Molecular Replacement in Action.
The NK1 fragment crystal structure solution by Molecular Replacement method

by Dima Chirgadze

9th July 2001

Abstract

When identical or similar structures exist in different crystallographic environments (like space group, unit cell parameters), similarities between their diffraction patterns, which are directly associated with the Fourier transforms of these structures, would be expected. The technique of Molecular Replacement in protein X-ray crystallography utilises this similarity in order to derive the initial phase information. This paper introduces the theory underlying the Molecular Replacement method as applied to the solution of protein crystal structures. It also describes the application of this method to the NK1 fragment crystal structure solution using the NMR model of HGF/SF (Hepatocyte Growth Factor/Scatter Factor) N-domain and the crystal structure of human plasminogen kringle 1 domain as molecular replacement search probes.
Contents

1 The theory of the Molecular Replacement method 3
   1.1 Basic concepts of the molecular replacement method ............................................. 3
   1.2 Basic algebra of X-ray crystallography ........................................................................ 3
      1.2.1 Structure factors ................................................................................................. 3
      1.2.2 Electron density and Fourier transform ................................................................ 4
      1.2.3 Patterson function .............................................................................................. 4
   1.3 The Rotation function ............................................................................................ 6
      1.3.1 The Rossmann-Blow Rotation function .................................................................. 6
      1.3.2 The angular definition of the rotation matrix ....................................................... 7
      1.3.3 The Crowther fast rotation function ...................................................................... 8
   1.4 The translation function ....................................................................................... 9
      1.4.1 The T and T1 translation functions .................................................................... 10
      1.4.2 The T2 Translation Function ............................................................................. 10
      1.4.3 The TO/O Translation Functions ....................................................................... 11
   1.5 Factors affecting the molecular replacement solution ........................................... 11

2 The application of the molecular replacement method to the structure solution of the NK1 fragment. 13
   2.1 Molecular replacement search probes ................................................................. 13
      2.1.1 The kringle domain probe ................................................................................... 13
      2.1.2 Molecular replacement search probes: the N-domain probe .............................. 14
   2.2 The use of the 'Automated Molecular Replacement' (AMoRe) package in the solution of the NK1 fragment structure ............................................................... 16
      2.2.1 The preparation steps .......................................................................................... 16
      2.2.2 The rotation function search .............................................................................. 18
      2.2.3 The translation function search .......................................................................... 20
      2.2.4 The integration radius of kringle and N-domain search probes ....................... 21
      2.2.5 Results of the kringle probe rotation search ....................................................... 21
      2.2.6 Results of the N-domain probe rotation search .................................................. 22
      2.2.7 The kringle and the N-domain probes translation function search .................... 22
      2.2.8 Refinement of the Molecular Replacement Solutions ......................................... 25
   2.3 Molecular replacement solution validation ........................................................... 26
      2.3.1 Analysis of the crystal packing .......................................................................... 26
      2.3.2 Non-crystallographic symmetry and self-rotation function ............................... 26

3 Conclusions .............................................................................................................. 28
1 The theory of the Molecular Replacement method

The number of proteins that have been structurally characterised, either in the solid or in the liquid phase (by means of X-ray diffraction or NMR techniques respectively), has increased exponentially over the last 40 years. The large body of information concerning the three-dimensional structure of macromolecules differing in their folding properties, as well as in their biological activity, has been organised in the PDB\(^1\). From this database the coordinates of protein structures can be readily retrieved.

As the number of structures present in the PDB increases, the possibility that an unknown structure has some features in common with one or more already characterised proteins (not necessarily involved in similar biological processes) becomes greater. Often nowadays, one may wish to determine the crystal structure of a molecule when the structure of a similar or homologous molecule is already available. The structure of enzyme-effector/inhibitor complexes may be of interest, for example, or changes in protein folding and/or packing caused by specific mutations of amino acid sequences.

In cases where the crystal under investigation is isomorphous with the known one (having the same space group and cell constants within experimental error), analysis can proceed directly by difference Fourier methods. However, more commonly, isomorphism does not exist, and it becomes necessary to seek other ways of utilising the known structure information to facilitate the target structure determination. To this end, the Molecular Replacement (MR) method has proved to be particularly successful (reviewed in: \[24, 27\]).

1.1 Basic concepts of the molecular replacement method

The prerequisites for using the MR method are:

- an observed diffraction pattern – intensities – for the unknown structure, or the target;
- the atomic coordinates of an homologous protein structure, or the probe.

The MR task involves positioning the probe within the unit cell of the target crystal in such a way that the theoretical diffraction pattern that would result from this model closely matches the experimental one.

If only one molecule is present in the asymmetric unit, then six parameters (three rotational and three translational), which fully describe how the probe is placed in the unit cell, would have to be determined. In principle, we could simply do a search on these six parameters to determine the position of the probe that gives the best agreement between observed and calculated structure factor. However, such a search would be computationally too demanding, although in special cases (in which symmetry reduces the number of parameters) such searches have been successful \[1\]. From the theoretical analysis of the properties of the Patterson function \[9, 22\] it became obvious that such a six-parameter search could be reduced to two three-dimensional problems. The first is the determination of the correct orientation of the probe, and the second is the determination of the position of the correctly-oriented molecule within the unit cell.

In figure 1 a pictorial representation of the MR problem is shown. We have a probe molecule \(A\) and the unknown molecule \(A'\) similar to \(A\). The position of \(A'\) is different from \(A\). To superimpose the molecule \(A\) with \(A'\) we have, firstly, to apply the rotation \(R\), and then the translation \(T\). Therefore, the main aim of the MR method is to find these two operators, or, in other terms, to solve the Rotation and the Translation functions.

Once the probe is correctly positioned, phases can be calculated from this model and combined with the observed amplitudes to give an approximate Fourier synthesis of the structure of interest. This model will then have to be carefully analysed and, through several refinement and model rebuilding steps, modified to account for the differences between the probe and the target molecule, in order to provide a model that maximises the agreement with the experimental data.

1.2 Basic algebra of X-ray crystallography

1.2.1 Structure factors

The structure factor for the hkl-reflection is a complex number defined by:\[fff\]

\(^1\)Brookhaven Protein Structure Data Bank
A = \begin{bmatrix} R \end{bmatrix} + T

A' = A[R] + T

\[ F(\vec{h}) = \sum_{\text{atoms}} f_j e^{2\pi i \vec{h} \cdot \vec{\alpha}_j} = |F(\vec{h})| e^{i\alpha(\vec{h})} \] (1)

This is a simple summation which extends over the whole set of atoms \( j \) in the unit cell; \( \vec{\alpha}_j \) and \( \vec{h} \) are the vectors that describe the fractional coordinates of the \( j \)-th atom \( ((x_j, y_j, z_j) \) and position \( (hkl) \) of the \( hkl \)-reflection. The \( f_j \) is the scattering factor of atom \( j \) and depends on the kind of atom, and the diffraction angle of the corresponding reflection. The structure factor equation can also be written in terms of structure factor amplitudes \( |F(\vec{h})| \) and phases \( \alpha(\vec{h}) \).

### 1.2.2 Electron density and Fourier transform

The complex exponential function in equation 1 is periodic and limited between \(-1, 1\) for its real part and \(-i, i\) for the imaginary part. To such a periodic function we can apply a Fourier transformation, the result of which represents the electron density in the point of the real space described by the vector \( \vec{x}(x, y, z) \):

\[ \rho(\vec{x}) = \frac{1}{V} \sum_{\vec{h}} F(\vec{h}) e^{-2\pi i \vec{h} \cdot \vec{x}} \] (2)

where \( V \) is the volume of the unit cell and the \( F(\vec{h}) \) are the Fourier coefficients. The diffraction pattern, then, is the Fourier transform of the protein structure in the crystal, and the structure (i.e. the electron density) is, in turn, the Fourier transform of the diffraction pattern.

### 1.2.3 Patterson function

The Patterson function is given by the product of the electron density at positions \( \vec{x} \) and \( \vec{x} + \vec{u} \), where \( \vec{u}(u, v, w) \) is the translation vector that relates two identical copies of the electron density map. It takes the form:

\[ P(\vec{u}) = \int \rho(\vec{x}) \rho(\vec{x} + \vec{u}) \, dv \] (3)
Using equation 2 for the electron density, the Patterson function can be rewritten as:

\[
P(\mathbf{u}) = \frac{1}{V^2} \sum_{h} \sum_{h'} F(\mathbf{h}) F(\mathbf{h'}) e^{-2\pi i \mathbf{h} \cdot \mathbf{u}} \int_{V} e^{-2\pi i (\mathbf{h} \cdot \mathbf{u})} d\mathbf{v}
\]  

(4)

The integration is equal to zero, unless \( \mathbf{h} = -\mathbf{h'} \) when it is equal to \( V \). Thus, equation 4 becomes:

\[
P(\mathbf{u}) = \frac{1}{V} \sum_{h} \sum_{h'} F(\mathbf{h}) F(\mathbf{h'}) e^{-2\pi i \mathbf{h} \cdot \mathbf{u}} = \frac{1}{V} \sum_{h} |F(\mathbf{h})|^2 e^{-2\pi i \mathbf{h} \cdot \mathbf{u}}
\]  

(5)

From equation 5 it can be seen that the Patterson function can be calculated using the observed amplitudes without knowing the phases. In fact, the Patterson function is a Fourier transform of the \( |F(\mathbf{h})|^2 \). The Patterson function will show a peak when the vector relating the two identical copies of the electron density, coincides with the distance between two atoms in the unknown structure. In other words, when the Patterson function is calculated over a broad range of translation vectors \( \mathbf{u}(u, v, w) \), it will correspond to the distribution of the interatomic vectors in the unit cell.

The Patterson function can be represented graphically (figure 2). Consider a two-dimensional example. There are two peaks of electron density with widths \( D_1 \) and \( D_2 \), separated by the distance \( U \). The corresponding Patterson peak will be on the distance \( U \) from the origin with width \( D_1 + D_2 \). The Patterson function can also be described as the convolution of the structure and its inverse. Convolution involves placing one object at every point of another. The example of Patterson function construction by convolution is presented in figure 3.

To summarise, the Patterson function has the following properties:

- the map has peaks at the ends of vectors \( \mathbf{u} \) equal to vectors between atoms in the real cell;
- for every pair of atoms in the real cell there exists a unique peak in the Patterson map, i.e. if there are \( n \) atoms in the real cell, then there are \( n^2 - n \) peaks in the Patterson map. There are \( n \) peaks at the origin, which has a height \( \sum f_j^2 \) (see equation 1);
- a Patterson map is always centrosymmetric;
- a screw axis in a real cell becomes a normal axis in a three-dimensional Patterson space;
- symmetry elements cause concentration of peaks along specific Harker lines or planes.
Figure 3: Formation of the Patterson peaks by the convolution. Schematic diagram showing the formation of Patterson peaks from a two-dimensional structure and its inverse (a). The convolution involves placing the inverse of the structure at every point of the structure itself, thus producing the Patterson function map (b).

1.3 The Rotation function

1.3.1 The Rossmann-Blow Rotation function

Consider a structure of two identical molecules, which are in different orientations. The Patterson function of such a structure will consist of three sets of vectors:

1. a set of self-Patterson vectors of one molecule. These are interatomic vectors which can be formed within that molecule;
2. a set of self-Patterson vectors of the other molecule. This set will be identical to the first, but they will be rotated away from the first due to different orientation;
3. finally, there will be the cross-Patterson vectors, or set of interatomic vectors which can be formed from one molecule to another. The self-Patterson vectors of two molecules will all be located in a volume extending from the origin by the overall dimensions of the molecules. Some or all of the cross-Patterson vectors will lie outside this volume.

Suppose that the Patterson function is now superimposed onto a rotated version of itself. There will be no particular agreement except when one set of self-Patterson vectors of one molecule has the same orientation as the self-Patterson vectors from the other molecule. In this position the agreement or 'overlap' between the two will be at a maximum. The function describing this overlap has been termed by Rossmann and Blow as the Rotation function [22], and is defined as follows:

\[ RF(\vec{R}) = \int_U P_1(\vec{\xi}) P_2(\vec{\xi}) dV \]  

(6)

\( P_1 \) and \( P_2 \) are Patterson functions, \( \vec{R} \) is the rotation operator that relates the coordinate system of \( P_2 \) to that of \( P_1 \), and \( U \) is a volume of integration, usually spherical, centred at the origin. A maximum in the rotation function \( RF(\vec{R}) \) indicates a potentially correct orientation of the search probe in the target cell.

From equation 5, the rotation function can be rewritten as:

\[ RF(\vec{R}) = \frac{U}{V^3} \sum_h \sum_p |F(\vec{h})|^2 |F(\vec{p})|^2 G_{hlp} \]  

(7)
where $F(\mathbf{h})$ and $F(\mathbf{p})$ are the structure factors of the target and the probe molecules respectively, $G_{\text{int}}$ is the so-called **interference function**. For an integration sphere of radius $r$ it equals:

$$G_{\text{int}} = \frac{3(\sin2\pi Hr - 2\pi Hr \cos2\pi r)}{(2\pi Hr)^3}$$

(8)

where $H = |\mathbf{h} + \mathbf{p}|$. $\mathbf{p}$ is the vector that describes the position of vector $\mathbf{p}$ after applying the $\mathbf{R}$ rotation matrix. In other words, $H$ represents the distance of the point with coordinates $(h_1 + h'_1, (h_2 + h'_2), (h_3 + h'_3))$ from the origin of reciprocal space. The maximum value of the interference function is 1.00 (see figure 4), and it is never greater than 0.086 outside the range $-0.725 < x < 0.725$. Hence all terms in equation 8 for which $|Hr| > 0.725$ may be neglected if $U$ is assumed to be a sphere with radius $r$.

Two major problems affect the efficiency of the Rossmann-Blow rotation function:

- the relatively long CPU time required to compute it;
- the lack of precision owing to truncation errors associated with the $G_{\text{int}}$ function [24].

### 1.3.2 The angular definition of the rotation matrix

The rotation matrix $\mathbf{R}$ can be defined with respect to two alternative geometrical systems. The most common one makes use of the set of Eulerian angles, $\theta_1$, $\theta_2$ and $\theta_3$, as the variables of this matrix. Conventionally, $\theta_1$ is a rotation angle about the $z$ axis of an orthogonal system, $\theta_2$ is the rotation about the moved $x$ axis, and $\theta_3$ is the rotation about the moved $z$ axis. These rotations are shown in figure 5a. For all three angles, positives values are defined by anti-clockwise rotations when looking down the positive axes toward the origin.

Some of the commonly available MR programs use a slightly different definition of Eulerian angles: $\alpha$, $\beta$ and $\gamma$. These angles behave in the same way as $\theta_1$, $\theta_2$ and $\theta_3$, except that the second rotation axis is $y$ instead of $x$. The program ROTMAT from the CCP4 crystallographic package [4] can perform conversions of rotation angles to any other format used by different programs. The other geometrical system utilises polar angles figure 5. Here, any rotation can be accomplished by an appropriate spin about a properly chosen axis. The angles $\phi$ and $\psi$ specify the longitude and latitude of this axis, while $\chi$ quantifies the spin about it. This system is useful whenever the order or direction of the rotation axis can be anticipated.
When the Patterson function shows rotational symmetry, the rotation function also has symmetry, and the full range of angles need not be explored [21]. The full description of matrix $R$ elements in terms of Eulerian and spherical polar angles can be found in [15].

### 1.3.3 The Crowther fast rotation function

The rotation function is generally used to correlate a spherical volume of a given Patterson density with a rotated version of another Patterson density. Since we are dealing with rotations of spherical volumes, it is likely that a more natural form for the rotation function than that given by Rossmann and Blow should be derivable. This new form was described by Crowther [5] – the fast rotation function (nowadays, it simply called the rotation function).

Crowther’s fast rotation function makes use of spherical coordinates $(r, \theta, \phi)$ instead of Cartesian ones $(x, y, z)$. The equation for the electron density (equation 2) in spherical coordinates will look like this:

$$\rho(r, \theta, \phi) = \frac{1}{V} \sum_{h} F(h) e^{-2\pi i R \cos \gamma}$$

Here, $\gamma$ is the angle between the vectors $x(r, \theta, \phi)$ and $\tilde{h}(R, \Theta, \Phi)$.

Crowther has expanded the electron density in normalised spherical Bessel functions $J_l$ and normalised spherical harmonics $Y_l^m$ (equation 10). The normalised spherical harmonics are given in terms that are associated by the Legendre polynomials [14].

$$\rho(r, \theta, \phi) = \sum_{l} \sum_{m} \sum_{n} p_{lmn} J_l(2\pi k r) Y_l^m(\theta, \phi)$$

Here the coefficients $p_{lmn}$ are represented by:

$$p_{lmn} = \sum_{h} F(h) T_{lmn}(\tilde{h})$$

where $T_{lmn}$ is the Fourier transform (equation 2) of the $J_l$ and $Y_l^m$ expansion functions. Thus, we can also expand in a similar way the Patterson function (Equation 3) within the spherical volume $r < a$ in the following form:

$$P(r, \theta, \phi) = \sum_{l} \sum_{m} \sum_{n} a_{lmn} J_l(2\pi k r) Y_l^m(\theta, \phi)$$
Here \( J_{ln} \) is such that \( J_{ln}(2\pi k_n a) = 0 \ (n = 1, 2, \ldots) \), where \( a \) is the radius of the sphere of the Patterson density. After some mathematical transformations the equation for the fast rotation function can be obtained [5], and will take the form:

\[
RF(\theta_1; \theta_2; \theta_3) = \sum_{m} \sum_{m'} A_{mm'}(\theta_2) e^{i(m'\theta_2 + m\theta_1)}
\]

(13)

where

\[
A_{mm'}(\theta_2) = \sum_{l} C_{lmm'} d_{lmm'}(\theta_2)
\]

(14)

and

\[
C_{lmm'} = \sum_{n} a_{lmmn} b_{lmmn}
\]

(15)

The \( a_{lmmn} \) and \( b_{lmmn} \) in equation 15 are the coefficients for the spherical expansions of two Patterson functions as described in equation 12, and they can be calculated using equation 11 by substituting \( F' \)'s with \( F \). In Equation 14, \( d_{lmm'} \) represents the matrix elements related to the rotation of spherical harmonics [5]. Thus, the algorithm of Crowther’s fast rotation function is reduced to the following operations:

- The \( a_{lmmn} \) coefficients are computed using the target structure factor amplitudes, which are observed from the X-ray diffraction pattern;
- The \( b_{lmmn} \) coefficients are computed for the probe molecule. If these coefficients equal to \( a_{lmmn} \) the self-rotation function is calculated. When more than one molecule is present in the asymmetric unit, the interpretation of the self-rotation function can provide us with the information about the nature of the non-crystallographic (NCS) symmetry elements relating these molecules;
- Using equation 15 the \( c_{lmm} \) coefficients are computed;
- The \( A_{mm'} \) coefficients are computed for each value of \( \theta_2 \) using equation 14. The matrix elements \( d_{lmm} \) are calculated at this step by recurrence relations;
- Finally the rotation function is calculated by equation 13. As we can see the calculation process uses only two variables \( \theta_1, \theta_3 \), and the calculation is re-initialised for each \( \theta_2 \).

Relative to the Rossmann-Blow rotation function, Crowther’s fast rotation function has the following advantages:

- It is about 100 times faster in terms of CPU usage;
- The calculation is more precise, since the truncation of \( G \) function by using only large \( |F|^2 \) [22] need not be made;
- It is easier to take the symmetry of the Patterson function into consideration. For example, if we have a symmetry axis of the \( N \)-th order along the \( z \) axis, then the non-zero coefficients among the coefficients \( a_{lmmn} \) will be only \( a_{lmmn} \) \( k = 0, 1, 2, \ldots \); 
- Finally, we can remove the origin peak in a very simple way by omitting the \( a_{000} \) coefficient from the calculation.

### 1.4 The translation function

Once the orientation of a probe molecule is known, the actual position of the molecule must be found. The main concept of the translation function is the same as for the rotation function. Here, we look at the correlation between the observed intensities and the Patterson cross-vectors of the symmetry-related molecules of the probe as it is moved within the cell. When the probe is correctly positioned, the translation function should have peaks at values corresponding to the translation vectors between the symmetry related molecules. There are several forms of the translation function, a brief account of which will be given in the following sections. A more detailed description of various translation functions is given in [26].
### 1.4.1 The T and T1 Translation Functions

Translation functions build models of the target crystal by moving properly oriented probes about the cell. For each position of the test molecule described by the vector, say \( x \), one can calculate the intensities (structure factor amplitudes) for the resultant model \( |F_{\text{calc}}(\vec{h})|^2 \) and see how they agree with observed intensities \( |F_{\text{obs}}(\vec{h})|^2 \). Crowther and Blow have initially represented this agreement by the T translation function [23]:

\[
T(\vec{x}) = \sum_{\vec{h}} |F_{\text{obs}}(\vec{h})|^2 |F_{\text{calc}}(\vec{h}, \vec{x})|^2 \tag{16}
\]

The calculated intensity can be expressed in terms of the structure factors of the individual symmetry-related molecules.

Let us consider a crystal with two molecules in the unit cell that are related by a two-fold axis. The calculated intensity is given by:

\[
|F_{\text{calc}}(\vec{h})|^2 = |F_{\text{calc}}(\vec{h})|^2 F^*_1(\vec{h}) = (F_{M1} + F_{M2})(F^*_1 + F^*_2) \tag{17}
\]

where \( F_{M1} \) and \( F_{M2} \) are the structure factors of the two molecules at the point \( \vec{h} \) in reciprocal space, and \( F^* \) values are their complex conjugated counterparts. For a molecule, which has been shifted by a vector \( \vec{x} \) from the original position, only the phase of its structure factor changes, and \( F_{M1} \) becomes \( F_{M1} e^{2\pi i \vec{x} \cdot \vec{h}} \). The calculated intensities can now be expressed as follows:

\[
|F_{\text{calc}}(\vec{h})|^2 = (F_{M1})^2 + F_{M2}^* F_{M1} e^{2\pi i \vec{x} \cdot \vec{h}} + F_{M1} F_{M2} e^{-2\pi i \vec{x} \cdot \vec{h}} \tag{18}
\]

Here \( x_1 \) and \( x_2 \), the positions of the molecular centres, are related by crystal symmetry. By substituting equation 18 into equation 16 one obtains:

\[
T(\vec{X}) = \sum_{\vec{h}} |F_{\text{obs}}(\vec{h})|^2 \left( |F_{M1}|^2 + |F_{M2}|^2 + F_{M1} F_{M2} e^{2\pi i \vec{X} \cdot \vec{h}} + F_{M1} F_{M2} e^{-2\pi i \vec{X} \cdot \vec{h}} \right) \tag{19}
\]

where \( \vec{X} \) equals \( x_1 - x_2 \). We are only interested in terms determining the cross-vectors, i.e., those which are dependent on the relative positions of the molecules. We can, therefore, omit the first two \( \langle F \rangle \) terms in the brackets, as they represent the self-vectors for the two molecules. The remaining term is known as the T translation function:

\[
T(\vec{X}) = \sum_{\vec{h}} |F_{\text{obs}}(\vec{h})|^2 F_{M1} F_{M2} e^{-2\pi i \vec{X} \cdot \vec{h}} \tag{20}
\]

Besides being independent from the molecular positions, the self-vectors would contribute to the background noise. In an attempt to improve the signal-to-noise ratio we may try to subtract the coefficients for the self-vectors from the observed intensities:

\[
|F_{\text{calc}}(\vec{h})|^2_{\text{cross}} = |F_{\text{calc}}(\vec{h})|^2 - k(F_{M1}^2 + F_{M2}^2) \tag{21}
\]

Here, \( k \) is a coefficient which scales the observed and the calculated intensities together. The self-vectors from the search probes have been subtracted from the observed Patterson function to leave, approximately, an observed cross-vector set. The T translation function calculated by means of equation 21 is normally referred to as the T2 translation function [8, 26].

### 1.4.2 The T2 Translation Function

The T2 translation function is the most commonly used in molecular replacement applications. While the T1 translation function has to be calculated for every pair of symmetry related molecules, the T2 function takes the whole symmetry into account. It has the form:

\[
T2(\vec{X}) = \sum_{\vec{h}} |F_{\text{obs}}(\vec{h})|^2_{\text{cross}} \sum_i \sum_j F_{Mi} F_{Mj}^* e^{-2\pi i \vec{X} \cdot \vec{h}} \tag{22}
\]

The T2 translation function applies phase shifts so that all peaks are piled up at the same place, defined by a single translation vector \( \vec{X} \) for the reference molecule, so the signal-to-noise ratio is much greater than for a single T1 function.
1.4.3 The TO/O Translation Functions

A number of improvements have been made to the original translation function. One of these consists of the combination of a three-dimensional translation function with a packing analysis [8]. The resulting function has the following form:

$$T_{H}(X) = \frac{TO(X)}{O(X)}$$  \hspace{1cm} (23)

where $TO(X)$ is the same as the T translation function (equation 20) but normalised:

$$TO(X) = \frac{\sum_{h}[F_{\text{obs}}(\vec{h})]^{2}[F_{\text{calc}}(\vec{h}, \vec{X})]^{2}}{\sum_{h}|F_{\text{obs}}(\vec{h})|^{2}}$$  \hspace{1cm} (24)

$O(X)$ is a packing function that measures the interpenetration of the molecules:

$$O(X) = \frac{\sum_{h}[F_{\text{calc}}(\vec{h}, \vec{X})]^{2}}{N \sum_{h} F_{M}}$$  \hspace{1cm} (25)

1.5 Factors affecting the molecular replacement solution

Solving a structure by the molecular replacement method is not always a straightforward task. Depending on the complexity of the problem, a number of important factors must be taken into consideration in order to ensure a correct structure determination. In this section we consider the factors affecting the molecular replacement solution.

The crystal. The crystal form most favourable for the molecular replacement method is the one with only a single molecule in the asymmetric unit. The presence of multiple copies of the protein in the asymmetric unit reduces the signal-to-noise ratio for the correct peaks on the rotation and translation function. Furthermore, the presence of non-crystallographic symmetry between these copies can introduce difficulties in determining correct solutions. However, there are many cases in which the structure was solved by molecular replacement method where there are several molecules in the asymmetric unit.

The X-ray data. It is very important that the experimental data are as complete as possible, ideally 100%. Also, the data should be of high quality. Possible problems include: systematically missing regions, ice ring, detector overloads, etc.

Molecular replacement search probe. The choice of a search probe has to be made by considering possible structural similarities with the target molecule (within 1.0 – 1.5 Å r.m.s.d.). The two structures should have, at least, around 30 - 40% amino-acid sequence identity and, in general, the higher the better. Also, when the choice for the search probe is given between a crystallographic structure and an NMR structure, the former, normally, provides a higher degree of success. This has mainly to do the imprecision of the NMR protein model. Furthermore, the search model does not necessarily have to be used in its integrity; some parts of it can be removed completely (at the expense of the amount of scattering matter that we are placing in the unit cell) or the side chains of some residues can be trimmed to alanine/glycine residues in the regions where the largest differences with the unknown structure are expected.

Rotation function integration radius. As the rotation function calculation has to be based only on the intra-molecular vectors and not on the inter-molecular ones, the choice of the integration radius is critical. Two main criteria can be used to decide what integration radius range should be explored:

- in the case of a spherical molecule, a value corresponding to about 75% of the minimum diameter should be used;
- in the case of elongated molecule, the average of the three semi-axes of the corresponding ellipsoid would be the value of choice. In both cases, deviations from these starting values should be examined in order to increase the signal-to-noise ratio corresponding to the correct solution.
Temperature factors sharpening. The Patterson peaks are much broader than the electron density peaks from which it is constructed (figure 2). Therefore, the Patterson space will have overlapping peaks affecting its clarity. Reducing (i.e. sharpening) of the atomic temperature factors of the molecular probe can be applied to avoid this problem. However, this parameter is highly correlated with the adopted resolution cut-off and, therefore, care should be taken.

The use of normalised structure factors $E(\vec{h})$. The normalised structure factors are given by the following formula,

$$E(\vec{h}) = \frac{F(\vec{h})}{\sqrt{\langle |F(\vec{h})|^2 \rangle}}$$

where $\sqrt{\langle |F(\vec{h})|^2 \rangle}$ is the root mean square value of structure factor amplitudes with sin $\theta$ values close to that of $F(\vec{h})$. Consequently the average value of $E^2$ is 1. The physical meaning of the normalised structure factors is that they are the structure factors corresponding to a structure of stationary point atoms. The use of normalised structure factors in the molecular replacement calculations can make the correct solutions more obvious and is another way of sharpening the Patterson peaks.

Resolution limits and other parameters. The choice of the resolution limits is a trial-and-error problem. Reflections with a resolution worse than 15 – 25 Å have to be omitted from the calculation, since they contain more information about solvent vectors than about protein vectors. The cut-off at high resolution depends on the degree to which the search probe molecule is similar to the target molecule; the greater the similarity the higher the resolution that can be used. Including high resolution terms, in turn, increases computational time, so a resolution cut-off of around 4 – 3 Å is usually used. The Patterson function has a large peak at the origin. Removal of this peak can give a greater contrast to the correct peak. The sampling of the Patterson or reciprocal space is also important, since the correct Patterson function peak can be missed by improper sampling. A frequently used criterion for sampling is 1/3 of the maximum resolution used in the calculation.
Figure 6: The sequence alignment of kringle domains. The sequence alignment of possible molecular replacement kringle search probe candidates with the target kringle of the NK1 fragment. The colouring of highlighted residues corresponds to the various degrees of sequence conservation. The Cys residues involved in disulphide bridge formations are indicated by arrows. The sequence consensus is shown underneath.

<table>
<thead>
<tr>
<th>Candidate</th>
<th>PDB code</th>
<th>Resolution, Å</th>
<th>R-value (1)</th>
<th>Sequence identities</th>
<th>Sequence positives (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasminogen Kringle 1</td>
<td>1PKR</td>
<td>2.5</td>
<td>15.9%</td>
<td>51%</td>
<td>67%</td>
</tr>
<tr>
<td>Plasminogen Kringle 4</td>
<td>1PK4</td>
<td>1.9</td>
<td>14.9%</td>
<td>51%</td>
<td>56%</td>
</tr>
<tr>
<td>Prothrombin Fragment 1</td>
<td>2PF1</td>
<td>2.25</td>
<td>17.5%</td>
<td>48%</td>
<td>60%</td>
</tr>
</tbody>
</table>

(1) R-value = \[ \sum |F_o - F_c| / \sum F_o \], where \( F_o \) is the observed structure factor amplitude and \( F_c \) is the calculated structure factor amplitudes;

(2) 'Sequence positives' value is calculated taking into an account the properties (like charge or hydrophobicity) of the residues in sequences aligned.

2 The application of the molecular replacement method to the structure solution of the NK1 fragment.

2.1 Molecular replacement search probes

2.1.1 The kringle domain probe

The kringle domains have been under intensive investigation for the past decade. At present, the protein structural databank contains 13 entries on the kringle domains. These are the structures of prothrombin fragment 1, t-PA activator kringle 2, plasminogen kringle 1 and plasminogen kringle 4 domains and their complexes with different chemical compounds. Sequence alignments as well as spectroscopic studies suggest that kringles are structurally homologous. After eliminating the NMR structures and the structure of complexes, three possible candidates for a kringle domain search probe are left to consider. These are human plasminogen kringle 1 (PDB code: 1PKR; [29]); human plasminogen kringle 4 (PDB code: 1PK4; [17]); and bovine prothrombin fragment 1 (PDB code: 2PF1; [20]). The sequence alignment of these structures with the kringle domain of the NK1 fragment is shown in figure 6, and the statistics of each candidate are presented in table 1.

As can be seen from the alignment, all candidates share high sequence identities with the NK1 fragment.
The structural homology was also confirmed by making a three-dimensional superposition of given structures using the COMPARER program [25]. The results of this superposition are shown in the figure 7. The analysis performed by COMPARER showed that the root mean square deviation (r.m.s.d.) of the Cα-atom positions in all three structures is within 0.5 Å. Therefore, taking into account the sequence identities with the kringle domain, one can say that each of these structures can be used as a molecular replacement probe. However, of the three kringle structures, the plasminogen kringle 1 has a slightly higher percentage of positive sequence matches. For this reason, plasminogen kringle 1 was chosen as the search probe.

The main feature of all kringle structures is the lack of secondary structural elements, apart from two small antiparallel β-sheet-like structures. The ribbon representation picture of plasminogen kringle 1 domain with disulphide bridges is shown in the figure 8. However, as the above analysis shows, kringles are globular and compact domains, with a very high degree of three-dimensional similarity. The fold of any kringle structure is stabilised by at least three disulphide bridges. One of them connects the amino-terminus with the carboxy-terminus, thus restricting the mobility of the structure termini.

A more careful analysis of the three-dimensional superposition showed that there are several small regions in plasminogen kringle 1 that deviate significantly (1 – 2 Å) from the other two structures. These regions are residues 32 – 34, 39 – 45, 55 – 59. Later, one can consider removing them, if the MR problem appears to be difficult to solve. Additionally, there is an option of ’trimming’ (replacing by Ala or Gly residues) the side chains that differ from the target protein. However, removing large parts of the search model has the affect of reducing the amount of scattering matter placed in the unit cell. As previously described, this is not advisable in MR cases when several molecules are present in the asymmetric unit. A third possibility is to replace by Ala any long residues of the search probe which have much shorter counterparts in the target protein. The list of such residues is shown in table 2. These replacements were introduced using the O molecular graphics package [11].

2.1.2 Molecular replacement search probes: the N-domain probe

The project was initiated when no structural information concerning the N-domain of HGF/SF protein was available. Therefore, it was initially decided to solve the phase problem by using a combination of the MR and MIR techniques. Unfortunately, all attempts at obtaining heavy atom derivatives failed, largely...
Figure 8: The ribbon representation of human plasminogen kringle 1. The wide-eye figure of human plasminogen kringle 1. Secondary structure elements, as defined in the PDB file (PDB code: 1PKR; [29]), are shown. The disulphide bridges are shown in yellow. The Cys residues involved in the disulphide bridges formation are labelled. The amino- and carboxy- termini are also labelled.

Table 2: The list of residues in the kringle search probe that were replaced by Ala or Gly residues.

<table>
<thead>
<tr>
<th>Residue Number (as in 1PKR)</th>
<th>Residue Type (as in 1PKR / NK1)</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>LYS / ILE</td>
<td>Replace by ALA</td>
</tr>
<tr>
<td>13</td>
<td>MET / VAL</td>
<td>Replace by ALA</td>
</tr>
<tr>
<td>15</td>
<td>LYS / ILE</td>
<td>Replace by ALA</td>
</tr>
<tr>
<td>24</td>
<td>LYS / PRO</td>
<td>Replace by ALA</td>
</tr>
<tr>
<td>32</td>
<td>ARG / GLU</td>
<td>Replace by ALA</td>
</tr>
<tr>
<td>34</td>
<td>ARG / SER</td>
<td>Replace by ALA</td>
</tr>
<tr>
<td>43</td>
<td>SER / GLY</td>
<td>Replace by GLY</td>
</tr>
<tr>
<td>56</td>
<td>ASN / GLY</td>
<td>Replace by GLY</td>
</tr>
<tr>
<td>59</td>
<td>GLN / GLY</td>
<td>Replace by GLY</td>
</tr>
<tr>
<td>70</td>
<td>LYS / VAL</td>
<td>Replace by ALA</td>
</tr>
<tr>
<td>74</td>
<td>TYR / VAL</td>
<td>Replace by ALA</td>
</tr>
<tr>
<td>78</td>
<td>LEU / PRO</td>
<td>Replace by ALA</td>
</tr>
</tbody>
</table>
owing to the non-isomorphism of the derivatives. Later, the NMR structure solution of human HGF/SF N-domain was reported [30]. This NMR structure (coordinates of a set of final 23 NMR models, supplied by Dr R.A. Byrd, NCI-Fredrick, Cancer Research and Development Centre) was used as the search probe for the molecular replacement.

Figure 9a shows the superposition of the 23 NMR models of the HGF/SF N-domain. The overall structure is well defined except for the Asn77 – Gly79 region and the seven residues of the amino-terminus, for which the NMR data were not observed. The middle of the structure is a five-stranded antiparallel β-sheet (figure 9b). This sheet is flanked on one side by a two-turn α-helix, a short 310-helix, and the amino-terminal residues. On the other side it is flanked by two extended loops joining the opposite ends of a β-strand. This organisation results in a three-layered sandwich topology.

The NMR structure of the N-domain shows a very well-defined core region. The plot of the positional r.m.s.d of all the atoms of the restrained minimised averaged structure is shown in figure 10. The average r.m.s.d. for the backbone atoms of the well defined region is 0.37 Å.

Several potential molecular replacement search probes of the N-domain were prepared. These included: a minimised averaged structure with atoms having a r.m.s.d. of more than 2.0 Å removed; main chain atoms only; and a minimised averaged structure with the loops removed. The 23 structures were analysed and clustered into four different groups using the program NMRCLUST [12]. The best representative of each cluster was identified and also used as an MR probe. In all probes the residues in regions 31 – 38, 77 – 79, and 127 were removed. As the NMR technique cannot determine individual temperature factors (B-factors), two different schemes for assigning their values were used:

- all B-factors set to 20 or 30 Å²;
- individual B-factors set proportionally to the r.m.s.d of each atom:

\[ B - \text{factor}_j = 20 + k \times (\text{r.m.s.d}_j) \]

where \( j \) is the atom identifier, and \( k \) is a scale factor, chosen to be between 20 and 10.

2.2 The use of the 'Automated Molecular Replacement' (AMoRe) package in the solution of the NK1 fragment structure

In the crystals of NK1 there are two molecules in the asymmetric unit, each with two different domains. Thus, we have to find solutions for four rotational and four translational problems. There is a number of crystallographic software packages that can perform the molecular replacement procedure. The most popular include: MERLOT [7], XPLOR [2], AMoRe [18] and CNS [3]. The chosen software for solving these problems was the Automated Molecular Replacement package (AMoRe) [18]. The flow-chart of the AMoRe procedure is shown in figure 11. This package can use both structure factor amplitudes and normalised structure factor amplitudes in its calculations. The following sections describe the various steps and operations of the AMoRe package that were applied to the NK1 fragment MR problem using both types of structure factor amplitudes in the calculation.

2.2.1 The preparation steps

Two preliminary programs must be run in order to format the data into a suitable representation. The first is SORTING, which transforms the measured X-ray data into the AMoRe internal format. The second program is TABLING, which calculates the continuous Fourier coefficients, corresponding to the kringle and N-domain search probes. The continuous Fourier coefficients will be used in the structure factor calculation by interpolation [18]. These coefficients are obtained by Fourier transforming the electron density based on the atomic coordinates of each search probe, which is placed within a cell with linear dimensions about four times the size of the molecule. To avoid creating an unnecessary large cell, caused by an unfavourable positioning of the probe, the molecule is shifted and rotated so that its centre of mass is at the origin and its principal axes of inertia are aligned with the cell axis. The resulting position is taken as the initial reference, and modified atomic coordinates of each probe are written out. One should keep in mind that all non-protein entities such as water molecules are removed from the input coordinates.
Figure 9: The NMR structure of the HGF/SF N-domain as a search probe.  

a) The superposition of a set of 23 NMR structures [30]. The amino-terminal is clearly disordered as well as residues Asn77 – Gly79 (shown by arrow).  
b) A ribbon representation of a restrained averaged NMR structure of N-domain showing secondary structure elements. The disulphide bridges are shown in yellow. The regions not used in molecular replacement calculations are shown in green.
Additional parameters can be tuned at this stage of the molecular replacement procedure. In the SORTING program a resolution limits cut-off can be applied to the experimentally measured data. In the case of the NK1 fragment crystals X-ray diffraction data, these were from 50 to 2.6 Å. In the TABLING program several parameters can be specified. They include:

- **Scale** regulates the size of the probe’s cell, which is given by the size of the search molecule multiplied by the scale value. This controls how finely the model structure factors sample the reciprocal space. This value had been set to 4.0 for both the kringle and the N-domain search probes.

- **Badd** adds the specified value to all input temperature factors. If badd is given a negative value the temperature factors are ‘sharpened’ (reduced). In the first runs of molecular replacement with the kringle and the N-domain probes Badd was set to zero, and it was changed to $-10$ or $-20$ Å² in the following runs.

- **Shann** specifies the rate of sampling of the coordinate space. Sampling uses a grid of approximately \( \text{resolution}/(2 \times \text{shann} \text{ value}) \). The coordinate map should not be sampled at a grid less than 0.75 Å, and if the B-factors have been sharpened, it is advisable to use a finer grid. This factor was chosen to be equal to 2.5 for all probes.

### 2.2.2 The rotation function search

The next step is the calculation of the rotation function using the program ROTING. First, the structure factor amplitudes of a search model have to be computed in a P1 cell (the size of which is controlled by the **Scale** parameter). This cell is substantially smaller than that used to compute the continuous Fourier coefficients. The size of this cell can be calculated in two ways. The first is to define a cubic cell with the edges of double the maximum distance from the centre of mass, plus the integration radius, plus a small safety term. The second is to take the size of this cell equal to the smallest box containing the search
Figure 11: The AMoRe molecular replacement procedure flow-chart. The step-by-step guide of AMoRe practice. The individual programs used at each step are indicated in bold letters. For the explanation of each step see the following sections.
probe, plus the integration radius, plus the resolution used for calculation. In any case the cell should be big enough to accommodate the molecule.

Next, the spherical harmonic coefficients for the target and probe Patterson distributions are computed. Finally, the rotation function is calculated. The optional control of the rotation function calculation includes:

- sharpening of the temperature factors;
- low and high order limits of the spherical harmonic function terms;
- limit of the Eulerian $\beta$-angle.

The sharpening of the temperature factors can increase the signal-to-noise ratio of the correct solution. The low-order terms of the spherical harmonics depend on the crystal symmetry. Removing them may reduce the peak heights, but make the solution more precise and make multiple solutions have more equal heights. The upper cut-off is governed by the ratio between the integration radius and the resolution. The lower cut-off has a similar effect to the inner cut-off radius for the Patterson vectors. To eliminate the symmetry-related peaks, the limit on the $\beta$-angle can be adjusted. In the case of the NK1 fragment crystals, the $\beta$-angle limit can be set to 180°, as the crystal lattice is monoclinic with a two-fold screw axis along the $b$ direction (P21). In this way we exclude the solutions that are related by this axis.

The result of the rotation function search, relative to the chosen molecular model, consists of a list of possible solutions sorted by the peak heights. Each output line contains the three Eulerian angles which identify the different orientations of the probe molecule in the cell and the peak height.

### 2.2.3 The translation function search

The top peak in the list of the rotation function solutions, produced by AMoRe, is not necessarily the correct one. If the rotation function solution is correct, it has to give the highest peak in the translation function as well. Since we know the orientation and the position of the search probe within the crystal unit cell, we can compute the structure factors from this model. Thus, we can judge the correctness of the solution by looking at the correlation coefficient and R-factor computed between calculated and observed structure factor amplitudes. In the case of a correct solution, the correlation coefficient should be as high as possible and the R-factor as low as possible.

A translation function can be computed, within the AMoRe package, by means of the TRAING program. It uses, as input, the solutions of the rotation function. Optional control parameters include: Badd (as defined as in the ROTING program) and the resolution limits adopted. Several rotation solutions for one search probe can be tested at once. For each rotation function solution, the search probe, identified by its label, is rotated by the specified angles and translated within the translation function cell.

The program TRAING also computes the correlation coefficient (equation 27) and the standard crystallographic R-factor (equation 28) for each translation solution. These two indicators are usually used to assess the agreement between observed and calculated structure factor amplitudes.

\[
\text{corr. coef.} = \frac{\sum(F_o - <F_c>) |(F_c - <F_c>)|}{\sqrt{\sum(F_o - <F_c>)^2 \sum(F_c - <F_c>)^2}}
\]  

(27)

\[
R\text{-factor} = \frac{\sum|F_o - F_c|}{\sum F_o}
\]

(28)

where $F_o$ are the observed and $F_c$ are the calculated structure factor amplitudes.

When two molecules are present in the asymmetric unit, once the translation for one molecule is found, it can be kept fixed while searching for the position of the second molecule. The fixed solution will provide a constant contribution to this second translation function calculation. This is very helpful because when several molecules are present in the asymmetric unit, one of the solutions will dominate in the rotation function solutions list. The second molecule often has a low peak height and may be hidden down below on the list of rotation function solutions. In this case, after determining the position of the first molecule, it is possible to find the second solution by fixing the first and scanning all the other rotation function solutions.
Table 3: The size of the kringle (KRG) and the N-domain (PAP) molecules.

<table>
<thead>
<tr>
<th>Probe</th>
<th>Minimal Box, Å</th>
<th>Maximal distance from centre of mass, Å</th>
<th>75% of minimal diameter, Å</th>
</tr>
</thead>
<tbody>
<tr>
<td>PAP</td>
<td>25.9 32.5 32.9</td>
<td>19.9</td>
<td>19.4</td>
</tr>
<tr>
<td>KRG</td>
<td>27.0 26.2 30.9</td>
<td>20.5</td>
<td>19.6</td>
</tr>
</tbody>
</table>

Table 4: The list of the first 10 peaks rotation function performed with the kringle probe (the two correct solutions are highlighted).

<table>
<thead>
<tr>
<th>Peak Number</th>
<th>Probe</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>Peak height</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kringle</td>
<td>30.57</td>
<td>33.29</td>
<td>4.21</td>
<td>16.0</td>
</tr>
<tr>
<td>2</td>
<td>Kringle</td>
<td>321.50</td>
<td>34.11</td>
<td>44.06</td>
<td>13.4</td>
</tr>
<tr>
<td>3</td>
<td>Kringle</td>
<td>279.79</td>
<td>79.94</td>
<td>278.30</td>
<td>11.8</td>
</tr>
<tr>
<td>4</td>
<td>Kringle</td>
<td>238.72</td>
<td>23.18</td>
<td>174.24</td>
<td>11.4</td>
</tr>
<tr>
<td>5</td>
<td>Kringle</td>
<td>95.27</td>
<td>75.77</td>
<td>231.58</td>
<td>11.4</td>
</tr>
<tr>
<td>6</td>
<td>Kringle</td>
<td>3.66</td>
<td>67.47</td>
<td>9.63</td>
<td>11.0</td>
</tr>
<tr>
<td>7</td>
<td>Kringle</td>
<td>47.94</td>
<td>78.50</td>
<td>37.51</td>
<td>10.8</td>
</tr>
<tr>
<td>8</td>
<td>Kringle</td>
<td>76.50</td>
<td>22.77</td>
<td>144.01</td>
<td>10.6</td>
</tr>
<tr>
<td>9</td>
<td>Kringle</td>
<td>300.48</td>
<td>90.00</td>
<td>94.79</td>
<td>10.5</td>
</tr>
<tr>
<td>10</td>
<td>Kringle</td>
<td>240.21</td>
<td>90.00</td>
<td>274.38</td>
<td>10.5</td>
</tr>
</tbody>
</table>

Resolution limits: 8 - 3.5 Å; probe’s cell: 55, 55, 55 Å; integration radius: 20 Å.

solutions together with the fixed one. The correct solution in this case will have a correlation coefficient greater than the first one. The output of the translation function search contains the peaks sorted by the correlation coefficient.

2.2.4 The integration radius of kringle and N-domain search probes

The choice of the integration radius in the calculation of rotation functions depends on the size of the search probe. As previously mentioned, a value corresponding to 75% of the minimal diameter of the molecule normally represents a good approximation. The TABLING program outputs the size of the minimal box containing the search probe and the distance of the furthest atom from the centre of mass. These values, as well as the length corresponding to 75% of the minimal diameter for the kringle (KRG) and the N-domain (PAP) search probes, are listed in table 3.

2.2.5 Results of the kringle probe rotation search

The rotation function search for the kringle domain clearly shows two peaks. The parameters for the calculation were: integration radius of 20 Å, resolution limits from 8 to 3.5 Å, a cubic search probe cell with axis equal to 55 Å. No structure factor sharpening was used. In table 4 the first 10 peaks from the rotation function calculation are listed. tables 5 and 6 show the effect of different resolution limits and radius of integration on the signal (peak height minus the highest background peak) divided over sigma (noise) of these two peaks (i.e. correct solutions). Negative values for the signal-to-noise ratio mean that the highest background peak was in the first position of the rotation function solutions list.

Different combinations of resolution limits, integration radii and B-factor values of the structure factors did not improve the signal-to-noise ratio. Furthermore, the peak corresponding to the first solution was always observed on top of the list, while the second kept changing its position relative to the set of incorrect peaks. Other kringle domains were tested as the kringle search probe, and the peak of the first solution was invariably observed on top of all the others. These observations suggest that the solution corresponding to the first peak of the kringle probe rotation function is correct; it describes the orientation of the first kringle domain in the crystal unit cell. Although no conclusive evidence in favour of the correctness of the second peak can be obtained, it seems to be a very good candidate for the actual orientation of the second kringle
Table 5: The effect of the resolution limits on the signal of the correct solutions for the kringle domain search probe (the best combination of resolution limits is highlighted).

<table>
<thead>
<tr>
<th>Resolution Limits, Å</th>
<th>Solution #1</th>
<th>Solution #2</th>
<th>The highest background peak</th>
<th>Sigma</th>
<th>Signal/noise Solution #1, σ</th>
<th>Signal/noise Solution #2, σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 - 4.0</td>
<td>17.4</td>
<td>16.0</td>
<td>15.0</td>
<td>3.5</td>
<td>0.68</td>
<td>0.28</td>
</tr>
<tr>
<td>8 - 4.0</td>
<td>17.2</td>
<td>16.2</td>
<td>14.4</td>
<td>3.4</td>
<td>0.82</td>
<td>0.53</td>
</tr>
<tr>
<td>20 - 3.5</td>
<td>16.8</td>
<td>12.9</td>
<td>13.2</td>
<td>3.1</td>
<td>1.16</td>
<td>-0.09</td>
</tr>
<tr>
<td>8 - 3.5</td>
<td>16.0</td>
<td>13.4</td>
<td>11.8</td>
<td>2.9</td>
<td>1.45</td>
<td>0.56</td>
</tr>
<tr>
<td>20 - 3.0</td>
<td>16.8</td>
<td>11.8</td>
<td>12.6</td>
<td>2.8</td>
<td>1.50</td>
<td>-0.28</td>
</tr>
<tr>
<td>8 - 3.0</td>
<td>16.1</td>
<td>11.8</td>
<td>10.8</td>
<td>2.7</td>
<td>1.96</td>
<td>0.37</td>
</tr>
</tbody>
</table>

The integration radius of 20 Å and the probe’s cell of 55, 55, 55 Å were used for calculations.

Table 6: The effect of the integration radius changes on the signal of the correct solutions for the kringle domain probe (the best radius is highlighted).

<table>
<thead>
<tr>
<th>Integration Radius, Å</th>
<th>Solution #1</th>
<th>Solution #2</th>
<th>The highest background peak</th>
<th>Sigma</th>
<th>Signal/noise Solution #1, σ</th>
<th>Signal/noise Solution #2, σ</th>
</tr>
</thead>
<tbody>
<tr>
<td>18</td>
<td>16.9</td>
<td>13.4</td>
<td>12.5</td>
<td>3.1</td>
<td>1.41</td>
<td>0.29</td>
</tr>
<tr>
<td>20</td>
<td>16.0</td>
<td>13.4</td>
<td>11.8</td>
<td>2.9</td>
<td>1.45</td>
<td>0.56</td>
</tr>
<tr>
<td>22</td>
<td>15.3</td>
<td>12.0</td>
<td>11.5</td>
<td>2.7</td>
<td>1.40</td>
<td>0.18</td>
</tr>
</tbody>
</table>

Resolution limits: 8 - 3.5 Å; probe’s cell: 55, 55, 55 Å.

2.2.6 Results of the N-domain probe rotation search

The results of the rotation function search using the N-domain search probe were more difficult to interpret than those of the kringle probe rotation search. The output list did not contain any strong (more than 0.2σ signal-to-noise ratio) peaks. The noisiness of the rotation function was somehow expected, as a relatively imprecise NMR model was used here as a search probe. The rotation function search was performed using different N-domain search probes, as they were described in section 2.1, but still no clear solutions could be found. However, as will be shown in the sections below, the peaks corresponding to the correct solutions were determined to be on the fifth and seventh positions in the list produced by the rotation function search using, as the search probe, the minimised averaged structure of the N-domain lacking residues 31 – 38, 77 – 79, 127. Table 7 shows the first 10 rotation peaks; the peaks corresponding to the correct solutions are highlighted.

The parameters for the calculation were: integration radius 20 Å; resolution limits from 8 to 3.0 Å; a cubic search probe cell with axes of 60 Å. The temperature factors of the search structure were set to 20 Å². Different combinations of resolution limits, integration radii, and temperature factor distributions were also tested, but peaks 5 and 7 never reached a higher position in the list. Table 8 shows the effect of different resolution limits and radius of integration on the position of the correct solution.

2.2.7 The kringle and the N-domain probes translation function search

Once a list of rotation function peaks has been obtained, the highest peak is selected and the corresponding orientation of the search model retained in the translation function calculation. The translation solution which shows the highest correlation coefficient between observed and calculated structure factor amplitudes will then have to be picked and kept fixed while performing a translation search for the other rotation function solutions. The peak corresponding to the correct solution must give a higher correlation coefficient and a lower R-factor than the first one. Then these two solutions must be fixed and the
Table 7: The list of the first 10 peaks rotation function performed with the N-domain probe (the two correct solutions are highlighted).

<table>
<thead>
<tr>
<th>Peak Number</th>
<th>Probe</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>Peak height</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PAP</td>
<td>277.48</td>
<td>76.07</td>
<td>233.34</td>
<td>18.9</td>
</tr>
<tr>
<td>2</td>
<td>PAP</td>
<td>308.93</td>
<td>57.82</td>
<td>226.06</td>
<td>17.0</td>
</tr>
<tr>
<td>3</td>
<td>PAP</td>
<td>233.88</td>
<td>42.49</td>
<td>295.57</td>
<td>16.1</td>
</tr>
<tr>
<td>4</td>
<td>PAP</td>
<td>298.60</td>
<td>54.72</td>
<td>356.36</td>
<td>15.3</td>
</tr>
<tr>
<td>5</td>
<td>PAP</td>
<td>237.22</td>
<td>81.68</td>
<td>3.30</td>
<td>15.2</td>
</tr>
<tr>
<td>6</td>
<td>PAP</td>
<td>145.59</td>
<td>48.10</td>
<td>198.50</td>
<td>14.3</td>
</tr>
<tr>
<td>7</td>
<td>PAP</td>
<td>257.33</td>
<td>64.92</td>
<td>8.58</td>
<td>14.3</td>
</tr>
<tr>
<td>8</td>
<td>PAP</td>
<td>1.00</td>
<td>32.73</td>
<td>334.00</td>
<td>14.1</td>
</tr>
<tr>
<td>9</td>
<td>PAP</td>
<td>31.50</td>
<td>58.76</td>
<td>253.84</td>
<td>14.0</td>
</tr>
<tr>
<td>10</td>
<td>PAP</td>
<td>314.50</td>
<td>47.59</td>
<td>142.40</td>
<td>14.0</td>
</tr>
</tbody>
</table>

Resolution limits: 8 - 3.0 Å; probe’s cell: 60, 60, 60 Å; integration radius: 20 Å.

Table 8: The effect of the resolution limits and integration radii changes on the signal for the correct solutions (i.e. place in the list of peaks) for the N-domain probe (the best combination of resolution limits is highlighted).

<table>
<thead>
<tr>
<th>Resolution Limits, Å</th>
<th>Radius 24, A</th>
<th>Radius 22, A</th>
<th>Radius 20, A</th>
<th>Radius 18, A</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 - 4.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8 - 4.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>20 - 3.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8 - 3.5</td>
<td>-</td>
<td>22</td>
<td>29</td>
<td>14</td>
</tr>
<tr>
<td>20 - 3.0</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8 - 3.0</td>
<td>-</td>
<td>10</td>
<td>-</td>
<td>5</td>
</tr>
</tbody>
</table>

Probe’s cell of 60, 60, 60 Å was used for all calculations. * - * sign indicates that the correct solution was not present in the top 30 peaks.
Table 9: The peak for the correct translation function solution for the first orientation of the kringle probe.

<table>
<thead>
<tr>
<th>Rotation peak Number</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>Tx</th>
<th>Ty</th>
<th>Tz</th>
<th>Corr. Coef</th>
<th>R-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>30.57</td>
<td>33.29</td>
<td>4.21</td>
<td>0.0263</td>
<td>0.0000</td>
<td>0.0625</td>
<td>34.2</td>
<td>53.9</td>
</tr>
</tbody>
</table>

Resolution limits of 20 - 3.0 Å were used for the calculation.

Table 10: The list of the translation function peaks for the 7th rotation function peak of the N-domain probe (kringle solution #1 is fixed). The translation peak of the correct solution is highlighted.

<table>
<thead>
<tr>
<th>Rotation peak Number</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>Tx</th>
<th>Ty</th>
<th>Tz</th>
<th>Corr. Coef</th>
<th>R-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>257.33</td>
<td>64.92</td>
<td>8.58</td>
<td>0.5515</td>
<td>0.0352</td>
<td>0.9106</td>
<td>34.9</td>
<td>52.9</td>
</tr>
</tbody>
</table>

Resolution limits of 20 - 3.0 Å were used for the calculation.

The translation function search performed on the remaining rotation function peaks. The next correct solution (peak) must have a better correlation coefficient (i.e. higher) and a better R-factor (i.e. lower) than the previous one. This procedure is repeated with the rest of the rotation function solutions until the positions of all molecules are found.

The translation function performed on the best peak of the rotation function of the kringle search probe was calculated using a 20 – 3.0 Å resolution range. The height of the produced peaks was limited to half the height of the maximum peak. The output list contained only one peak, and it had a correlation coefficient of 34.2 % and a R-factor of 53.9 % (table 9). The NK1 fragment crystallises in space group P2₁, where a two-fold screw axis lies along y. Therefore, the position of the first molecule probe in the crystal unit cell can be chosen arbitrarily on this direction. The translation search is only performed in a two-dimensional space of coordinates x and z. The position of the first molecule will thus automatically fix the origin of the system, and the positions of the rest of the molecules will be defined relative to this origin.

Good correlation coefficient and R-factor values in a successful molecular replacement application are normally of the order of 50 – 60 % and 40 – 45 % respectively. However, in our case we are using only 1/4 of the overall scattering matter and this accounts for the lower quality of the statistics. The presence of only one peak in the output, however, means that this peak dominates the translation function map, which, in turn, suggests that the rotational angles and the translational vector are correct (in the following text this solution is referred as the kringle solution #1). This solution was then fixed, and the translation function of the N-domain probe was calculated for each rotation function solution. Since the correct rotation function peak for this probe had not been unambiguously determined, all the peaks from the list had to be checked.

The translation function calculated using the N-domain rotation solution #7 showed a peak with correlation coefficient of 34.9 % and R-factor of 52.9%, which are both better then the values of the kringle solution #1 (table 10). Other N-domain rotation function solutions gave translation function peaks having only R-factor values and correlation coefficients of around 54 – 55 % and 25 – 30 % respectively. Therefore, the correctness of this solution (N-domain solution #7) was assumed.

This solution was then kept fixed together with the kringle solution #1 and the translation function for all the other rotation function peaks of the N-domain probe was computed. The rotation solution #5 gave a translation function peak with correlation coefficient of 37.2 % and R-factor of 52.0 %, both values being better than the previously observed ones. This is, therefore, a good indication that this peak is the correct solution (in the following text this solution is referred as the N-domain solution #5). Table 11 shows the list of translation function peaks for the rotation function solution #5.

Finally, the translation for the second orientation of the kringle domain probe was required. The kringle solution #1, N-domain solution #7 and N-domain solution #5 were fixed, and the translation function for the kringle rotation function solution #2 was calculated. The list of peaks of this calculation is given in table 12.
Table 11: The list of the translation function peaks for the 5th rotation function solution of the N-domain probe (kringle solution #1 and N-domain solution #7 are fixed). The peak of the correct translation function solution is highlighted.

<table>
<thead>
<tr>
<th>Rotation peak Number</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>$T_x$</th>
<th>$T_y$</th>
<th>$T_z$</th>
<th>Corr. Coef</th>
<th>R-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>237.22</td>
<td>81.68</td>
<td>3.30</td>
<td>0.9519</td>
<td>0.5757</td>
<td>0.3769</td>
<td>37.2</td>
<td>52.0</td>
</tr>
<tr>
<td>5</td>
<td>237.22</td>
<td>81.68</td>
<td>3.30</td>
<td>0.9442</td>
<td>0.1980</td>
<td>0.8818</td>
<td>34.1</td>
<td>53.0</td>
</tr>
<tr>
<td>5</td>
<td>237.22</td>
<td>81.68</td>
<td>3.30</td>
<td>0.0015</td>
<td>0.1890</td>
<td>0.7021</td>
<td>33.8</td>
<td>53.2</td>
</tr>
<tr>
<td>5</td>
<td>237.22</td>
<td>81.68</td>
<td>3.30</td>
<td>0.7053</td>
<td>0.2582</td>
<td>0.9017</td>
<td>33.7</td>
<td>53.1</td>
</tr>
</tbody>
</table>

Resolution limits of 20 - 3.0 Å were used for the calculation.

Table 12: The list of the translation function peaks for the 2nd rotation function solution of the kringle probe (kringle solution #1 and N-domain solution #5 and #7 are fixed). The peak of the correct translation function solution is highlighted.

<table>
<thead>
<tr>
<th>Rotation peak Number</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>$T_x$</th>
<th>$T_y$</th>
<th>$T_z$</th>
<th>Corr. Coef</th>
<th>R-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>321.50</td>
<td>34.11</td>
<td>44.06</td>
<td>0.3500</td>
<td>0.3498</td>
<td>0.4663</td>
<td>38.3</td>
<td>51.8</td>
</tr>
<tr>
<td>2</td>
<td>321.50</td>
<td>34.11</td>
<td>44.06</td>
<td>0.9254</td>
<td>0.0224</td>
<td>0.7337</td>
<td>36.9</td>
<td>52.8</td>
</tr>
<tr>
<td>2</td>
<td>321.50</td>
<td>34.11</td>
<td>44.06</td>
<td>0.0069</td>
<td>0.4868</td>
<td>0.9604</td>
<td>36.9</td>
<td>52.6</td>
</tr>
<tr>
<td>2</td>
<td>321.50</td>
<td>34.11</td>
<td>44.06</td>
<td>0.2792</td>
<td>0.8548</td>
<td>0.2544</td>
<td>36.9</td>
<td>52.7</td>
</tr>
</tbody>
</table>

Resolution limits of 20 - 3.0 Å were used for the calculation.

The top translation function peak showed a correlation coefficient of 38.3 % and a R-factor of 51.8 %, and again these values are better than those computed before, confirming that this solution is the correct one for the orientation and the translation of the second kringle domain in the unit cell.

The final statistics of all four solutions of the N-domain and the kringle search probes are presented in table 13. The correctness of the solutions can be confirmed by the fact that the correlation coefficient increases and the R-factor decreases as the number of domains which are kept fixed gets bigger.

2.2.8 Refinement of the Molecular Replacement Solutions

After finding the solutions to the rotation and translation functions, the rotation angles and translation parameters are usually refined in AMoRe to minimise the quadratic misfit described by the following formula [18]:

$$Q(k, B, \alpha, \beta, \gamma, x, y, z) = \sum_{h} |F_{obs}(\vec{r})| e^{-2}\theta |2 - k| F_{calc}^{\alpha,\beta,\gamma,x,y,z}(\vec{r})|^2$$  \hspace{1cm} (29)

where $k$ is the scale factor and $B$ is the overall temperature factor (they are also refined together with the rotational and the translational parameters). The refinement algorithm is performed in AMoRe by the program FITTING. This rigid-body refinement is considered to be another checking procedure, to prove

Table 13: The statistics of molecular replacement solutions after sequential addition of the domains.

<table>
<thead>
<tr>
<th>P</th>
<th>RF#</th>
<th>α</th>
<th>β</th>
<th>γ</th>
<th>$T_x$</th>
<th>$T_y$</th>
<th>$T_z$</th>
<th>Corr. Coef</th>
<th>R-factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>1</td>
<td>30.57</td>
<td>33.29</td>
<td>4.21</td>
<td>0.0263</td>
<td>0.0000</td>
<td>0.0625</td>
<td>34.2</td>
<td>53.9</td>
</tr>
<tr>
<td>N</td>
<td>+7</td>
<td>257.33</td>
<td>64.92</td>
<td>8.58</td>
<td>0.5515</td>
<td>0.0352</td>
<td>0.9106</td>
<td>34.9</td>
<td>52.9</td>
</tr>
<tr>
<td>N</td>
<td>+5</td>
<td>237.22</td>
<td>81.68</td>
<td>3.30</td>
<td>0.9519</td>
<td>0.5757</td>
<td>0.3769</td>
<td>37.2</td>
<td>52.0</td>
</tr>
<tr>
<td>K</td>
<td>+2</td>
<td>321.50</td>
<td>34.11</td>
<td>44.06</td>
<td>0.3500</td>
<td>0.3498</td>
<td>0.4663</td>
<td>38.3</td>
<td>51.8</td>
</tr>
</tbody>
</table>

Abbreviations: P – probe; RF# – rotation solution number; K – kringle domain ; N - N-domain.
the correctness of the solutions. If the solutions are correct, the refinement should result in an increase of the correlation coefficient and a decrease of the R-factor.

When several molecules are present in the asymmetric unit, at each cycle of the refinement the optimal scale and temperature factors are determined. The least-squares minimisation, with respect to the rotation and position parameters, is alternately performed for each solution while the others are kept fixed.

A rigid-body refinement of the complete kringle and N-domain molecular replacement model yielded the final correlation coefficient of 49.7 % and R-factor of 49.2 % between the observed and the calculated structure factor amplitudes (resolution limits 20 – 3.0 Å). These values are much lower that the initial ones, confirming that these solutions are the correct ones.

2.3 Molecular replacement solution validation

To verify the correctness of the obtained solutions several validation procedures were used:

- The molecular replacement protocol was repeated using the probes that were rotated and shifted according to the solutions obtained. The resulting peaks on the rotation and translation functions were located at the origin.

- The additional search was performed by trying to find other solutions by fixing the four domains in a different order. The obtained results were consistent with the initial solutions.

- Normalised structure factor amplitudes were also tested. The obtained results were also consistent with the initial solutions. The following sections describe further checks on the conclusions reached.

2.3.1 Analysis of the crystal packing

The N-domain and the kringle probes were rotated and shifted according to the obtained molecular replacement solutions, using the PDBSET (CCP4). The molecular packing of these domains was analysed using graphical software, such as the MOLPACK [28] and O [11] programs. Molecules were evenly distributed through the whole unit cell. No clashing between molecules or close contacts were observed. The crystal packing within the unit cell is shown in figure 12.

2.3.2 Non-crystallographic symmetry and self-rotation function

The self-rotation function for the NK1 fragment crystals X-ray diffraction data was calculated using the POLARRFN (CCP4). It calculates the self-rotation function in polar angles. If there is a non-crystallographic two-fold or two-fold screw axis, a peak additional to those given by the proper symmetry elements will be found. A resolution range of 8 – 3.0 Å and an integration radius of 20 Å were used for this calculation. A graphical representation of the NK1 self-rotation function, with $\chi = 180^\circ$, is given in figure 13. On this section a peak close to the crystallographic two-fold screw axis is clearly shown. Thus, the two molecules in the asymmetric unit of the NK1 fragment crystals are related by a non-crystallographic two-fold symmetry.

How does this experimental evidence fit the obtained molecular replacement solutions? The two-fold axis cannot be clearly seen when the packing diagram of four domains within just one unit cell is shown (see figure 12). The packing of the molecules was, therefore, extended to two unit cells in each direction, and it was analysed using graphical programs. The two-fold axis could be immediately spotted, and to visualise it more clearly the crystallographic two-fold screw operator and unit cell translations were applied to the initially positioned domains, as listed in the table 14. The domain arrangements after the applied transformations are shown in figure 14 and the packing view of these arrangements within the unit cell in figure 15. The carboxy-termini of the N-domains are close enough to the amino-termini of the kringle domains to fit the linker region (two residues in length). This is yet another good indication that the molecular replacement solutions are correct.

To find out the precise relationship between the domains, the following procedure was used. The N-domain A and kringle domain C were considered to be one NK1 molecule, and the N-domain B and kringle domain D another (figure 14). Then the first NK1 molecule was superimposed onto the second using the LSQMAN program from the RAVE crystallographic package [10]. The output from this program
Figure 12: Packing of the obtained molecular replacement solution. The crystal packing arrangement of the N- and the kringle domains in the crystal unit cell. The two-fold screw axis is along the b direction. The projection along the a direction is shown (a) as well as the projection along the b direction (b). The N-domains are shown in green and blue colours, whereas the kringle domains in yellow and magenta. Same colouring of the molecules indicate that these domains are related by the crystallographic two-fold screw axis. Both pictures were produced using program SETOR [6].
Figure 13: Self-rotation function for the NK1 fragment crystal. The $\chi = 180.0^\circ$ section of the self-rotation function for the NK1 fragment crystal. The peak corresponding to non-crystallographic two-fold symmetry as well as the crystallographic two-fold screw symmetry are indicated by the arrows. A resolution range of 8 - 3.0 Å and an integration radius of 20 Å were used for the calculation. The map is contoured at 0.5. Position of this NCS peak in polar angles is $\omega = 77.4^\circ$, $\phi = 80.1^\circ$, $\chi = 180.0^\circ$.

Table 14: The operators that bring the kringle and N-domains in the asymmetric unit close to each other, revealing the two-fold non-crystallographic axis.

<table>
<thead>
<tr>
<th>The domain</th>
<th>Rotation</th>
<th>Translation (x y z), in fractions of the unit cell</th>
</tr>
</thead>
<tbody>
<tr>
<td>KRINGLE #1</td>
<td>2_1</td>
<td>-</td>
</tr>
<tr>
<td>KRINGLE #2</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>N-DOMAIN #1</td>
<td>2_1</td>
<td>1.000 0.000 1.000</td>
</tr>
<tr>
<td>N-DOMAIN #2</td>
<td>-</td>
<td>-1.000 0.000 0.000</td>
</tr>
</tbody>
</table>

contained the description of the operator that needed to be applied to achieve such superposition. The spherical polar angles of the operator were: $\omega = 77.8^\circ$, $\phi = 78.6^\circ$, $\chi = 179.8^\circ$; these values are very similar to those obtained from the self-rotation function, $\omega = 77.4^\circ$, $\phi = 80.1^\circ$, $\chi = 180.0^\circ$. Thus, the molecular replacement solutions are consistent with the self-rotation function results.

The proximity of the termini makes it impossible to determine which of the domains actually belong to one molecule. The described calculation was, therefore, repeated in the situation where the N-domain A and the kringle domain D were considered to be one NK1 molecule, and domains C and B another. The results of this calculation were similar to those obtained in the first instance.

3 Conclusions

The molecular replacement method represents a powerful technique for protein crystal structure determination, if the structure of an homologous protein is known. However, some of the factors can introduce problems in the successful application of this method. These factors include: the presence of multiple copies of a protein in the asymmetric unit, the presence of non-crystallographic symmetry, an incomplete or NMR model of the search probe.

This chapter presents a description of how the molecular replacement method was applied to determine the initial model of the NK1 fragment crystal structure. Calculations were performed with the state-of-the-art molecular replacement software package AMoRe. The search probes were as follows: the restrained minimised structure of 23 NMR models and the crystal structure of plasminogen kringle 1.

Prior to the calculation, the search probes were edited using a molecular graphics program in order to alter (remove or replace to Ala/Gly) residues that differ in the target structure. The residues 31 – 38,
Figure 14: The non-crystallographic symmetry of the NK1 molecules. The non-crystallographic two fold relationship between the N- and kringle domains of the two NK1 molecules is shown. The view is down the non-crystallographic axis (shown as black oval). The termini of N-domain A and kringle domain C are labelled. The figure was produced using the programs MOLSCRIPT [13] and RASTER 3D [16].
Figure 15: Packing of the NK1 dimers in the unit cell. The crystal packing arrangement of the NK1 dimers in the crystal unit cell. The two-fold screw axis is along the b direction. The projection along the a direction is shown (a) as well as the projection along the b direction (b). The N-domains are shown in green and blue colours, whereas the kringle domains in yellow and magenta. Same colouring of the molecules indicate that these domains are related by the crystallographic two-fold screw axis. Both pictures were produced using program SETOR [6].
77 – 79, and 127 of the N-domain probe were removed owing to their high r.m.s.d. in the NMR model, and the atomic B-factors of this probe were set to 30 Å².

The rotation function search using the kringle domain probe showed two peaks at the top of the list with signal-to-noise ratios of 1.45σ and 0.56σ respectively (see table 5). The output list of possible solutions of the rotation function search using the probe for the N-domain did not contain any strong (more than 0.2σ signal-to-noise ratio) peaks. However, analysis of the translation function calculated for each of the rotation function solutions from the list showed that the correct rotation function peaks for the N-domain probe were at the fifth and seventh positions in the list.

The N-domain and the kringle domain probes were placed in the crystal cell according to the rotational and translational parameters obtained, yielding the values for the R-factor and correlation coefficient between observed and calculated structure factor amplitudes of 51.8 and 38.3 % respectively. The refinement of the molecular replacement rotational and translational parameters decreased the R-factor to 49.2 % and increased the correlation coefficient to 49.7 % respectively, for the data set of resolution range 20 – 3.0 Å.

The analysis of the domains’ packing in the crystal unit cell revealed that the molecules are evenly distributed through the whole unit cell, and no clashing between the molecules was observed. The non-crystallographic two-fold relationship between the NK1 molecules in the asymmetric unit was consistent with the results of the self-rotation function.

The analysis described above indicates that the obtained solution for the NK1 fragment molecular replacement problem is the correct one.

References


