Tyr52-Pro56 at about 7.0 Å seperation in this region with a turn T8 in between. This is followed by the curves T9 and T10 which carry some of the residues implicated in lysine binding. The structure then enters into the β3 and β4 regions (Table 15) which contain super conserved residues many of which are aromatic. The T11 turn produces the β4 structure. Finally the kringle 4 structure concludes with another sharp turn (T12) in the carboxy terminal interkringle tail.

B. DIFFERENCES BETWEEN THE TWO INDEPENDENT MOLECULES

The crystallographically independent kringle 4 molecules are engaged in quite different molecular contacts. The transformation giving the optimal superposition of molecule B (X2,Y2,Z2) on molecule A (X1,Y1,Z1) is :

$$\begin{array}{lcrrrcrl} X2 & = & 1.0117 & 0.0142 & -0.0702 & X1 & & \\ Y2 & = & 0.0138 & -0.9999 & 0.0005 & Y1 & + 73.683 & ....(29) \\ Z2 & = & 0.3018 & 0.0021 & -1.0105 & Z1 & 23.652 & \end{array}$$

In Figure 24, the Cα structures of molecules A and B are optimally superimposed and presented in the orientation originally used to discuss the prothrombin fragment 1 kringle [30]. The the r.m.s deviations of homologous α-carbon atoms of molecules A and B are plotted against the residue number in Figure 25.

Figure 24. α-carbon drawing of the kringle 4 molecule A (blue) optimally superimposed on molecule B (amber).

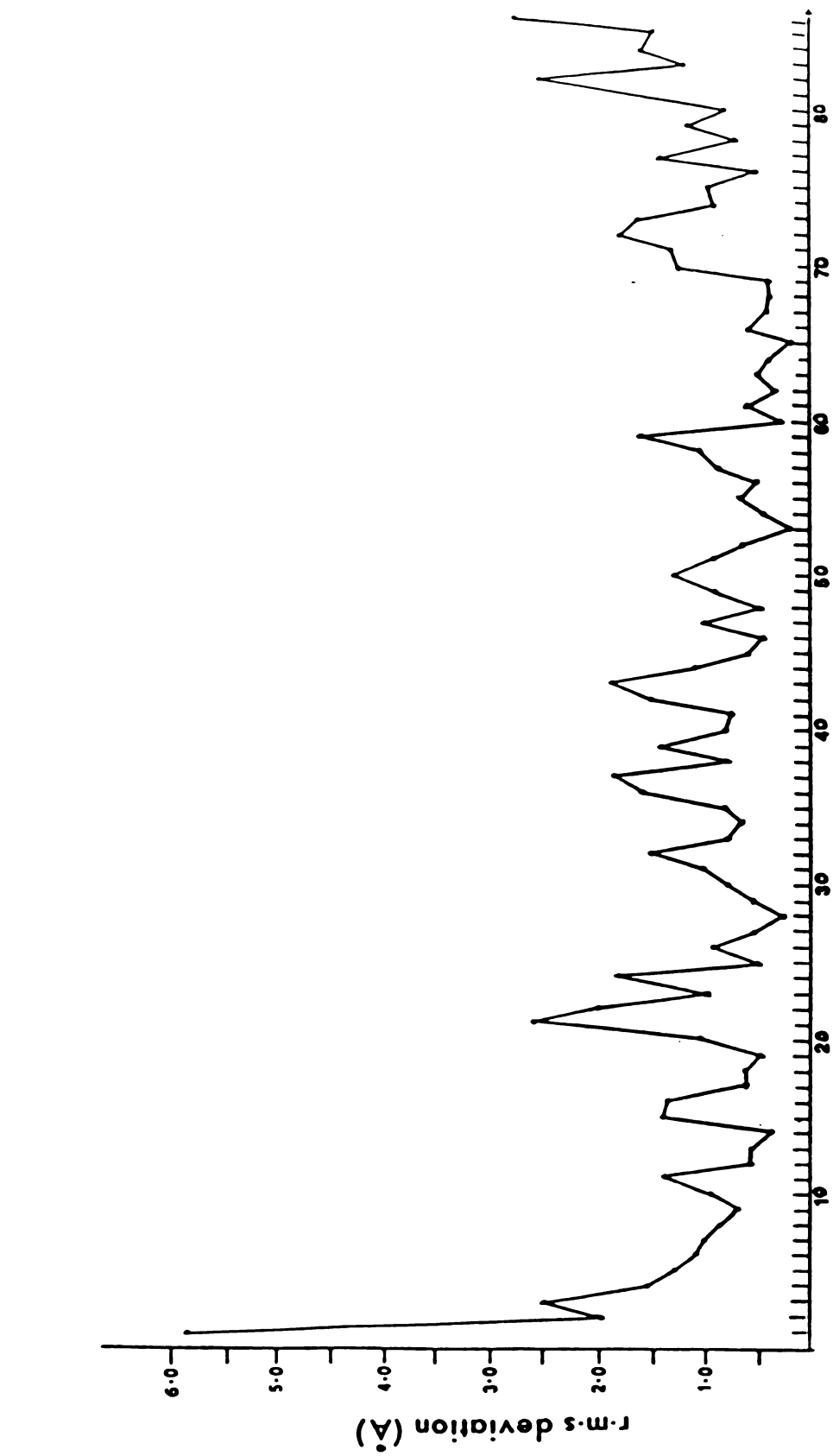


Figure 25. r.m.s deviations between the chemically identical α-carbon atoms of molecules A and B.

From Figure 24 the similarity of the independent kringle 4 molecules is clearly evident. Only four segments, from His6 to Asp8 (b-turn T1, Table 15), Thr21 to Lys24, Thr32 to Tyr43, and Val71 to Trp73 show considerable deviations of their main chain atoms. In addition, there are some differences between the two molecules in the amino terminal and carboxy terminals. The r.m.s deviation of the homologous main chain atoms of the two molecules is 0.70 Å. For all the atoms except the two terminal segments (Val1-Asp3 and Gly83-Ala86), which are different in the two molecules, the r.m.s deviation is 1.90 Å. Residues Thr21, Lys24, Thr32, His36, Gln37, Asp59, and Arg72 exhibit the largest deviations. The conformational differences of the main chain segments in both the kringle 4 molecules must be due to different crystal contacts, since most of the residue segments described above are involved in intermolecular contacts. The packing of the molecules in the crystal lattice is described in Chapter IX.

Relatively large deviations are found near Tyr43, probably caused by different intermolecular contacts. As a consequence, Tyr43 is well defined from the electron density in molecule A, but only partially in molecule B. It is interesting to note that this residue is part of the B segment (Figure 16) where insertions or deletions occur frequently in kringle sequences (Table 1), thus suggesting a 'soft' property.

The side chains of the two molecules exhibit more differences. A qualitative comparision of the electron

density around some of the side chains which show differences between the two molecules is listed in Table 16 while Val1, Lys79, and Ser82 are not well defined in either of the two molecules. Apart from this and those in Table 16, all other side chains have good density in both the molecules.

C. COMPARISON OF K4 WITH K1

In general, the kringle 4 models are very similar in overall conformation to the refined prothrombin fragment 1 kringle [29]. The largest main chain deviations are observed in the segment Ala58-Gly61. It is interesting to note that in kringle 4 there is a deletion in this region compared to the fragment 1 kringle. This region embodies Asp59 which has been implicated in lysine binding of kringle 4 [9]. Relatively large deviations also occur in the segment His36-Thr39. Further, main chain deviations between the K4 molcule A and K1 occur around residues His6-Arg13, His34, Gln41-Tyr43, Trp73, and Lys79-Lys80 while molecule B differ largely around the residues Thr32, Asn45, Asp57, and Ser70. The carboxy terminal interkringle tail of kringle 4 folds similar to the folding of the interkringle tail of fragment 1 structure [29]. The overall r.m.s. deviation of all main chain atoms, excluding the segment 58-61, between K4 and K1 is 0.95 Å and 0.83 Å for molecules A and B respectively.

Table 16. Differences in side chain electron density between the independent molecules A and B.

Side chain	Molecule A	Molecule B
Gln2	No density	Good
Asp3	Good	No density
His6	Good	Weak
Arg13	Weak	Good
Thr20	No density	Good
Gln26	Weak	Good
Ser29	Weak	Good
Ser30	Weak	Good
Pro33	Good	Weak
Gln37	Good	Weak
Glu41	No density	Good
Tyr43	Good	Weak
Asn55	Good	Weak
Asp59	Weak	Good
Lys60	Good	No density
Ser70	No density	Good
Trp73	Weak	Fair
Thr84	No density	Fair
Glu85	No density	Fair

VIII LYSINE BINDING SITE OF KRINGLE 4

It has been postulated that plasmin(ogen) and some of the kringles attach to blood clots by interacting with lysine side chains exposed in the polymerized fibrin matrix [13,20]. This is based on affinity chromatography studies which show that plasminogen and at least some of its kringle containing proteolytic fragments bind strongly to lysine-conjucated support gels and can be efficiently eluted with lysine and with other ω-amino carboxylic acid lysine analogs [6,98-99]. It has also been shown that the ω-aminocarboxylic acid binding sites of plasminogen are involved in fibrin binding since fibrin-plasminogen interaction is decreased by ω-aminocarboxylic acids [13].

The three main elastase fragments of plasminogen, kringle 1+2+3, kringle 4 and miniplasminogen(kringle 5+ light chain) were all shown to have affinity for fibrin polymer [6]; the strongest fibrin binding site was located on the kringle 1+2+3 fragment [6]. The fibrin-binding of all three fragments was decreased by 6-aminohexanoic acid [13]. This decrease was most pronounced with kringle 1+2+3 [13]. Later, it was shown that the fibrin binding site of the kringle 1+2+3 fragment resides on the kringle 1 domain [6] which also binds very strongly to plasminogen activation peptide, a potential effector of fibrinolysis activity. The fibrin-binding of kringle 4, on the other hand, is relatively weak, though it binds very strongly to lysine analogs [13]. Kringle 4 also has a relatively weak anti-plasmin binding site [13].

Because of their apparent roles in the regulation of fibrinolysis, the lysine binding sites have been the subject of a number of studies in recent years. Several investigations of the ligand specificity of the lysine binding sites have revealed the nature of the various residues involved in ligand binding. In the case of kringle 4, it has been shown that the residues Asp57, Asp59, Trp63, Phe65,Arg71, and Trp73 are either involved in ligand binding or are close to the binding site [9,24,27,32].

The lysine binding region is located on the surface of one of the oblate faces of the kringle, bounded by the residues 34-38,56-60,62-65, and 72-76 (Figure 26) [31-32]. This region essentially corresponds to the inner loop of the kringle sequence (Figure 16). The site is below the three dimensional loop formed by the C segment of the inner kringle loop and runs parallel to the antiparallel β-strand loop formed by the D seqment [31].

The binding site can be described as a relatively shallow depression, with cationic and anionic centers that extends from the Pro56-Lys60 peptide stretch which contains the Asp57 and Asp59, to a short segment where Arg72 neighbors His34 (Figure 26) [32]. The depression is lined by Trp63 near the Asp57-Asp59 stretch, which is consistent with the observed acid/base titrability of the Trp63 indole H2 singlet [25]. The depression is lined with Phe65 near the Arg72/His34 positively charged end. The indole ring of Trp73 is very close to the site on one side of the depression and is in a very favorable position to interact

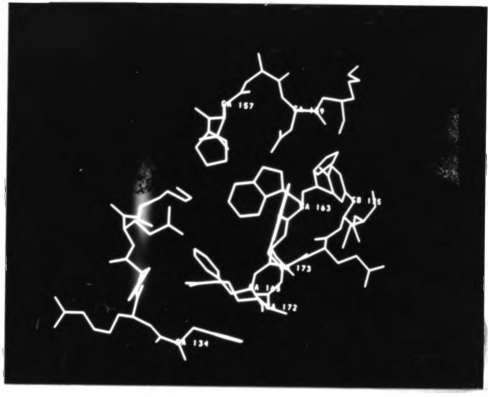


Figure 26. Lysine binding site of kringle 4 (molecule B).

with the hydrophobic part of the ligand [32]. It is interesting to note that the OG of the side chain of Tyr75 points towards the negatively charged end of the binding site, as shown previously by high field NMR experiments [32]. The three side chains Trp63, Phe65, and Trp73, thus expose a lipophilic surface which seperates the protruding anionic and cationic centers defined by residue pairs Asp57/Asp59 and His34/Arg72, respectively (Figure 26). These cationic and anionic centers can bind to the complementary polar end groups of the ω-aminoacid ligand [9,100]. The distance between the charges is consistent with a ligand of dipole length ~6.8 Å being optimal in terms of ligand binding [98].

It has been shown by NMR experiments that the presence of aromatic or aliphatic ligands induce large shifts in the ring resonances of several aromatic residues [23-24,101-102]. In the case of kringle 4, Trp63, Phe65, and Trp73 are the most perturbed by ligand presence [32]. This is in excellent agreement with the crystal structure; these are the aromatic residues which form the lipophilic surface in ligand binding (Figure 26). The ligand can be visualized as lying on the depression in an extended conformation [32]. Then it can interact very closely with the aromatic side chains mentioned above. It has been shown that the side chain of His34 which is very near Arg72 in the crystal structure also undergoes significant perturbation on ligand binding [32].

Somewhat removed from the center, but still an integral part of the binding pocket are Trp27 and Tyr75. The side chain of Trp28, besides lying proximal to Trp63, also establishes good contact with Leu48. It has been proposed that Leu48 centers at a hydrophobic core which is adjacent to, but not directly a part of the lysine-binding site [23,101]. It is very clear that the Trp28 indole group seperates Leu48 from the rest of the aromatic side chains at the binding site. The aromatic residues Tyr52 and Tyr43 are at the periphery of the binding site and this is in agreement with NOE experiments where it was shown that the resonances of these two residues are perturbed only slightly in presence of ligands [32]. These two aromatic residues also interact with Leu48, again in agreement with NOE experiments [23]. Residues Gln37, Lys38, and Pro56 are also in the vicinity of the binding site. The aromatic residues Tyr5, His6, and Tyr12 are further removed from the binding site. This has been suggested earlier based on NMR studies where the spectra of these residues are practically insensitive to ligand binding [32].

The binding site is relatively open and exposed, allowing for facile ligand access [32]. The feature is consistent with recent estimates of association rate constants for linear and cyclic ligands, suggesting that the kringle 4 complexation kinetics are essentially diffusion controlled [102]. This picture of the kringle 4 lysine-binding site is in agreement with a model of ligand binding to kringle 1 [34]. In kringle 1, the residues

Trp73,His34, and His36 are replaced by Tyr73, Arg34, and Phe36 respectively. Based on NMR experiments, and the crystal structure of the kringle of prothrombin fragment 1, it has been suggested that all three residues mentioned above participate in ligand binding in kringle 1 [34]. Thus it appears that in kringle 1, residues outside the inner loop segment also participate in ligand binding, unlike kringle 4. This suggests that although the binding site is conformed by a fixed set of residues, there appears to be some flexibility in ligand binding. The ligand binding affinity of kringle 1 is about twice as large as that for kringle 4 [21,100,103]. This could be due to the presence of Arg34 in kringle 1 which, along with Arg72, can provide a stronger cationic center towards various ligands [34]. The evidence for the participation of Arg34 in kringle 1 comes from chemical modification studies where it has been shown that modification of Arg34 with 1,2 cyclohexanediol abolishes both fibrin and lysine binding [11]. The area centered around His34 in kringle 4 may not be very important for lysine binding as glycosylation of Asn34 in kringle 4 of chicken plasminogen seems to have no effect on ligand binding [34-35].

A space-filling representation of the fibrin/lysine binding site of kringle 4 is shown in Figure 27. In this drawing, carbon atoms are a light grey color, oxygen atoms red, nitrogen atoms blue, and the sulfur atoms yellow. The most conspicuous aspect of this drawing is the segregation and apparent immiscibility of oxygen and nitrogen atoms

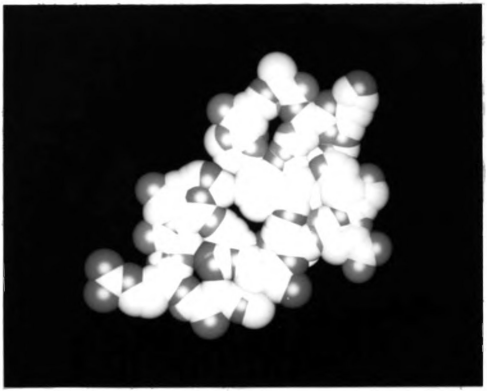


Figure 27. Space-filling view of the lysine binding site of kringle 4 (molecule B). The following color codes have been used for the atoms: carbon, light grey; nitrogen, blue; oxygen, red; sulfur, yellow.

surrounding the binding site on the surface [31]. The carboxylate group of the ligand can interact with the electropositive portion of this dipolar surface near the guanidinium group while the amino terminal of the ligand is near the negatively charged carboxylate groups Asp57/Asp59 of the protein [31]. The aromatic side chains of kringle 4 can then interact with the neutral methylene groups of the ligand. The dipolar surface could be of functional significance in recognizing at long range the marker of the fibrin ligand site and in aiding in the approximate orientation and alignment of the kringle [31]. This would lead to a better defined interaction upon close approach to fibrin and could also be a major contributing factor to the resultant stability of interaction of the docked fibrin-kringle complex.

As described in Chapter I, a model has been proposed for the lysine binding site of kringle 4 based on the three dimensional structure of kringle 1 of prothrombin fragment 1, high field NMR experiments and enery minimization [31]. The lysine binding site of the modeled kringle 4 and the binding site from the crystal structure are shown in Figure 28. It can be clearly seen that there is very good agreement between the model and the actual structure as observed crystallographically. The largest difference is in the orientation of Trp73. The side chain of Trp73 in the predicted model lies with the aromatic ring flat in the depression while in the crystal structure it is towards one side of the binding site. Apart from this Trp73, most of

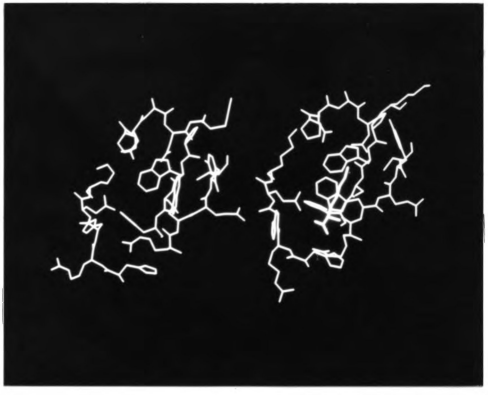


Figure 28. Comparision of the lysine binding sites. Left: Lysine binding site from the present study. Right: Lysine binding site from the modelling study (from [31]). In the model, the ligand ε-aminocaproic acid is also shown (white).

the other side chains near the binding site are in the same orientation and position in the predicted model and in the crystal structure. This is especially true for the residues Trp63, Phe65, Asp57, Asp59, Arg72, Tyr75, and His34, which have been implicated in lysine-binding. The residues Arg32, Lys60, and Lys35 which are away from the model also seem to have a different orientation in the model and the crystal structure.

IX CRYSTAL PACKING

The monoclinic crystal form of kringle 4 contains four molecules per unit cell or two crystallographically independent molecules per asymmetric unit. A schematic drawing of the four molecules in the unit cell is shown in Figure 15. The crystallographic 2_1 screw axis is shown by the conventional symbol and the non-crystallographic symmetry elements are indicated by the dotted lines. From an examination of the molecular arrangement in the crystal lattice (Figure 15), each independent kringle 4 molecule makes close contact with two other kringles giving rise to a three dimensional net work. This might be of relevance in kringle-kringle interactions in plasminogen which contains five kringles [4] and in apolipoprotein(a) which has 37 copies of kringle 4 and one copy of kringle 5 [5]. The molecular packing also gives rise to two "dimer" possibilities.

As described earlier in Chapter V, the choice of the asymmetric unit is quite arbitrary and since each kringle is in contact with two others, two different kinds of "dimers" are possible (Figure 15). Molecules A and B are related to one another by a non-crystallographic two fold axis. The axis is approximately parallel to the a^* axis. These two molecules are shown in Figure 29, viewed perpendicular to the non-crystallographic two fold axis.

In this "dimer", the maximal interaction between the two independent molecule involves the interkringle tail

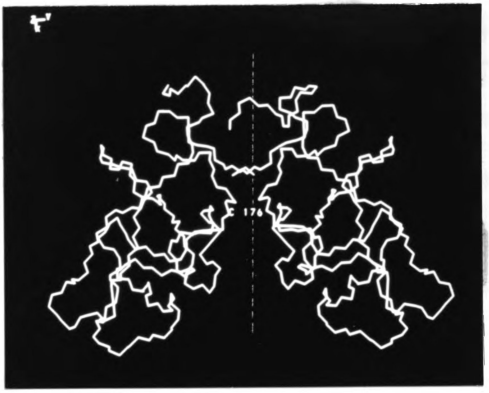


Figure 29. Cα, C, N main chain atoms of the two independent molecules A and B of the A-B "dimer". Non-crystallographic 2-fold rotation axis indicated.

portion and the segment Gln10-Thr15. The latter segment of molecule A participates in interactions with the corresponding residues of molecule B at the dimer interface. These residues are part of the flexible outer A segment of the kringle fold (Figure 16). This is very similar to the crystal packing of the kringle of fragment 1 where the majority of close contacts with neighboring molecules in the crystal involve the A loop [29]. In kringle 4, the amino terminal of molecule A is also involved in interactions with the amino terminal and the A loop of molecule B. The intermolecular contacts (less than 3.2 Å) at this dimer interface are listed in Table 17. The close packing of the two molecules can be seen from the space-filling representation shown in Figure 30.

Another kind of "dimer" is also possible. In this "dimer", the two molecules A and B' are related by an approximate rotation of 180 degrees, followed by a translation of about 22.0 Å along the c axis (Figure 15). This pseudo-screw axis runs almost parallel to the c axis of the crystal cell and intersects xy plane at x= 15.5/32.8 Å and y= 25.5/49.2 Å. The translational component of 22.0 Å is slightly less than one-half of the length of the c axis (0.48). Thus the packing of kringle molecules in the monoclinic crystals is of psuedo orthorhombic $P22_12_1$. The intensity of the reflections with l=odd along the c axis is generally weak compared to the even integer reflections, as expected from such a packing arrangement described above.

Table 17. Intermolecular contacts at the A-B "dimer" interface (< 3.2 Å) and the possible interaction(s).

Molecule A	Molecule B	Distance (Å)	Possible
Val1 N	Glu10 O	2.81	H-bond
Val1 CB	Asn45 ND2	3.12	
Glu2 N	Ser11 CA	2.99	
Glu2 O	Ser11 O	2.90	H-bond
Glu2 CB	Ser11 CB	2.81	
Ser11 CA	Val1 CG2	3.16	
Arg13 O	Arg13 O	2.79	H-bond/local 2-fold interaction
Arg13 O	Glu14 CA	2.53	local 2-fold interaction
Arg13 CB	Asp3 N	3.02	
Arg13 NH1	Thr15 O	2.40	H-bond
Glu14 CA	Arg13 O	2.53	local 2-fold interaction
Glu14 C	Arg13 O	2.93	
Thr15 OG1	Arg54 NH2	3.11	H-bond
Met50 CE	Tyr52 OH	3.03	
Tyr52 OH	Tyr52 OH	2.74	H-bond/local 2-fold interaction

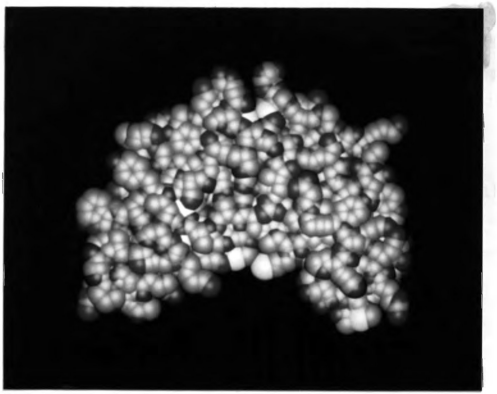


Figure 30. Space-filling view of the two independent molecules A and B of the A-B "dimer". Orientation same as Figure 29.

The "dimer" between A and B′ is shown in Figure 31 from which it can be seen that it is the outerloop segment B and the innerloop segment D which are involved in intermolecular interactions. There are very few short contacts involving the A and C segments of the kringles (Figure 16). The segment Ser29-Pro33 of the B loop of molecule A is involved in interactions with Gly22-Lys23 (A loop), Thr67-Asp68 (D loop), and Pro40-Asn42 (B loop) of molecule B′. All these residues belong to β-turns of the kringle structure (Table 15). The residues Thr67-Asp68 of the D loop of molecule A are also involved in interactions with the segment Thr39-Pro40 of the B loop of molecule B′. All intermolecular contacts at this dimer-interface which are less than 3.2 Å are listed in Table 18. A space-filling representation of this "dimer" is shown in Figure 32. The very tight packing of the molecules (66% protein fraction) can be clearly seen from Figure 32.

From the Tables 17 and 18 and the Figures 30 and 32, it is clear that in both the "dimeric species", there are many close contacts. Some of these interactions involve atoms which usually participate in hydrogen bonding. From Tables 17 and 18, it would suggest that the "dimer" A-B would have a better binding energy as it involves many more such interactions and the non-crystallographic relationship is a proper two-fold rotation which is fairly well known. On the other hand, the arrangement of kringles with a translation component in addition to a two-fold rotation might afford very efficient packing and favorable interactions in large

Figure 31. Cα, C, N main chain atoms of the two independent molecules A and B′ of the A-B′ "dimer". Non-crystallographic symmetry element (approximate 2_1 screw axis) indicated.

Table 18. Intermolecular contacts at the A-B′ "dimer" interface (< 3.2 Å) and the possible interaction(s).

Molecule A	Molecule B′	Distance (Å)	Possible
Thr21 O	Ser30 OG	3.19	H-bond
Gly22 O	Ser29 O	2.99	H-bond
Lys23 N	Ser29 O	2.73	H-bond
Lys23 CA	Ser29 O	2.30	
Lys23 CA	Ser29 C	2.92	
Lys23 CB	Ser29 CB	3.01	
Lys23 CB	Ser29 OG	3.20	
Lys23 CG	Ser29 OG	2.60	
Lys24 NZ	His36 O	3.10	H-bond
Lys24 NZ	Glu41 OE2	3.60	ion pair
Ser27 OG	Glu41 OE1	2.50	H-bond
Thr67 O	Pro40 CG	2.60	

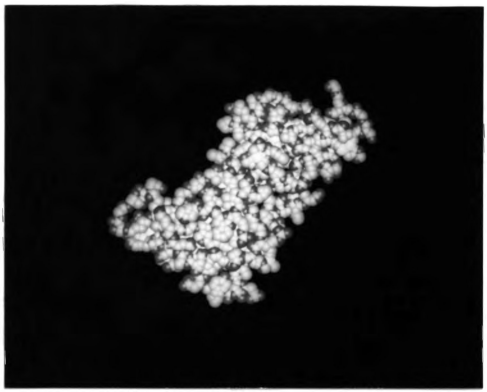


Figure 32. Space-filling view of the two independent molecules A and B' of the A-B' "dimer". Orientation same as Figure 31.

multi-kringle arrays such as occurs in apolipoprotein(a). At this stage, however, it is still difficult to reconcile the dimeric species described above to relevant kringle-kringle interactions in physiological systems. On the other hand, it is clear that the "dimers" will serve as models for the latter until a multi-kringle structure is determined.

In addition to the dimer interactions, there are a large number of contacts with neighboring molecules in the three dimensional crystal lattice. A schematic of the crystal packing is shown in Figure 33 with the symmetry elements and their positions indicated appropriately. In the case of molecule A, all the four loop segments (A,B,C, and D; Figure 16) are involved in interactions with neighboring molecules and the carboxy terminal of molecule A is also involved in packing interactions. In the case of molecule B, the residues from the outer loop segment B are involved in most of the interactions. In addition, other contacts occur between molecules A and B2, A′ and B, A′ and A2, A and A2 and B and B′.

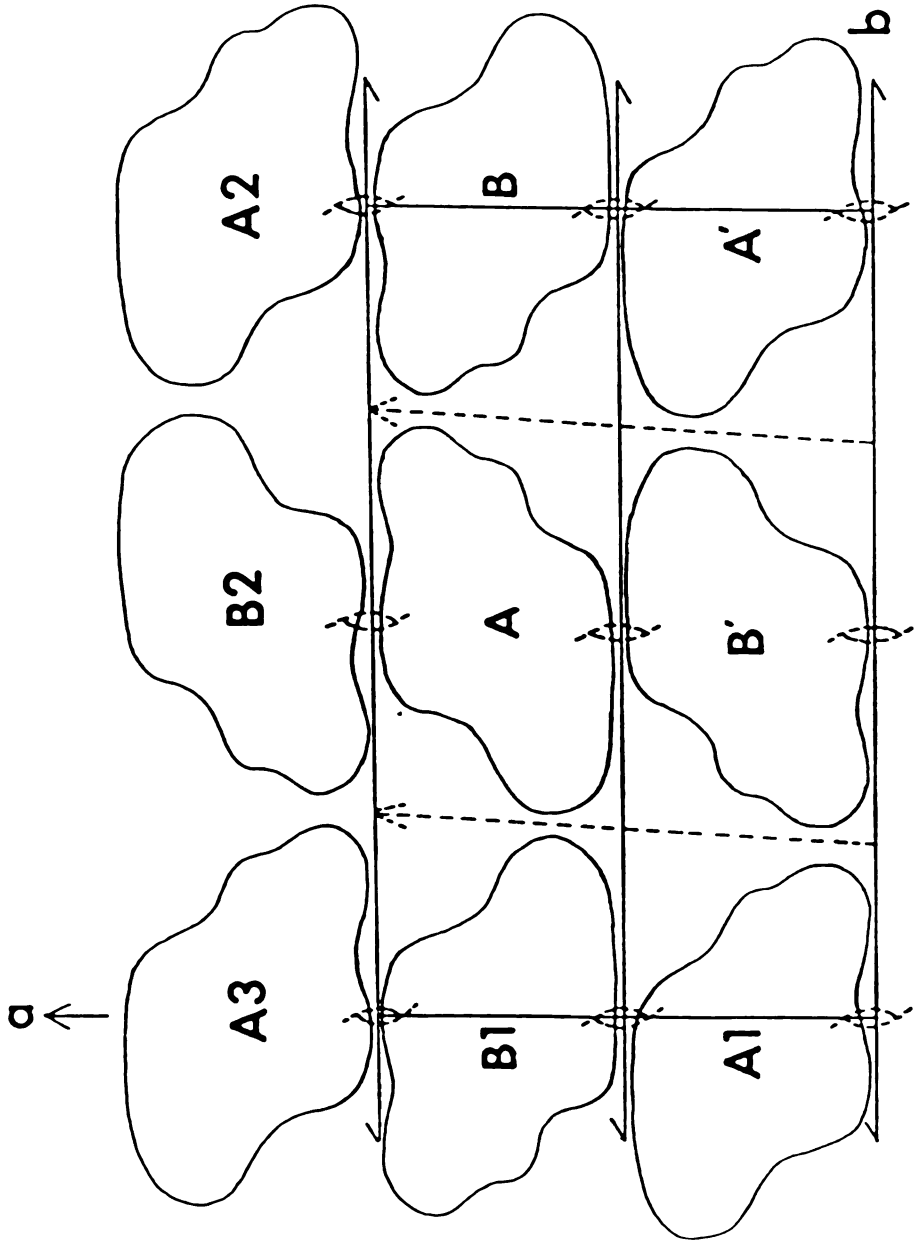


Figure 33. Schematic of crystal packing of kringle 4. Non-crystallographic symmetry elements are indicated by the dotted lines. Crystallographic symmetry elements and positions designated appropriately.

REFERENCES

LIST OF REFERENCES

1. Steffens,G.J.,Gunzler,W.A.,Otting,F.,Frankus,E., Flohe,L.,**Hoppe-Seylers's Z.Physiol.Chem.** 363: 1043-1058,1982.
2. Gunzler,W.A.,Steffens,G.J.,Otting,F.,Kim,S-M.A., Frankus,E.,Flohe,L. **Hoppe-Seylers's Z.Physiol.Chem.** 363:1155-1165,1982.
3. McMullen,B.A.,Fujikawa,K. **J.Biol.Chem.** 2: 5328-5340,1985.
4. Magnusson,S.,Peterson,T.E.,Sottrup-Jensen,L.,Claeys,H. in **Proteases and Biological Control.** Reich,E., Rifkin,D.B.,Shaw,E.,Eds. Cold Spring Harbor, New York: Cold Spring Harbor Laboratories:123-149,1975.
5. Pennica,D.,Holmes,W.E.,Kohr,W.J.,Harkins,R.N., Vehar,G.A.,Ward,C.A.,Bennet,W.F.,Yelerton,E., Seeberg,P.H.,Heynober,H.L.,Goeddel,D.V.,Collen,D. **Nature** 201:214-221,1983.
6. Sottrup-Jensen,L.,Claeys,H.,Zajdal,M.,Petersen,T.E., Magnusson,S. **Prog.Chem. Fibrinolysis Thrombolysis** 3: 191-209,1978.
7. Mclean,J.W.,Tomlinson,J.E.,Wun-Jing,K.,Eaton,D.L., Chen,E.Y.,Fless,G.M.,Scanu,A.M.,Lawn,R.M. **Nature** 330:132-137,1987.
8. Patthy,L.,Trexler,M.,Vali,Z.,Banyai,L.,Varadi,A. **FEBS Lett.** 171:131-136,1984.

8a. Holland,S.K.,Harlos,K.,Blake,C.C.F. **EMBO J.** 6: 1875-1880,1987.

9. Vali,Z.,Patthy,L. **J.Biol.Chem.** 259:13690-13694, 1984.

10. Lerch,P.G.,Rickli,E.E.,Largier,W.,Gillenssen,D. **Eur.J.Biochem.** 107:7-13,1980.

11. Thorsen,S.,Clemmensen,I.,Sottrup-Jensen,L.,Magnusson,S. **Biochim.Biophys.Acta.** 668:377-387,1981.

12. Trexler,M.,Vali,Z.,Patthy,K. **J.Biol.Chem.** 257: 7401-7406,1982.

13. Varadi,Z.,Patthy,L. **Biochem.Biophys.Res.Commun.** 103:97-102,1981.

14. Esmon,C.T.,Jackson,C.M. **J.Biol.Chem.** 249: 7791-7796,1974.

15. Bajaj,S.P.,Butkowski,R.J.,Mann,K.G. **J.Biol.Chem.** 250:2150-2156,1975.

16. van Zonneveld,A.J.,Veerman,H.,Pannekoek,H. **Proc.Natl.Acad.Sci., U.S.A.** 83:4670-4675,1986.

17. Wallen,P. in **Progress in Chemical Fibrinolysis and Thrombolysis** (Davidson,J.F.,Rowan,R.M.,Samama,M.M., Desnnoyers,P.C. eds.) Raven Press, New York 3: 167-181,1978.

18. Groskopf,W.R.,Hsieh,B.,Summaria,L.,Robbins,K.C. **J.Biol.Chem.** 244:359-365,1969.

19. Magnusson,S.,Sottrup-Jensen,L.,Petersen,T.E.,Dudek-Wojciechowska,G.,Claeys,H. in **Proteolysis and Physiological Regulation** (Ribbons,D.W.,Brew,K. eds.) **Academic Press**, New York: 203-238,1976.

20. Wiman,B.,Collen,D. **Nature** 272:549-550,1978.

21. Hochschwender,S.M.,Laursen,R.A.,De Marco,A.,Llinas,M. **Arch.Biochem.Biophy.** 223:58-67,1983.

22. Trexler,M.,Banyai,L.,Patthy,L.,Pluck,N.D., Williams,R.J.P. **Eur.J.Biochem.** 152:439-446,1985.

23. Llinas,M.,Motta,A.,De Marco,A.,Laursen,R.A. **J.Biosci.** 8:121-139,1985.

24. De Marco,A.,Laursen,R.A.,Llinas.,M. **Arch.Biochem.Biophys.** 224:727-741,1986.

25. De Marco,A.,Pluck,N.D.,Banyai,L.,Trexler,M., Laursen,R.A.,Patthy,L.,Llinas,M.,Williams,R.J.P. **Biochemistry** 24:748-753,1985.

26. De Marco,A.,Laursen,R.A.,Llinas,M. **Biochem.Biophys. Acta.** 787:275-280,1984.

27. Hochschwender,S.M.,Laursen,R.A. **J.Biol.Chem.** 256: 11172-11176,1981.

28. Ramesh,V.,Gyenes,M.,Patthy,L.,Llinas,M. **Eur.J.Biochem.** 159:581-595,1986.

29. Tulinsky,A.,Park.C.H.,Skrzypczak-Jankun,E. **J.Mol.Biol.** 202:885-901,1988.

30. Park,C.H.,Tulinsky,A. **Biochemistry** 25:3977-3982, 1986.

31. Tulinsky,A.,Park,C.H.,Mao,B.,Llinas,M. **Proteins: Structure,Function, and Genetics** 3:85-96,1988.

32. Ramesh,V.,Petros,M.Llinas,M.,Tulinsky,A.,Park,C.H. **J.Mol.Biol.** 198:481-498,1987.

33. Thewes,T.,Ramesh,V.,Simplaceanu,E.L.,Llinas,M. **Biochem.Biophys.Acta.** 912:254-269,1987.

34. Motta,A.,Laursen,R.Llinas,M.,Tulinsky,A.,Park,C.H. **Biochemistry** 26:3827-3836,1987.

35. Gyenes,M.,Patthy,L. **Biochem.Biophys.Acta.** 832: 326-330,1985.

36. Blundell,T.L.,Johnson,L.N. **Protein Crystallography, Academic Press,** New York,1976.

37. **Rossmann,M.G. Ed. Molecular Replacement Method,** Gordon and Breach, New York,1972.

38. Argos,P.,Rossmann,M.G. **Acta Cryst.** A30:672-677, 1974.

39. Rayment,I. **Acta Cryst.** A39:102-116,1983.

40. Bloomer,A.C.,Champness,J.N.,Bricogne,G.,Staden,R., Klug,A. **Nature** 276:362-368,1978.

41. Harrison,S.C.,Olson,A.J.,Schutt,C.E.,Winkler,F.K., Bricogne,G. **Nature** 276:368-373,1978.

42. Abad-Zapatero,C.,Abdel-Meguid,S.S.,Johnson,J.E., Leslie,A.G.W.,Rayment,I.,Rossmann,M.G.,Suck,D., Tsukihara,I. **Nature** 286:33-39,1980.

43. Abad-Zapatero,C.,Abdel-Meguid,S.S.,Johnson,J.E., Leslie,A.G.W.,Rayment,I.,Rossmann,M.G.,Suck,D., Tsukihara,I. **Acta Cryst.** B37:2002-2018,1981.

44. Unge,T.,Liljas,L.,Strandberg,B.,Vaara,I.,Kannan,K.K., Fridborg,K.,Nordman,C.E.,Lentz,P.J.Jr. **Nature** 285:373-377,1980.

45. Bricogne,G. **Acta Cryst.** A30:395-405,1974.

46. Bricogne,G. **Acta Cryst.** A32:832-847,1976.

47. Rossmann,M.G.,Blow,D.M. **Acta Cryst.** 15:24-31,1962.

48. Rossmann,M.G.,Blow,D.M. **Acta Cryst.** 17:338-348, 1964.

49. Crowther,R.A.,Blow,D.M. **Acta Cryst.** 23:544-548, 1967.

50. Joynson,M.A.,North,A.C.T.,Sarma,V.R.,Dickerson,R.E., Steinrauf,L.K. **J.Mol.Biol.** 50:137-142,1970.

51. Dodson,E.,Harding,M.M.,Hodgkin,D.C.,Rossmann,M.G. **J.Mol.Biol.** 16:227-241,1966.

52. Tollin,P. **J.Mol.Biol.** 45.:481-490,1969.

53. Fujinaga,M.,Read,R.,Sielecki,A.,Ardelt,W., Laskowski,M.Jr.,James,M.N.G. **Proc.Natl.Acad.Sci., U.S.A.** 79:4868-4872,1982.

54. McPhalen,C.A.,Svendsen,I.,Jonassen,I.,James,M.N.G. **Proc.Natl.Acad.Sci., U.S.A.** 82:7242-7246,1985.

55. Dijkstra,B.W.,Van Nes,G.J.H.,Kalk,K.H.,Brandenburg,N.P.,Hol.,W.G.J.,Drenth,J. **Acta Cryst.** B38: 793-799,1982.

56. Dacheng,W.,Bode,W.,Huber,R. **J.Mol.Biol.** 185: 595-624,1985.

57. Phizackerly,R.P.,Wishner,B.C.,Bryant,S.H.,Amzel,L.M., Lopez De Castro,J.A.,Poljak,R. **J. Mol.Immunol.** 16: 841-850,1979.

58. Schirmer,T.,Huber,R.,Schneider,M.,Bode,W.,Miller,M., Hackert,M.L. **J.Mol.Biol.** 188:651-676,1986.

59. Stout,G.H.,Jensen,L.H. **X-ray Structure Determination, A Practical Guide**, The Macmillan Company, New York,1968.

60. Lattman,E.E.,Love,W.E. **Acta Cryst.** B26:1854-1857, 1970.

61. Fitzgerald,P.M.D. **J.Appl Cryst.** 21:273-278,1988.

62. Huber,R. **Acta Cryst.** 19:353-356,1965.

63. Nordman,C.E.,Nakatsu,K. **J.Amer.Chem.Soc.** 85: 353-357,1963.

64. Steigemann,W. **Ph.D. Thesis, TU Muenchen**,1974.

65. Goldstein,H. **Classical Mechanics**, Addison Wesley Publishing Company, Massachusets,1981.

66. **Patterson,A.L. International Tables for X-ray Crystallography**, Kynoch Press,Birmingham 2:52,1959.

67. Tollin,P.,Main,P.,Rossmann,M.G. **Acta Cryst.** 20: 404-407,1966.

68. Rao,S.R.,Jih,J-H.,Hartsuck,J.A. **Acta Cryst.** A36: 878-884,1980.

69. Tollin,P. **Acta Cryst.** 21:613-614,1966.

70. Langs,D.A. **Acta Cryst.** A31:543-550,1975.

71. Nordman,C.E.,Schilling,J.W. in **Crystallographic Computing** (Ahmed,F.R.,Hall,S.R.,Huber,C.P. eds.), Munksgaard,Coppenhagen 110-114,1969.

72. Cutfield,J.T.,Cutfield,S.M.,Dodson,E.J.,Dodson,G.G., Sabesan,M. **J.Mol.Biol.** 87:23-30,1974.

73. Ral,A.D. **Acta Cryst.** A33:423-425,1977.

74. Nixon,P.E.,North,A.C.T. **Acta Cryst.** A32: 320-325,1976.

75. Harada,Y.,Lifchitz,A.,Berthou,J.,Jolles,P. **Acta Cryst.** A37:398-406,1981.

76. Rius,J.,Miravitlles,C. **Acta Cryst.** A42: 402-404,1986.

77. Fujinaga,M.,Read,R. **J.Appl.Cryst.** 20:517-521,1987.

78. Cygler,M.,Anderson,W.F. **Acta Cryst.** A44: 300-308,1988.

79. Rossmann,M.G. in **Theory and Practice of Direct Methods in Crystallography** (Ladd,M.F.C.,Palmer,R.A., eds.), Plenum Press,New York, 361-417,1980.

80. Bott,R.,Sarma,R. **J.Mol.Biol.** 106:1037-1046,1976.

81. Hendrickson,W.A.,Ward,K.B. **Acta Cryst.** A32: 778-780,1976.

82. Ward,K.B.,Wishner,B.C.,Lattman,E.E.,Love,W.E. **J.Mol.Biol.** 98:161-177,1975.

83. Winn,E.S.,Hu,S.P.,Hochschwender,S.M.,Laursen,R.A. **Eur.J.Biochem.** 257:7401-7406,1982.

84. Mulichak,A.,Park,C.H.,Tulinsky,A.,Llinas,M. submitted to **J.Biol.Chem.**

85. Matthews,B.W. **J.Mol.Biol.** 33:491-497,1968.

86. Tulinsky,A.,Mani,N.V.,Morimoto,C.N.,Vandlen,R.L. **Acta Cryst.** B29:1309-1322,1973.

87. Wyckoff,H.W.,Doscher,M.,Tsernoglou,D.,Inagami,T., Johnson,L.N.,Hardman,K.D.,Allewell,N.M.,Kelly,D.M., Richards,F.M. **J.Mol.Biol.** 27:563-578,1967.

88. North,A.C.T.,Phillips,D.C.,Mathews,F.S. **Acta Cryst.** A24:351-359.

89. Yeates,T.O. **Acta Cryst.** A44:142-144,1988.

90. Fisher,R.G.,Sweet,R.M. **Acta Cryst.** A36: 755-760,1980.

91. Murray-Rust,P. **Acta Cryst.** B29:2559-2566,1973.

92. Britton,D. **Acta Cryst.** A28:296-297,1972.

93. Rees,D.C. **Acta Cryst.** A36:578-581,1980.

94. Jones,T.A. In omputational Crystallography Ed. Sayre,D. Oxford:Clarendon Press 303-308,1982.

95. Wilson,A.J.C. **Acta Cryst.** 2:318-321,1949.

96. Castellino,F.J.,DeSerrano,V.S.,Powell,J.R., Johnson,W.R.,Beals,J.M. **Arch.Biochem.Biophys.** 247:312-320,1986.

97. Trexler,M.,Patthy,L. **Proc.Natl.Acad.Sci.U.S.A.** 80: 2457-2461,1983

98. Winn,E.S.,Hu,S.P.,Hochschwender,S.M.,Laursen,R.A. **Eur.J.Biochem.** 104:579-586,1980.

99. Deutsch,D.,Mertz,E.T. **Science** 170:1095-1096,1970.

100. Lerch,P.G.,Rickli,E.E. **Biochim.Biophys.Acta.** 625: 374-378,1980.

101. De Marco,A.,Laursen,R.A.,Llinas,M. **Biochim.Biophys. Acta.** 827:369-380,1985.

102. De Marco,A.,Petros,A.M.,Laursen,R.A.,Llinas,M. **Eur.Biophys.J.** 14:359-368,1987.

103. De Marco,A.,Hochschwender,S.M.,Laursen,R.A.,Llinas,M. **J.Biol.Chem.** 257:12716-12721,1982.