

# Direct Determination of Crystal Structure

## II. Sign Determination

Hidewo TAKAHASHI

(Received October 7, 1985)

Faculty of Education  
Kagoshima University  
Kagoshima 890, Japan

### Abstract

Sign determination by convergence of calculated structure factors to observed ones is examined. The coincidence between calculated and true signs becomes more than 80% in 3 sets with 316 reflections among 25 sets of 195 reflections.

### 1. Introduction

There are two ways to direct determination of crystal structure, direct determination of atomic coordinates by Monte Carlo method and phase determination. The author tried to program of direct determination of crystal structure by Monte Carlo method (Takahashi (1983)). Although the method is not perfected yet, the method is expected to become the method of direct determination of the crystal structure in near future. Monte Carlo method is economically expensive by the experience of the author.

Determination of phases of structure factors of crystals for X-ray diffraction intensities is the most important method in practice. At present, the two methods to solve the phase problem were proposed and have been used practically. These methods, although being over-simplification, depend essentially on the validity of probability theoretical interpretation of a phase determined by the two other phases (triple product phase relationship). However, we can not determine the two phases correctly, so that the validity of the third phase can not be assured even the probability is large. Symbolic addition method extends the triple product phase relationship self-consistently to the set of phases of all reflections. A doubt arises if we can determine all phases from comparatively small number of phases. Multiple solution method selects starting set of phases which consists of very few phases and determine all the phases by the triple product phase relationship. Hence, we can not avoid the above doubt even the multiple solution method.

The experience of the author shows about the Monte Carlo method that the crystal structure with R-value less than 0.3 determined by using 316 reflections is only one from 61 initial structures. When the number of reflections decreases, the number of structures with small R-value increases. The phases of structures with smaller R-value satisfy the triple product phase relationship. This seems to the author that the phases determined comparatively small number of triple product phase relationship are not reliable.

A structure factor can be approximated by the sum of products of the structure factors of reflections satisfying triple product phase relationship (Karle and Karle (1966)). If phases are given by using random number done by Furusaki (1979), we can calculate structure factors and may judge the validity of the phases by the difference between observed absolute values of structure factors and those of calculated structure factors. Then converging the calculated structure factors to observed ones, we may determine the phases of the structure factors. This idea was concretized for electronic computer program and calculated the signs of pyrenocine used for Monte Carlo method (Takahashi (1983)). The result was that 3 structures with coincided number of 80% with true phases were obtained among 25 initial sets.

## 2. Method

The outline of the program of the sign determination is as follows. The calculation of signs is repeated by the following 4 steps, and if the signs are no more changed by the calculation, the calculation is stopped.

### *Preparatory step*

$N_{ref}$  strong reflections are chosen from total reflections, and the set of the  $N_{ref}$  reflections is called  $H$ . There are several pairs of reflections  $k$  and  $h-k$  for a reflection  $h$  in  $H$ . The number of pairs and the pairs of reflections  $k$  and  $h-k$  for every reflection in  $H$  are decided at first. When the number of pairs is zero, the sign of the reflection can not be determined. Then the reflection is excluded from  $H$ . Signs are assigned to the reflections of  $H$  by using random numbers (Furusaki, (1979)). Structure factors,  $E_h(calc)$ , are calculated by using the following eq. (1) (Karle and Karle (1966)),

$$E_h(calc) = \sigma_2^{3/2} \sigma_3^{-1} \langle E'_k E'_{h-k} \rangle_k \quad (1)$$

where absolute value of  $E'_h$  is equal to  $E_h$ . The set of the signs of the calculated structure factors is called  $A$ . The degree of the difference between  $|E_h|$  and  $|E'_h|$  is estimated for every reflection by

$$DF_h = ||E_h| - |E'_h|| / |E_h|, \quad (2)$$

and the  $N_{ref}$  reflections are numbered according to the values of  $DF_h$ .

*1st step of a cycle of calculation*

This step is jumped to the 2nd step when calculation is at start. The coincidence between  $A$  and  $A'$  which is obtained in 4th step is examined. If the coincidence is better than the criterion previously set, the calculation is stopped and the result is printed out. If the coincidence is not sufficient,  $A'$  is put to be  $A$ .

*2nd step of a cycle of calculation*

The  $L_{ref}$  reflections with large  $DF_h$  are chosen and the set of signs of the  $L_{ref}$  reflections in  $A$  is called  $B$ . The set of signs necessary to determine the signs in  $B$  is called  $C$ . The signs of  $N_{ref}$  structure factors are calculated by using the signs belonging to the set  $A - (B \cup C)$  and the absolute values of structure factors. If  $L_{ref}$  is too large, it happens not to determine all signs, then  $L_{ref}$  is decreased by 1 and the calculation in this step is done over again from the start.

*3rd step of a cycle of calculation*

The  $N_{ref}$  signs and structure factors,  $E_h''$ s, are calculated by using the signs obtained in the 2nd step and the absolute values of the structure factors. The set of the signs in this step is called  $A'$ .

*4th step of a cycle of calculation*

The  $N_{ref}$  structure factors calculated in the 3rd step are scaled by using following relation,

$$\sum_h E_h = \sum_h E_h'' \quad (3)$$

The degree of the difference of structure factors  $E_h$  and  $E_h''$  is also calculated for every reflection and the  $N_{ref}$  reflections are numbered according to the values of  $DF_h$ . When this step of calculation is over, the calculation comes back to the 1st step.

Another program to examine coincidence between the calculated and true signs was made for pyrenocine. Since the space group of pyrenocine is  $Pbcm$ , there are 8 set of signs of  $E_h$ s for the 8 positions of possible origins. The correctness of the determined signs were estimated by the largest number of the coincidence between the obtained signs and the 8 sets of true signs. The degree of the coincidence ( $dc$ ) is represented by the number of the coincided signs divided by  $N_{ref}$ .

Before full calculation, the following problems were to be solved.

- (1) How to estimate the coincidence between  $A$  and  $A'$  in the 1st step?
- (2) How large the size of  $B$ ,  $L_{ref}$ , should be?
- (3) How to select small number of reliable sets from large sets of signs?

The first problem was easily solved. By putting  $cc$  to be the number of the coincidence of signs between  $A$  and  $A'$ , the program was set to stop calculation if  $cc \geq EPS$ . Among the results for  $EPS = 0.8, 0.9, 0.95$  and  $1.0$ , the best is for  $EPS = 1.0$ . Hence,  $EPS$  was set to be  $1.0$  in all the calculation.

To solve the 2nd problem, the calculations were made for  $L_{ref}=4, 6, 9, 19, 32,$  and  $97$  in the case of  $N_{ref}=195$  and 5 sets of  $A$ . The CPU times were 59.48, 52.50, 43.52, 48.08, 62.70 and 75.25 sec, respectively. The value of  $dc$  increases when  $L_{ref}$  was smaller than 19, but the differences were very small. The number of cycles and CPU time increases when  $L_{ref}$  become smaller. The increase in CPU time for larger  $L_{ref}$  means that the useless time for settle of the value of  $L_{ref}$  to determine all the signs of  $B$  increases at the first cycle of calculation, this time is comparatively short and the increase in total CPU time is not remarkable. Although the problem could not be solved, we can easily choose the value of  $L_{ref}$ , since the variable range in CPU time is comparatively narrow.

The 3rd problem is the most important. The author's method depends on the accuracy of structure factors calculated by using eq. (1) to the true structure factors. The accuracy depends on the number of reflections,  $N_{ref}$ . If  $N_{ref}$  becomes small, the accuracy becomes bad. Hence,  $N_{ref}$  should be set to be large as possible. But, if  $N_{ref}$  becomes large, CPU time increases remarkably. When  $N_{ref}=316$ , CPU time for the calculation of structure factors and R-value by using the true structure factors was 57.2 sec and R-value was 0.347. It is not possible to determine directly the signs for the large number of reflections. There is no other way for large  $N_{ref}$  than to extend signs determined for smaller  $N_{ref}'$  to the large  $N_{ref}$  signs by using eq. (1). Although smaller R-value does not correspond necessarily to larger degree of coincidence, but there is a tendency that the sets of signs with smaller R-values have large  $dc$ . Hence, the following way may be best. The signs of the large number of sets with small  $N_{ref}'$  are determined at first. The sets of signs giving smaller R-values are chosen among them. Then, the sets of signs with larger  $N_{ref}$  are determined from the obtained sets of signs. Thus, the signs of the small number of sets with smaller R-values and large  $N_{ref}$  are determined. At final stage, we should choose reasonable set of signs by Fourier transform of the structure factors with determined signs.

### 3. Results and discussions

The number of starting sets with 195 reflections was 25. The value of  $L_{ref}$  is 6 at first. There were

Table I

No.	R-value	Number of signs coincided	$dc$
3	0.57831	283	0.90
3*	0.41158	279	0.88
7	0.43696	165	0.52
9	0.41297	165	0.52
13	0.41382	188	0.59
14	0.40847	264	0.84
18	0.46202	169	0.53
20	0.38760	263	0.83
21	0.49627	173	0.55

8 sets with R-value less than 0.6 after calculation. The 8 sets of signs of  $N_{ref}=316$  were calculated 10 cycles with  $L_{ref}=15$ . The results are listed in Table I. The set of No. 3 was re-calculated by using  $L_{ref}=6$ . The cause of irregularity in R-value in No. 3 is not clear. It may be remarkable that there are 3 sets with  $dc$  more than 0.8.

This method is an antithesis against symbolic addition method and multiple solution method in the following two meanings.

- (1) Total phases should be determined from not very few triple product phase relationship but many (more than  $3N$ , where  $N$  is the number of atoms) relationships.
- (2) So-called probability of phase is not the absolute probability but variable if set of signs is changed.

Many structures with very small R-value obtained from very few reflections by Monte Carlo method were not extended to have small R-value in case of large number of reflections. The signs of these structures satisfy triple product phase relationship. There are many not correct structures satisfying triple product phase relationship for small number of reflections.

Even a phase is calculated by triple product phase relationship and the probability of the phase is also calculated, the probability does not represent the true probability of the phase. The probability means that if the other two phases are so then the probability of the phase should be so. But, there is no guarantee that the other phases are so. The author programmed the sign determination to be led by the probability at first, but failed to obtain reasonable set of signs. This was the reason why the author programmed to make converge the calculated structure factors to observed one.

The author tried to extend the program to determine phases of complex structure factors. The trial does not succeed yet. Structure factor convergence is very good criterion for real structure factors, but the condition seems insufficient for complex structure factors.

#### References

- Furusaki, A. (1979). Acta. Cryst. **A35**, 220.  
Germain, G. and Woolfson, M.M. (1968). Acta Cryst. **B24**, 91.  
Germain, G. Main, P. and Woolfson, M.M. (1970). Acta Cryst. **B26**, 274.  
Karle, J. and Karle, L.L. (1966). Acta Cryst. **21**, 849.  
Takahashi, H. (1983). Bull. Kagoshima Univ. (Nat. Sci.). **35**, 7.