MARKSTEINER, P., BLAHA, P. & SCHWARZ, K. (1986). *Z. Phys.* B64, 119-127.
NAGEL, S. (1985). *J. Chem. Phys. Solids* 46, 743-756.
PETCOV, A. (1989). DESY F41. Internal Report 89-08. DESY, Hamburg, Germany.
PETCOV, A., KIRFEL, A. & FISCHER, K. (1990). *Acta Cryst.* A46, 754-763.
RESTORI, R. & SCHWARZENBACH, D. (1987). *Acta Cryst.* B43, 251-257.
TEMPLETON, D. H. & TEMPLETON, L. K. (1980). *Acta Cryst.* A36, 237-241.
TEMPLETON, D. H. & TEMPLETON, L. K. (1982). *Acta Cryst.* A38, 62-67.
TEMPLETON, D. H. & TEMPLETON, L. K. (1985*a*). *Acta Cryst.* A41, 133-142.
TEMPLETON, D. H. & TEMPLETON, L. K. (1985*b*). *Acta Cryst.* A41, 365-371.
TEMPLETON, D. H. & TEMPLETON, L. K. (1986). *Acta Cryst.* A42, 478-481.
TEMPLETON, D. H. & TEMPLETON, L. K. (1987). *Acta Cryst.* A43, 573-574.
TEMPLETON, D. H. & TEMPLETON, L. K. (1988). *Acta Cryst.* A44, 1045-1051.
THULER, M. R., BENBOW, R. L. & HURYCH, Z. (1982). *Phys. Rev. B*, 26, 669-677.

Acta Cryst. (1991). A47, 195-204

Solution of a Fab (26-10)/Digoxin Complex by Generalized Molecular Replacement

BY AXEL T. BRÜNGER

The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511, USA

(*Received* 11 *July* 1990; *accepted* 22 *October* 1990)

Abstract

The structures of a Fab fragment of a monoclonal murine antidigoxin antibody (26-10) complexed with digoxin and of a mutant of the Fab itself have been solved by molecular replacement. The solution strategy employed a generalization of molecular replacement. Prior to translation searches, a large number of the highest rotation-function peaks were subjected to a rigid-body refinement against the linear correlation coefficient between intensities of observed and calculated structure factors in which first the overall orientation of the model and then the orientations and translations of the individual domains (V_H, V_L, C_H1 and C_L) were refined. This procedure clearly identified the correct orientation of the search model. Furthermore, it produced a significant and unambiguous solution for the translation search. After rigid-body refinement, the *R* factor was in the low forties at 8-2·5 and 8-2·7 Å resolution for the Fab mutant and the Fab/digoxin complex, respectively. One round of simulated annealing refinement of all atomic positions reduced the *R* factor to the low twenties in both cases.

Introduction*

Molecular replacement, which is sometimes also referred to as Patterson search, involves the placement (*i.e.* rotation and translation) of the known structure of a search model in the unit cell of the target crystal in order to obtain the best agreement between calculated model diffraction data and the observed diffraction data (Hoppe, 1957; Rossmann & Blow, 1962; Huber, 1965). Unfortunately, molecular replacement often fails if the search model is too inaccurate. It does not require large structural changes to make the search model too inaccurate for molecular replacement. A case in point has been reported by Brünger, Campbell, Clore, Gronenborn, Karplus, Petsko & Teeter (1987) where molecular replacement had failed with a r.m.s. difference in backbone positions between the search model and the crystal structure of around 1·4 Å; variation of the parameters of the molecular replacement (number of reflections, resolution range, temperature factors and occupancy of atoms of the search model) made little difference.

Recently, we have generalized molecular replacement by introducing additional parameters *p* in order to make the search model more accurate (Brünger, 1990*a*). First, a conventional rotation search is carried out. The highest peaks of the rotation search are selected. Here we make the *ad hoc* assumption that the correct orientation is among the highest peaks of the rotation function. Then the rotation search is 'filtered' by employing refinements of the parameters *p* against the negative correlation coefficient PC for each selected orientation of the search model. Refinement is carried out against the negative correlation coefficient since minimization algorithms normally locate minima as opposed to maxima; a minimum of −PC corresponds to a maximum of PC

* Abbreviations: CDR, complementarity determining region; CPU, central processing unit; Fab, antigen binding fragment of an antibody; PC, standard linear correlation coefficient between $|E_{obs}|^2$ and $|E_{model}|^2$; r.m.s., root-mean-square; SA, simulated annealing; σ, standard deviation.

0108-7673/91/030195-10$03.00

Table 1. *Data sets*

Data set	Resolution range (Å)	Number of reflections* 15-3·5 Å, $\|F_{obs}\| > 1{\cdot}5\sigma$	Completeness† (%)	a (Å)	b (Å)	c (Å)	β (°)
3*r*9	61-2·5	11655	98·1	44·168	164·74	70·18	108·48
3*r*9*a*-det	42-2·5	8404	70	44·144	164·69	70·17	108·50
adig	29-2·7	9917	84	44·158	164·43	70·01	108·37
adig-det	21-2·7	8446	71	44·158	164·43	70·01	108·37

* Number of observed unique reflections with $|F_{obs}| > 1{\cdot}5\sigma$ for the resolution range 15-3·5 Å.
† Ratio of the number of observed reflections with $|F_{obs}| > 1{\cdot}5\sigma$ and the number of theoretically possible reflections for the resolution range 15-3·5 Å.

which is what one is really aiming at. Examples of parameters p are the orientation and position of groups of atoms represented as rigid bodies, atomic coordinates, temperature factors or occupancies. The refinement against $-$PC may be combined with an empirical energy function in order to improve the ratio of observables to parameters as is routinely done in R-factor refinement (Jack & Levitt, 1978; Hendrickson, 1985; Brünger, Kuriyan & Karplus, 1987). However, the major difference to R-factor refinement is that PC refinement is carried out in P_1, that is, without crystallographic symmetry. Crystallographic symmetry cannot be applied since the molecule has not yet been positioned correctly in the unit cell. Minimizing the negative correlation coefficient $-$PC is equivalent to minimizing the phase error (Hauptman, 1982). The final step of our strategy consists of translation searches using the refined search models that yield the highest correlation coefficients after PC refinement.

Here we report the successful application of our generalized molecular replacement strategy to the solution of the complex of the antigen-binding fragment (Fab) of a monoclonal murine anti-digoxin antibody (26-10) with digoxin and of a mutant of the Fab itself (Mudgett-Hunter, Margolies, Ju & Haber, 1982), both of which crystallized isomorphously in space group $P2_1$ with two molecules in the asymmetric unit. This paper contains the thorough description of the structure-solution process. Details of data collection, structure refinement and analysis of the refined structure are published elsewhere (Strong, 1990; Strong and co-workers, in preparation). It should be noted that two other Fab fragments have recently been solved by our generalized molecular replacement method (Brünger, Leahy, Hynes & Fox, 1991; Garcia, Ronco, Verroust & Amzel, 1989; Garcia and co-workers, work in progress). In both cases the Fab fragments crystallized in high-symmetry space groups ($P6_522$ and $P4_1$) which shows that our methodology is not restricted to low-symmetry space groups. The method can easily be applied to the solution of multidomain proteins other than Fabs, whose domains have flexible orientations relative to one another.

The strategy to solve the Fab 26-10 consisted of a self-rotation function to determine the existence of a non-crystallographic symmetry element relating the two molecules in the asymmetric unit, a cross-rotation function with the known structure of a homologous Fab to generate a large number of likely orientations of the two molecules, PC refinement to filter the correct orientations among those produced by the cross-rotation function, translation searches of the correctly oriented molecules and, finally, rigid-body refinement of the correctly placed molecules.

Materials and methods

Diffraction data

Table 1 lists relevant properties of the four data sets that were used in this report. Data set 3*r*9 is the native data from a mutant of the Fab 26-10 which has six substitutions in the heavy-chain variable region and exhibits altered binding affinity for hapten (Hudson, Mudgett-Hunter, Panka & Margolies, 1985). This data set is the merged product from two crystals *a*, *b*, where the *a* crystal accounts for more than 70% of the unique data. Data set adig is the native data from a single crystal of the Fab 26-10/digoxin complex. Data were collected on area detectors; details are published elsewhere (Strong, 1990). R_{sym} for the 3*r*9 and adig data sets is 4·5 and 3·2%, respectively. All crystals exhibit space-group symmetry $P2_1$ unique axis *b*, with two molecules in the asymmetric unit. Both data sets were twinned. A de-twinning procedure was employed to reduce the residual amplitudes from the weaker one of the twins. Details of the de-twinning procedure will be published elsewhere (Strong, in preparation). 3*r*9*a*-det is the de-twinned data set derived from the *a* crystal portion of data set 3*r*9. adig-det is the de-twinned data set derived from data set adig. Throughout this work reflections with $|F_{obs}| > 1{\cdot}5\sigma_{obs}$ were used.

Search model

The search model for molecular replacement was the Fab portion of the HyHEL-5 Fab/lysozyme complex (Sheriff *et al.*, 1987). In the numbering scheme of HyHEL-5, the four domains of the heavy and light chain of the variable and constant domains are defined as $V_L = 1$-106; $C_L = 107$-212; $V_H = 1001$-1116; $C_H1 = 1117$-1218. The elbow angle is

defined as the angle between the pseudo twofold axes of symmetry of the V_L-V_H and C_L-C_H1 domain pairs. The elbow angle for HyHEL-5 is 161·1°. The sequence identity between the Fabs 26-10 and HyHEL-5 is 43% for V_L, 61% for V_H, 100% for C_H1 and 100% for C_L.

The temperature factors of the search model were set to twice the values of the original HyHEL-5 structure. This was done in order to amplify the effect of the *B* factors, *i.e.* less well determined regions contribute less to the Patterson map than with the original *B* factors. All atoms were included in the search model. After molecular replacement and rigid-body refinement, residues of the HyHEL-5 V_L and V_H domains were substituted to account for the sequence of the Fab 26-10. Nine residues near or at the complementarity determining region 1 (CDR1) and 12 residues near or at CDR3 were excluded during the subsequent refinement.

Computational techniques

All calculations (molecular replacement, PC refinement and rigid-body refinement) were carried out with version 2.1 of the program *X-PLOR* (Brünger, 1990*b*). The new version of *X-PLOR* is available upon request from the author. All calculations were carried out on a Convex-C220 and on a Cray-YMP.

Results

Self-rotation search

Density-gradient measurements indicated the presence of two molecules in the asymmetric unit. A self-rotation search was performed to locate a possible non-crystallographic symmetry operation relating the two molecules in the asymmetric unit. A real-space Patterson-search method (Huber, 1985) was employed. The crystal Patterson map was calculated on a 1·1 Å grid using the observed reflections from 15 to 4 Å resolution. The 4000 strongest Patterson vectors with lengths between 24 and 5 Å resolution were selected. The selected Patterson vectors (P_{model}) were rotated using the spherical polar angles (ψ, φ, κ) and the product correlation

$$\langle P_{obs} P_{model}(\psi, \varphi, \kappa) \rangle, \qquad (1)$$

where the crystal Patterson map (P_{obs}) was computed by linear eight-point interpolation (Nordman, 1980). ψ is the inclination of the rotation axis *vs* the *y* axis, φ is the azimuthal angle, that is, the angle between the *x* axis and the projection of the rotation axis into the *x*, *z* plane, κ is the rotation around the rotation axis. The self-rotation search was carried out by varying ψ and φ in steps of 2° in the angular range $\psi = 0$ to 180° and $\varphi = 0$ to 180° while κ was held fixed at 180°. A contour plot of the self-rotation search for data set 3*r*9*a*-det is shown in Fig. 1. Apart from the crystallographic diad at $\psi = 0°$, a strong non-crystallographic symmetry element emerges at $\psi = 90$, $\varphi = 0°$ and at the symmetry-related position ($\psi = 90$, $\varphi = 90°$). The self-rotation functions of the other three data sets are very similar to Fig. 1. Thus, the Fab 26-10 data sets exhibit pseudo-orthorhombic symmetry.

Cross-rotation search

The real-space Patterson search method of Huber (1985) was employed for rotation searches with the HyHEL-5 search model against each of the four data sets listed in Table 1. The crystal Patterson map was computed from the observed intensities by fast Fourier transformation at 15–4 Å resolution on a 1·1 Å grid. The model Patterson maps were computed by placing the HyHEL-5 search model into a 120 × 120 × 150 Å orthorhombic box with the largest extent of the model approximately parallel to the *z* direction, evaluating the structure factors, and fast Fourier transformation of the squared amplitudes on a 1 Å grid. The box size was chosen such that it was about twice as large as the model in each direction to avoid overlap of self Patterson vectors. Model Patterson vectors were selected according to length (between 24 and 4 Å) and according to peak height (3σ above the mean of the model Patterson map). This resulted in about 3200 selected Patterson vectors. The selected Patterson vectors were rotated using the Eulerian angles (θ_1, θ_2, θ_3) as defined by Rossmann & Blow (1962). The product correlation with the crystal Patterson map was computed by linear eight-point

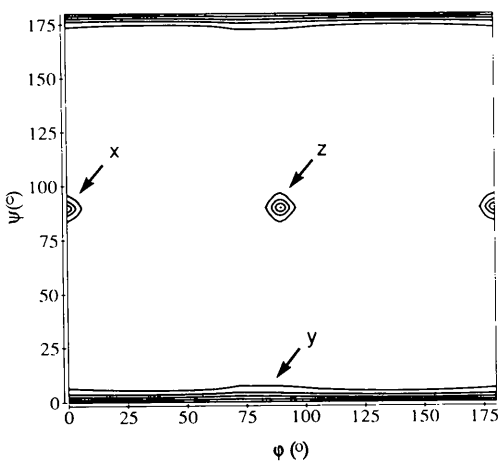

Fig. 1. Contour plot showing the self-rotation function of the 3*r*9*a*-det diffraction data at 15 to 4 Å resolution with κ held fixed at 180° (spherical polar coordinates, see text). The mean of the self-rotation function is 1·34 and σ is 0·51. The maximum (3·8) is assumed at the crystallographic diad (marked *y*), whereas the peak height for the non-crystallographic symmetry elements (marked *x*, *z*) is 3·3. The lowest contour is drawn at 1σ above the mean and the highest contour level is drawn at the maximum of the self-rotation function.

interpolation (Nordman, 1980). The orientations of the search models were sampled by using the pseudo-orthogonal Eulerian angles introduced by Lattman (1972):

$$\theta_+ = \theta_1 + \theta_3$$
$$\theta_- = \theta_1 - \theta_3 \qquad (2)$$
$$\theta_2.$$

To ensure proper sampling of the rotation function, the sampling interval Δ for θ_2 should be less than $\arcsin[d/(3r)]$ where d is the high-resolution limit of the Patterson maps and r is the maximum Patterson vector length of the search model. The interval Δ was set to 2·5° which is a rather conservative choice considering the selected maximum Patterson vector length of 24 Å. Following Lattman (1972) the interval for θ_+ is given by $\Delta/\cos(\theta_2/2)$ and the interval for θ_- is given by $\Delta/\sin(\theta_2/2)$. As the crystal symmetry is monoclinic $P2_1$, b axis unique, the rotation search could be restricted to the asymmetric unit $\theta_+ = 0, \ldots, 720°$, $\theta_2 = 0, \ldots, 90°$, $\theta_- = 0, \ldots, 360°$ as shown by Rao, Jih & Hartsuck (1980).

All grid points of the rotation function were sorted with respect to their rotation-function value and the peak search described by Brünger (1990*a*) was carried out. Briefly, for two given rotation matrices Ω^1 and Ω^2 a metric is defined as

$$m(\Omega^1, \Omega^2) = \min_{s=1,n_s} \{\mathrm{tr}[(\Omega^1 - \mathcal{O}_s\Omega^2)(\Omega^1 - \mathcal{O}_s\Omega^2)^t]^{1/2}\}, \qquad (3)$$

where n_s is the number of symmetry operators of the space group of the crystal and $\mathcal{O}_s$ is the rotational part of the symmetry operator s. The metric m is related to the rotation angle κ around some axis that transforms matrix Ω^1 into Ω^2 or one of its crystal symmetry mates by the relationship

$$m(\Omega^1, \Omega^2) = [4(1 - \cos\kappa)]^{1/2}. \qquad (4)$$

This metric allows one to define clusters of grid points and a peak belonging to each cluster of grid points. Going down the list of grid points sorted by the value of the rotation function, a grid point is defined as a peak if its corresponding orientation Ω^1 yields $m(\Omega^1, \Omega^2) > \varepsilon$ for all previously defined peaks with orientations Ω^2, otherwise it belongs to the cluster of grid points around the peak with $m(\Omega^1, \Omega^2) \leq \varepsilon$. ε has to be chosen such that it is less than twice the radius of convergence of the subsequent PC refinement. In this work ε has been set to 0·4, that is, two rotation matrices with $m(\Omega^1, \Omega^2) < \varepsilon$ have to be related by a rotation around some axis of less than 16°. Because of the relatively large radius of convergence of PC refinement (10 to 13°, see Brünger, 1990*a*), this was a rather conservative choice. We have extended the *X-PLOR* program (Brünger, 1990*b*) to compute the rotation function and carry out the peak search. This simplified the data transfer from the rotation-function routine to the subsequent PC refinement. However, in principle, any other conventional rotation function (*e.g.* Fitzgerald, 1988) could have been used to generate the orientations of the search model to be checked by subsequent PC refinement.

Table 2. *First ten peaks of the rotation function for data set 3r9a-det*

Peak index	Rotation-function value σ	θ_1	θ_2, θ_3 (°)	
1	3·4216	194·95	35	61·935
2	3·4094	191·81	35	244·2
3	3·2203	335·1	15	277·64
4	3·1997	210·78	22·5	31·374
5	3·1953	166·89	42·5	244·08
6	3·1823	188·38	17·5	252·56
7	3·1525	205·27	25·2	18·75
8	3·1359	209·36	27·5	243·3
9	3·115	194·01	37·5	46·239
10	3·107	330·51	15	100·67

15 000 grid points corresponding to the highest rotation-function values were analyzed by the peak search for each of the four data sets listed in Table 1. This resulted in 145 peaks for data set 3*r*9*a*-det, 103 peaks for data set 3*r*9, 195 peaks for data set adig-det and 202 peaks for data set adig (Figs. 2*a*, 3*a*, 4*a*, 5*a*). Many peaks of the rotation functions emerge as pairs related by the non-crystallographic symmetry. For instance, in the case of data set 3*r*9*a*-det, peak 1 is related to peak 2 and peak 3 is related to peak 10 by the non-crystallographic twofold symmetry (Table 2). Note that the non-crystallographic twofold axis is simply represented as $\theta_2 + 180°$ in Eulerian angle space. Peaks 1 to 3 are not even close to the correct orientation (Table 4). Thus, the appearance of a pair of peaks is a necessary but not a sufficient criterion to discriminate among correct and incorrect orientations.

Table 3 lists the orientations which correspond to the highest peaks of various rotation functions. In addition to rotation functions using the four data sets and the full atomic model we have also computed a rotation function using a partial model consisting only of the constant domains as was proposed by Cygler & Anderson (1988). The five orientations are quite different from each other. In retrospect, none of the highest peaks are close to the correct orientations shown in Table 4, except for data set 3*r*9 using the full model in which case the highest peak is about 10° off the correct orientation for molecule *A*. However, even in this case a subsequent translation search fails (see *Concluding remarks*). Rotation searches with *MERLOT* (Fitzgerald, 1988) were also inconclusive (Strong, 1990).

Table 3. *Highest rotation-function peaks*

Data set	Rotation function value (σ)	θ_1 (°)	θ_2 (°)	θ_3 (°)
3*r*9	3·44	258·94	40	348·49
3*r*9*a*-det	3·42	194·95	35	61·935
3*r*9*a*-det, C_L+C_H1	2·76	186·54	27·5	70·84
adig	3·27	250·44	70	344·56
adig-det	2·79	83·18	12·5	190·58

Table 4. *Correct orientations (after PC refinement with data set 3r9a-det)*

Molecule	θ_1 (°)	θ_2 (°)	θ_3 (°)
A	267·54	35·56	333·34
B	268·63	34·36	154·81

Patterson refinement

In order to determine the correct orientations of the two molecules in the asymmetric unit among the most likely peaks predicted by the cross-rotation function we carried out PC refinement as described by Brünger (1990*a*) of the orientations and positions of the four Fab domains V_H, V_L, C_H1, C_L. Briefly, this consists of minimization of a target function

$$E_{tot}(\Omega, \Omega_i, t_i) = 1 - \mathrm{PC}(\Omega, \Omega_i, t_i), \qquad (5)$$

were Ω is the overall orientation of the search model, Ω_i and t_i are the individual orientations and translations of the four Fab domains, respectively. PC is the standard linear correlation coefficient,

$$\begin{aligned}\mathrm{PC}(\Omega, \Omega_i, t_i) = [&\langle|E_{obs}|^2|E_m(\Omega, \Omega_i, t_i)|^2\\ &-\langle|E_{obs}|^2\rangle\langle|E_m(\Omega, \Omega_i, t_i)|^2\rangle\rangle]\\ &\times[\langle|E_{obs}|^4-\langle|E_{obs}|^2\rangle^2\rangle\\ &\times\langle|E_m(\Omega, \Omega_i, t_i)|^4\\ &-\langle|E_m(\Omega, \Omega_i, t_i)|^2\rangle^2\rangle]^{-1/2}. \qquad (6)\end{aligned}$$

The symbols $\langle\,\rangle$ denote an averaging over the set of observed reflections expanded to P_1. E_{obs} denote the normalized observed structure factors and $E_m(\Omega, \Omega_i, t_i)$ denote the normalized structure factors of the search model placed in a triclinic unit cell identical in geometry to that of the crystal.

PC refinements of the HyHEL-5 search model were carried out with the search model oriented according to each of the selected peaks of the rotation functions for the four data sets. It should be noted that a large number of the PC refinements are actually carried out for incorrect orientations of the search model; only the PC refinements starting close to the correct orientation are expected to yield a relatively large correlation coefficient after PC refinement. The PC refinements consisted of 15 steps conjugate gradient minimization of E_{tot} using the method of Powell (1977) for the overall orientation Ω of the molecule, followed by 100 conjugate gradient steps for the orientational and translational parameters (Ω_i, t_i) of the Fab domains V_H, V_L, C_H1, C_L. PC refinements were carried out at 15–4 Å resolution (Figs. 2*c*, 3*c*, 4*c*, 5*c*) and 15–3·5 Å resolution (Figs. 2*b*, 3*b*, 4*b*, 5*b*). The grid size for the fast Fourier transformation of the structure factors was set to one third of the high-resolution limit.

Data set 3*r*9*a*-det and data set adig produced the most significant peaks, labelled *A* and *B* (Figs. 2*b*,*c*, 5*b*). The two corresponding orientations are listed in Table 4; they are related by the non-crystallographic twofold symmetry. Less-significant peaks corresponding to orientations close to those of peaks *A* and *B* were produced by the other data sets (Figs. 3, 4, 5). It should be noted that some peaks converged to the same orientation after PC refinement which explains why multiple *A* and *B* peaks were observed in Figs. 3, 4, 5. These 'redundant' peaks can sometimes cause confusion in the manual interpretation of PC refinement. The program *X-PLOR* actually lists the orientations of the search model after PC refinement

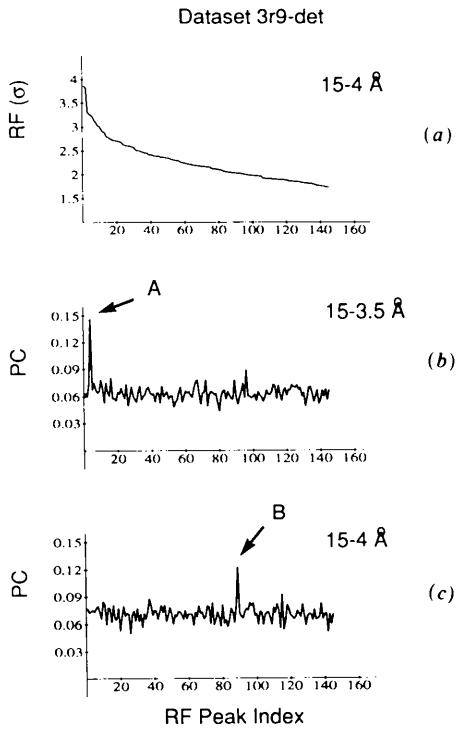

Fig. 2. Cross-rotation searches and PC refinements using the HyHEL-5 search model against the 3*r*9*a*-det diffraction data. (*a*) The values of the selected peaks of the conventional rotation function are shown in units of standard deviations (σ) above the mean sorted by value. Rigid-body PC refinements of the Fab domains were carried out for each orientation of the HyHEL-5 search model corresponding to the selected peaks of the rotation function. Shown are the correlation coefficients (PC) after the refinement at (*b*) 15–3·5 Å resolution and (*c*) 15–4 Å resolution.

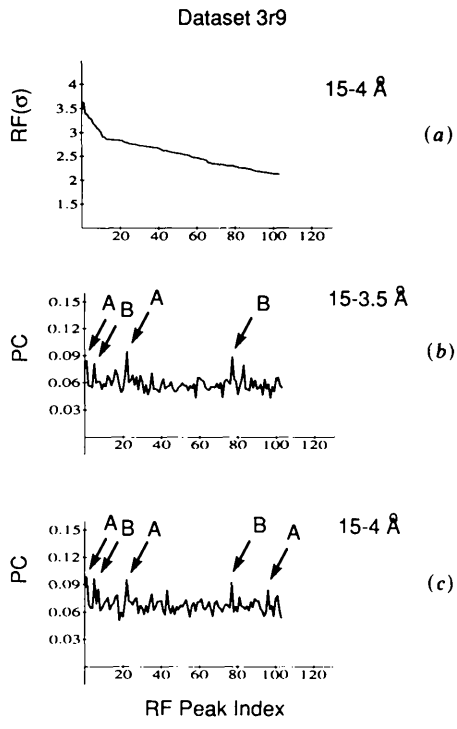

Fig. 3. Same as Fig. 2, but for the 3r9 diffraction data.

which could in principle make an automatic analysis possible.

The search model oriented according to peaks *A* and *B* is referred to as molecule *A* and *B*, respectively. The orientational and positional shifts for molecule *A* during PC refinement at 15–3·5 Å resolution (Fig. 2*b*) are listed in Table 5. During PC refinement of the overall orientation and position of molecule *A*, the orientation is changed by 25° around some axis. The large translational shift is due to numerical instabilities since the target function (5) is independent of the overall position of the molecule, but the current implementation of the minimization algorithm did not allow us to constrain positions independent of orientations. After PC refinement of the individual orientations and positions of the four Fab domains, the r.m.s. deviations between the original model and the PC-refined model are 2·3, 2·54, 2·36 and 4·22 Å for C^{α} atoms of the V_L, C_L, V_H and C_H1 domains, respectively, when the centers of the molecules are fitted. The r.m.s. deviation is 5·7 Å for C^{α} atoms when the constant domains of the Fab are fitted. This results in a change in the elbow angle by about 10°.

PC refinement is computationally expensive but is well within the limits of presently available supercomputing power. For instance, the 202 PC refinements for data set adig at 15–3·5 Å resolution required about 10 CPU h on a Cray-YMP which corresponds to roughly 60 CPU h on a Convex-C210.

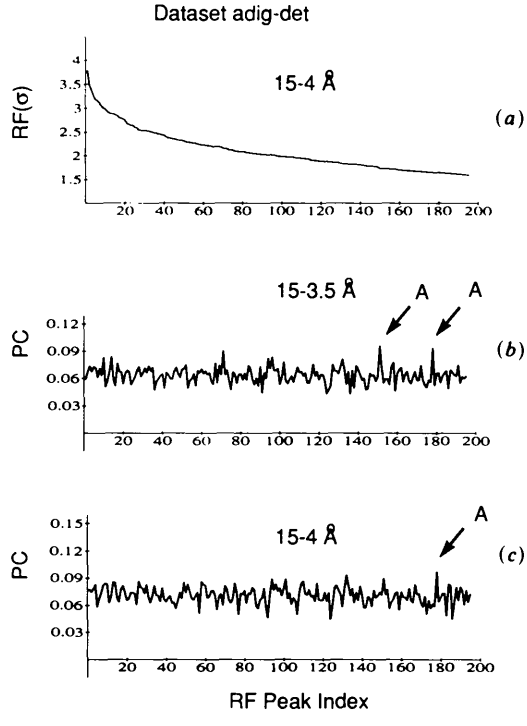

Fig. 4. Same as Fig. 2, but for the adig-det diffraction data.

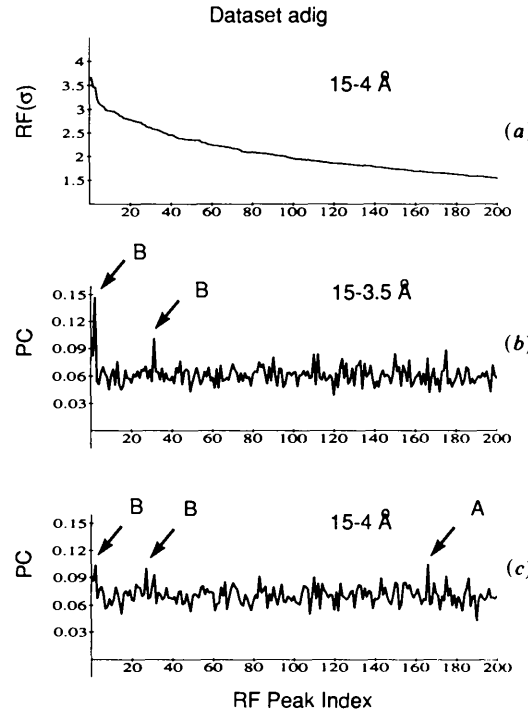

Fig. 5. Same as Fig. 2, but for the adig diffraction data.

Table 5. *Shifts of individual domains for molecule A produced by PC refinement*

Domain	θ_1 (°)	θ_2 (°)	θ_3 (°)	Δx (Å)	Δy (Å)	Δz (Å)	R.m.s. (Å)
$V_H+V_L+C_H1+C_L$	−7·1	−24·1	13·9	−3·7	−1·24	−0·11	11·62
V_L	5·83	1·26	4·97	−1·28	−0·96	0·30	2·30
C_L	−13·00	−4·99	−3·19	−0·82	0·76	0·53	2·54
V_H	3·38	2·09	9·03	−0·42	−0·62	−0·90	2·36
C_H1	−13·08	−6·45	−2·79	1·09	−2·15	−1·13	4·22

Listed are the Eulerian angles, the vectors and the corresponding atomic r.m.s. differences that represent the rotation and translations produced by rigid-body PC refinement of the search model at 15-3·5 Å resolution for data set 3*r*9*a*-det. The first row shows the result of PC refinement consisting of 15 conjugate gradients steps with the overall orientation of the search model ($V_H+V_L+C_H1+C_L$) (note: the translational shift is due to numerical inaccuracies since the target function (6) is independent of the overall position of the molecule). Rows two to five show the shifts during 100 conjugate gradient steps with the rotational and translational degrees of freedom of the individual four domains V_H, V_L, C_H1, C_L. The r.m.s. differences were computed for C^α atoms of the specified domains.

Translation search

After the correct orientation of the two molecules *A* and *B* had been determined, their positions were determined by translation searches. Translation searches using the method of Fujinaga & Read (1987) were carried out with the properly oriented and PC-refined HyHEL-5 search model. Briefly, the standard linear correlation coefficient between the normalized observed structure factors (E_{obs}) and the normalized calculated structure factors (E_{calc}) for the model at position *x*, *y*, *z* was computed,

$$\begin{aligned}\mathrm{TF}(xyz) = [&\langle|E_{obs}|^2|E_{calc}(xyz)|^2\\ &-\langle|E_{obs}|^2\rangle\langle|E_{calc}(xyz)|^2\rangle]\\ &\times[\langle|E_{obs}|^4-\langle|E_{obs}|^2\rangle^2\rangle\\ &\times\langle|E_{calc}(xyz)|^4-\langle|E_{calc}(xyz)|^2\rangle^2\rangle]^{-1/2} \qquad (7)\end{aligned}$$

where the symbols $\langle\,\rangle$ denote an averaging over the set of observed reflections. Equation (7) is identical to (6) except E_m is replaced by $E_{calc}(xyz)$, where the latter denotes the normalized calculated structure factors of the search model and *all of its symmetry mates.* Data were included between 15 and 3·5 Å resolution. The sampling interval in *x*, *y* and *z* was set to 1/3 of the high-resolution limit.

The translation-search strategy was as follows. First, the *x*, *z* position of the PC-refined molecules *A* and *B* (Table 4) were determined independently. The symmetry of the space group $P2_1$ with the *b* axis unique implies that the *y* component of the geometric center of the search model is arbitrary when searching with a single copy of the molecule. Figs. 6(*a*), (*b*) show the translation seraches starting with the properly oriented and PC-refined search model against the 3*r*9*a*-det data. Significant peaks emerge for both molecules *A* and *B*. For molecule *A*, the peak is at $x^A_{max}=0{\cdot}129$ and $z^A_{max}=0{\cdot}043$ in fractional coordinates, it is $9{\cdot}8\sigma$ above the mean and $6{\cdot}9\sigma$ above the next-highest peak. For molecule *B*, the peak at $x^B_{max}=0{\cdot}314$ and $z^B_{max}=0{\cdot}143$ is $7{\cdot}9\sigma$ above the mean and it is $5{\cdot}1\sigma$ above the next-highest peak.

Second, the relative position along *y* of molecules *A* and *B* was determined by varying the *y* position of molecule *B* while keeping molecule *A* fixed at x^A_{max}, $y=0$, z^A_{max}. Since the x^B_{max}, z^B_{max} position of molecule *B* is determined only up to an additive constant 0·5, four translation searches were required. The translation search that used $x^B_{max}+0{\cdot}5$, $z^B_{max}+0{\cdot}5$ produced an unambiguous and significant peak at $y=0{\cdot}39$ (shown in Fig. 6*c*). The peak has a height of about 9σ above the mean.

Rigid-body refinement

The final step of our structure-solution process consisted of rigid-body refinement of the properly oriented and positioned molecules in the unit cell of the crystal. Rigid-body refinement of the positions and orientations of the eight domains of molecules *A* and *B* were carried out by running 70 steps of conjugate gradient minimization at 8-3·5 Å resolution followed by 70 steps at 8-2·5 and 8-2·7 Å resolution for the 3*r*9, 3*r*9*a*-det and adig, adig-det

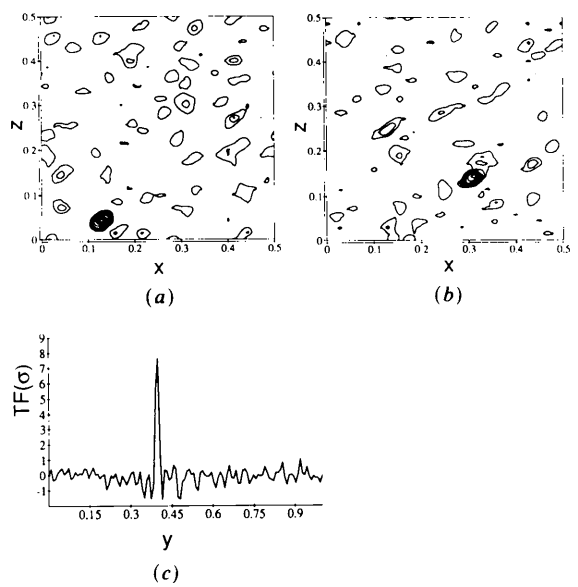

Fig. 6. Translation searches after successful PC refinement for (*a*) molecule *A*, (*b*) molecule *B* and (*c*) molecule *B* while *A* is held fixed at its correct position. Data set 3*r*9*a*-det was used.

Table 6. *R factor of the rigid-body-refined structure*

Data set	Resolution (Å)	*R* factor (%)
3*r*9	8–2·5	42·4
3*r*9*a*-det	8–2·5	42·0
adig	8–2·7	40·9
adig-det	8–2·7	42·8

data sets, respectively. Shifts produced by rigid-body refinement were less than 1° for angular parameters and less than 0·2 Å for translational parameters. The *R* factors after rigid-body refinement were in the low forties (Table 6). The *R* factor as a function of reciprocal resolution is shown in Fig. 7 for the 3*r*9*a*-det data set. The *R*-factor distribution is relatively uniform with values in the forties even at high resolution. The refined orientations and positions of the two molecules are listed in Table 7. At this stage of the refinement, the elbow angle for the Fab 26-10 was 171·5° compared to 161·1° for the HyHEL5 search model.

One round of simulated annealing (SA) refinement using the slow-cooling protocol described by Brünger, Krukowski & Erickson (1990) with a partial model (see *Materials and methods*) reduced the *R* factor to 21% at 6–2·5 Å resolution for data set 3*r*9*a*-det, and to 17% at 6–2·7 Å resolution for data set adig-det with good stereochemistry. At that point we considered the solution of 26–10 Fab and the 26-10 Fab/digoxin complex as complete. Further refinement of the structure is in progress (Strong, in preparation).

Concluding remarks

In conclusion, we address two questions that emerged during the structure solution process. First, why is the PC refinement of the orientations and positions of the four Fab domains resolution dependent? Second, could one have solved the structure with conventional molecular replacement?

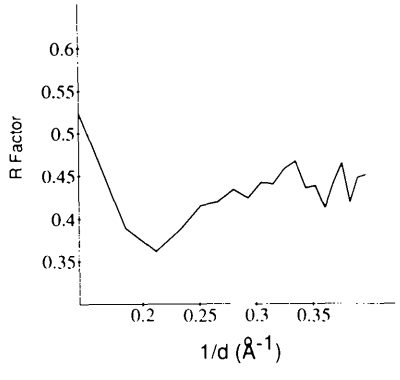

Fig. 7. *R* factor as a function of reciprocal resolution for the 3*r*9*a*-det data set after rigid-body refinement.

Table 7. *Orientations and positions after rigid-body refinement*

Molecule	θ_1, θ_2, θ_3 (°)			*x*, *y*, *z**		
A	267·44	35·84	333·38	0·067	0·077	0·100
B	266·98	35·64	155·27	0·843	0·321	0·740

* Fractional coordinates of the geometric center of the specified molecule.

PC refinement is subject to the multiple-minimum problem

Peaks *A* and *B* in Figs. 2(*b*),(*c*), 5(*b*) correspond to successful PC refinements with correlation coefficients around 0·14 whereas such high correlation coefficients are not achieved in the other cases which are marked in Figs. 2 to 5. Furthermore, the PC refinement for peak *A* is successful at 15–3·5 Å resolution, but not at 15–4 Å resolution, whereas *B* is successful at 15–4 Å resolution but not at 15–3·5 Å resolution (Figs. 2*b*,*c*). This reflects the multiple-minimum nature of PC refinement and the limited radius of convergence of conjugate gradient minimization. To illustrate this point, the correlation coefficient PC as a function of the elbow angle is shown in Fig. 8 for various resolution ranges. It appears that the global maximum of PC corresponding to the correct elbow angle (171·5 or 0° in Fig. 8) is resolution independent. There are, however, a number of local maxima that are resolution dependent. In reality, the situation is more complicated than shown in Fig. 8 since the PC refinement of the orientations and positions of the four Fab domains involves 24 parameters. This makes it even more probable to

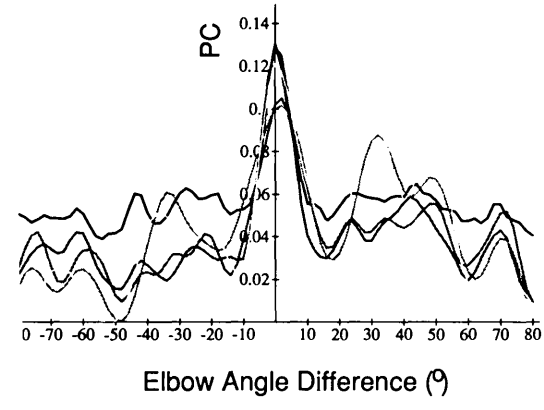

Fig. 8. PC as a function of the elbow-angle difference between an artificially modified Fab fragment and the correct Fab 26-10 structure with elbow angle 171·5°. The elbow angle is defined as the angle between the pseudo twofold axes of symmetry of the V_L–V_H and C_L–C_H1 domain pairs of the Fab. The elbow angle was modified by rotating the variable domains around an axis connecting the two linkage regions consisting of residue 106 of the light chain and residue 116 of the heavy chain. PC was evaluated at 15–2·5 (darkest line), 15–4, 15–5, 15–7, 15–8 (lightest line) Å resolution.

get trapped in a local maximum of PC. Since the original search model had an elbow angle of −10° it is conceivable that the conjugate gradient minimization got trapped in a local maximum at a certain resolution range but converged to the global maximum at a different resolution range. The detwinning has a similar effect to changing the resolution range: it shifts local maxima around whereas the global maximum is conserved (not shown). Efforts are under way to improve convergence behavior of PC refinement by using simulated annealing.

Could the structure have been solved without PC refinement?

As Tables 3 and 4 indicate, the correct orientation of molecule *A* is relatively close to the orientation corresponding to the highest peak of the conventional rotation function using data set 3*r*9. This poses the question whether it would have been possible to solve the structure by conventional methods without PC refinement. It is a customary procedure to test a rotation-function solution by carrying out a translation search. Fig. 9 shows the translation search for molecule *A* oriented according to the highest peak of the conventional rotation function for data set 3*r*9 listed in Table 3. The translation search is very noisy. The highest peak (indicated by an arrow), which is less than 1σ above other peaks, is relatively close to the correct position (*cf.* Fig. 6*a*). However, there is a 5 Å difference in position and 11° difference in orientation for molecule *A*. A similar situation arises for molecule *B* (not shown). The *R* factor after rigid-body refinement of molecules *A*, *B* placed according to these conventional translation/rotation searches is 49% at 8–2·6 Å. One round of SA refinement reduced the *R* factor to 31% at 6–2·6 Å, but the r.m.s. difference between this refined model and the correct structure is still 5 Å for backbone atoms. As a consequence, the refinement would have probably got 'stuck' and one would have abandoned this solution. Thus, PC refinement of the correctly oriented search model prior to the translation search was essential to obtain the correct position. This was also observed by Yeates & Rini (1990) for a number of other immunoglobulin structures who used a procedure similar to PC refinement prior to translation searches.

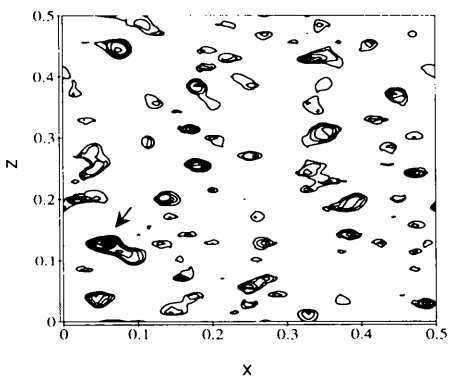

Fig. 9. Translation search using the orientation corresponding to the highest peak of the conventional rotation function for data set 3*r*9 (Table 3). The highest peak of the translation search is marked by an arrow.

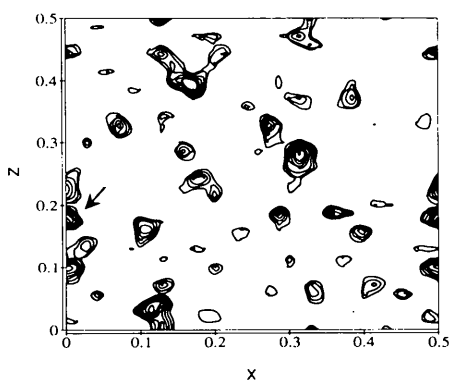

Fig. 10. Translation search using the correct orientation (Table 4) for data set 3*r*9*a*-det without prior PC refinement of the orientations and positions of the four Fab domains. The highest peak of the translation search is marked by an arrow.

Another question is whether PC refinement of the individual Fab domains, including the elbow angle, was necessary to solve the structure. Fig. 10 shows the translation search against data set 3*r*9*a*-det for molecule *A* in the correct orientation where only the overall orientation of the molecule was refined by PC refinement. The translation search clearly fails.

Packing considerations can be useful to identify the correct placement of the search model as was shown by Hendrickson & Ward (1976). The solvent content of the Fab 26-10 is about 60% in the unit cell of the crystal. Thus, packing considerations were of no help in solving the structure.

Considering the limited number of tests with conventional molecular replacement, we cannot exclude the possibility that the structure could have been solved by conventional rotation and translation functions without PC refinement by some particular choice of the parameters of the molecular replacement. But it is clear that it would have been a very difficult case. Thus, PC refinement of the orientations and positions of the Fab domains has made the solution of the Fab 26-10 structure an efficient and nearly straightforward process.

The author thanks Gregory A. Petsko and Roland K. Strong for the very fruitful collaborations to solve the Fab 26-10 structure. Support by the Pittsburgh Supercomputer Center is gratefully acknowledged (grant no. DMB890008P). The author thanks Anton Krukowski for producing Fig. 8.

References

BRÜNGER, A. T. (1990*a*). *Acta Cryst.* **A46**, 46-57.
BRÜNGER, A. T. (1990*b*). *X-PLOR Manual*, Version 2.1. Yale Univ., New Haven, USA.
BRÜNGER, A. T., CAMPBELL, R. L., CLORE, G. M., GRONENBORN, A. M., KARPLUS, M., PETSKO, G. A. & TEETER, M. M. (1987). *Science*, **235**, 1049-1053.
BRÜNGER, A. T., KRUKOWSKI, A. & ERICKSON, J. (1990). *Acta Cryst.* **A46**, 585-593.
BRÜNGER, A. T., KURIYAN, J. & KARPLUS, M. (1987). *Science*, **236**, 458-460.
BRÜNGER, A. T., LEAHY, D. J., HYNES, T. R. & FOX, R. O. (1991). *J. Mol. Biol.* Submitted.
CYGLER, M. & ANDERSON, W. F. (1988). *Acta Cryst.* **A44**, 38-45.
FITZGERALD, P. M. D. (1988). *J. Appl. Cryst.* **21**, 273-281.
FUJINAGA, M. & READ, R. J. (1987). *J. Appl. Cryst.* **20**, 517-521.
GARCIA, K. C., RONCO, P., VERROUST, P. J. & AMZEL, L. M. (1989). *J. Biol. Chem.* **264**, 20463-20466.
HAUPTMAN, H. (1982). *Acta Cryst.* **A38**, 289-294.
HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252-270.
HENDRICKSON, W. A. & WARD, K. B. (1976). *Acta Cryst.* **A32**, 778-780.
HOPPE, W. (1957). *Acta Cryst.* **10**, 750-151.
HUBER, R. (1965). *Acta Cryst.* **19**, 353-356.
HUBER, R. (1985). In *Molecular Replacement.* Proc. of the Daresbury Study Weekend, February 1985, pp. 58-61. SERC Daresbury Laboratory, Warrington, England.
HUDSON, N. W., MUDGETT-HUNTER, M., PANKA, D. J. & MARGOLIES, M. N. (1985). *J. Immunol.* **139**, 2715-2723.
JACK, A. & LEVITT, M. (1978). *Acta Cryst.* **A34**, 931-935.
LATTMAN, E. E. (1972). *Acta Cryst.* **B28**, 1065-1068.
MUDGETT-HUNTER, M., MARGOLIES, M. N., JU, A. & HABER, E. (1982). *J. Immunol.* **129**, 1165-1172.
NORDMAN, C. E. (1980). *Acta Cryst.* **A36**, 747-754.
POWELL, M. J. D. (1977). *Math. Programm.* **12**, 241-254.
RAO, S. N., JIH, J.-H. & HARTSUCK, J. A. (1980). *Acta Cryst.* **A36**, 878-884.
ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24-31.
SHERIFF, S., SILVERTON, E. W., PADLAN, E. A., COHEN, G. H., SMITH-GILL, S. J., FINZEL, B., DAVIES, D. R. (1987). *Proc. Natl Acad. Sci. USA*, **84**, 8075-8079.
STRONG, R. K. (1990). PhD thesis, Harvard Univ., Cambridge, USA.
YEATES, T. O. & RINI, J. M. (1990). *Acta Cryst.* **A46**, 352-359.

Acta Cryst. (1991). A**47**, 204-206

Random-Generation Model for Statistical Distribution of Point Groups

BY YOSHIAKI ITOH

The Institute of Statistical Mathematics, Minami-Azabu, Minato-ku, Tokyo, 106 *Japan*

AND TAKEO MATSUMOTO

Faculty of Science, Kanazawa University, Marunouchi, Kanazawa, 920 *Japan*

(*Received* 7 *July* 1990; *accepted* 23 *October* 1990)

Abstract

The point groups which have C_{2h} as a subgroup are frequently observed in crystal data. A random-generation model of point groups gives a possible reason for the statistical distribution among the point groups in the class of crystals containing elements and alloys. By the addition of possible symmetry operations to point group C_{2h}, new point groups of higher symmetry can be generated. The computer simulation of the random-generation model seems to explain the higher frequencies of occurrence of O_h and D_{6h}, the lower frequencies of C_{4h} and C_{6h} and the moderately high frequencies of the remaining five point groups which have C_{2h} as a subgroup.

The symmetry of three-dimensional periodic structure is described by the 230 space groups. The statistical distribution of the 230 space groups for crystals may aid the assignment of the symmetry of a crystal under study (Nowacki, Matsumoto & Edenharter, 1967; Bel'skii & Zorkii, 1971; Matsumoto & Nowacki, 1966; Matsumoto, 1988). The number of space groups that should be observed is estimated by a statistical study on ordered abundances (MacKay, 1967).

Statistical distributions on graphs or algebraic structures can be a convenient tool for various fields of science (Itoh, 1979). A random-generation model of groups gives a possible reason for the frequent observation of point groups in the structures of oxide and hydroxide crystals (Itoh, 1986). Here we give another random-generation model to explain the statistical distribution for the class of elements and alloys (Nowacki, Edenharter, & Matsumoto, 1967) whose structures are closely related to hexagonal closest packing or cubic closest packing. The symmetry of a crystal is determined by physical and chemical processes, which are composed of a large number of mutually interacting factors. These factors may give a reason for the use of the stochastic model for statistical distributions.

To carry out the statistical study we must make clear the concept of the species of a crystal. Although

0108-7673/91/030204-03$03.00