THE SIGNIFICANCE OF THE PHASE SYNTHESIS

By R. SRINIVASAN

(Department of Physics, University of Madras, Madras 25, India)

ABSTRACT

This paper gives an account of the results of the tests that have been carried out to assess the relative importance of the phase angles and structure amplitudes in a Fourier synthesis. For this, the phase factors, $e^{i\alpha}$, have been combined in different ways with structure amplitudes such as when the latter are (a) replaced by the r.m.s. of the structure amplitudes (i.e., by $\sqrt{\sum f_j^2}$), (b) randomly chosen from the set of correct structure amplitudes, (c) made us much anti-correlated (about the mean) with the correct ones as possible and (d) taken to be those of an entirely different structure. The above syntheses which are only modifications of the phase synthesis, $e^{i\alpha}$, have been tested on both the centro-symmetric and non-centrosymmetric projections of a known structure. They strongly establish the greater importance of the phase angles in a Fourier synthesis.

1. INTRODUCTION

IN a series of papers from this laboratory (Ramachandran and Raman, 1959; Raman, 1959; Raman, 1960; Srinivasan, 1960 a; Srinivasan and Aravindakshan, 1960; hereafter referred to as Parts I, II, III, IV and V respectively) the problem of deconvoluting the Patterson function when a part of the structure is known, was considered in detail. In the course of these studies it became necessary to analyse the significance of the standard syntheses which employ for their Fourier coefficients various standard functions of F, such as $|F|$, $e^{i\alpha}$, $1/F$, etc. It was pointed out in Part I and later proved in a more rigorous way in Parts II and IV that the phase synthesis $e^{i\alpha}$ resembles closely the Fourier synthesis while the modulus synthesis, $|F|$, resembles the Patterson. Though this indicated in a general way the importance of the phase angles, a closer study of the relative importance of the phase angles and the structure amplitudes was thought worthwhile. The results of such a study are presented in this paper.

252

2. DESCRIPTION OF THE VARIOUS SYNTHESES

The calculations were all made on the centrosymmetric and non-centro-symmetric projections of L-tyrosine hydrochloride whose structure has been thoroughly analysed in this laboratory (Srinivasan, 1959). The crystal data of this compound are: $a = 11\cdot1$, $b = 9\cdot1$, $c = 5\cdot1\text{A}$, $\beta = 92°$, two molecules of $C_9H_{12}NO_3\cdot HCl$ per cell, space group $P2_1$. In all the syntheses to be described below the phase angles referred to (or the signs in the centro-symmetric case) correspond to those calculated for the above structure (Srinivasan, 1959) unless otherwise stated.

(a) *The Normalised Phase Synthesis.* To start with, a modified form of the phase synthesis, $e^{i\alpha}$, was devised. The modification consisted in not using unit amplitudes for all reflections as in $e^{i\alpha}$, but a value equal to $\sqrt{\Sigma f_j^2}$ which is the mean for all reflections having the same value of $(\sin\theta/\lambda)$. Thus, in effect, the magnitudes of the geometrical structure factors of all reflections were made equal, but all of them were multiplied by a mean atomic scattering factor. We shall call this synthesis the "normalised phase synthesis", the normalisation being done such that $\Sigma|F|^2$ is the same for this synthesis and the Fourier synthesis. The syntheses for the non-centro-symmetric and the centrosymmetric projections are shown in Figs. 1 (b) and 2 (b) respectively along with the actual Fourier syntheses in Figs. 1 (a) and 2 (a).

In carrying out the calculations the value of $\sqrt{\sum_j f_j^2}$ for each reflection was obtained from a graph. The temperature factor was also applied to the amplitudes and it was the same as that employed for the actual crystal. It will be noticed that the agreement between the normalised syntheses [Figs. 1 (b) and 2 (b)] and the actual Fourier syntheses [Figs. 1 (a) and 2 (a)] is extremely good.

(b) *The Random Phase Synthesis.* The interesting question now arises as to what would happen if random values of $|F|$ are used along with the correct phase angles, but not equal values (apart from the 'f' factor variation) as in the above synthesis. Such a synthesis which may be called the "random synthesis" was calculated. However, in order to ensure the same statistical distribution of structure amplitudes as in the correct structure, the reflections were grouped into various ranges of $(\sin\theta/\lambda)$ and in each range the values of the actually observed $|F_o|$'s (Srinivasan, 1959) were randomly permuted amongst themselves and these were then used with the correct phase angles to calculate the synthesis. The syntheses for the non-centrosymmetric and the centrosymmetric cases are given in Figs. 1 (c) and 2 (c) respectively,

Surprisingly, even in these random syntheses, the structure is revealed in fairly good detail.
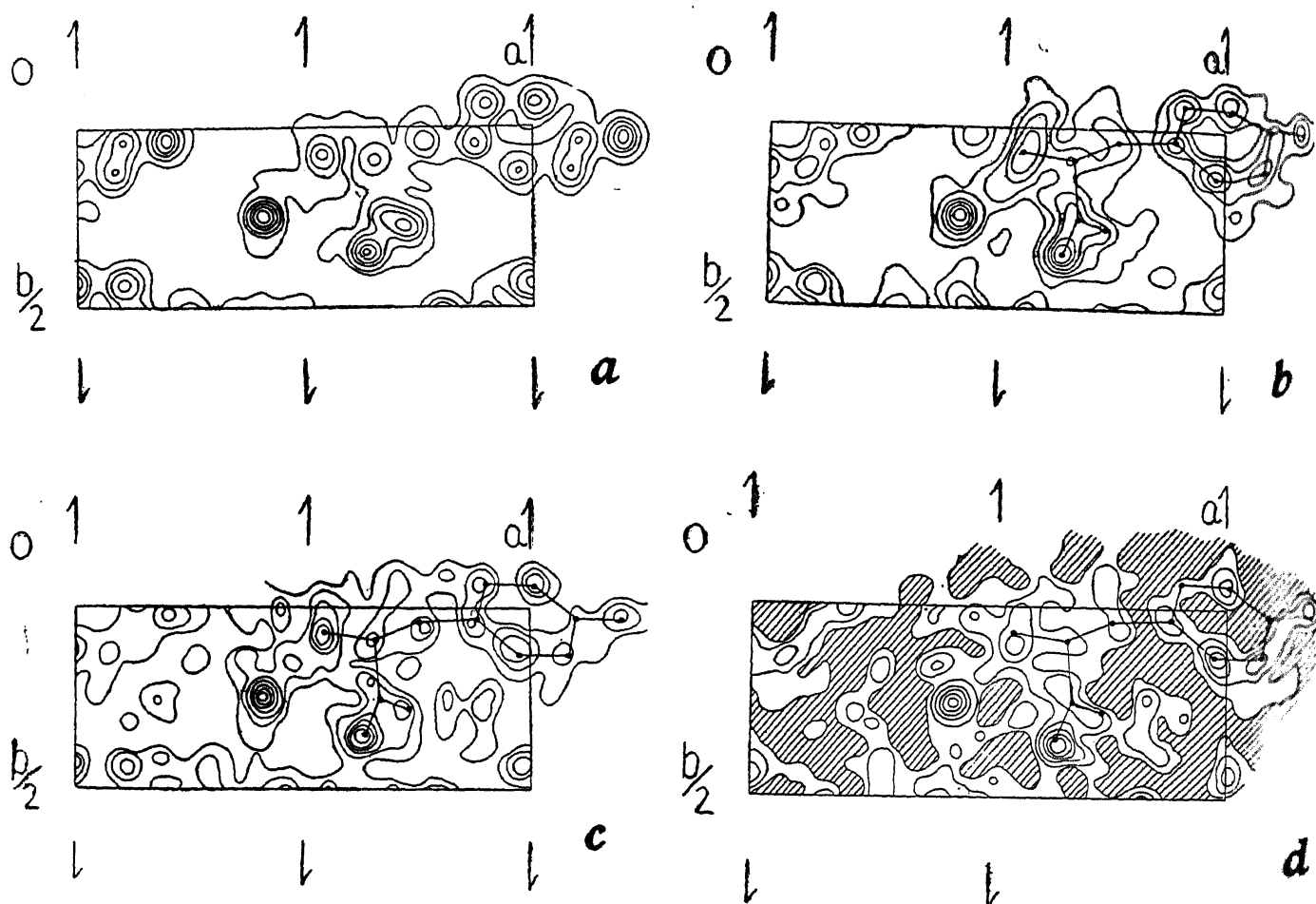

FIG. 1*.   (a) The Fourier synthesis, (b) the normalised phase synthesis, (c) the random phase synthesis and (d) the anti-correlated phase synthesis of L-tyrosine hydrochloride projected down the c-axis.   Contours are at intervals of $2e/A^2$ at chlorine and $1e/A^2$ elsewhere, starting from $2e/A^2$.

* The different kinds of atoms in the molecule of tyrosine are not marked in all the figures but only in Figs. 3 (b) and 4 (b) where small dots correspond to carbons big dots to oxygens and open circle to nitrogen.

(c) *The Anti-correlated Phase Synthesis*.—Now, it was thought worthwhile to go a step further and take the strong reflections weak and *vice versa*. This would be equivalent to making the new set of amplitudes and the correct set as much anti-correlated about their mean as possible.*   The method adopted for getting the *maximum* anti-correlation was as follows: In each range of (sin θ/λ) the various reflections were arranged in decreasing order of magnitude.   The whole sequence was now inverted so that the

* For a proper explanation of the various statistical terms used in the discussion *see* Appendix I.

strongest reflection would become the weakest and so on. However, it will not be possible to realise perfect anti-correlation about the mean by this process. The possibility of achieving this depends on the nature of the probability function and it can be shown that, unless the function is symmetrical about the mean, perfect anti-correlation is not possible. In our present case, the probability functions for the centrosymmetric and non-centrosymmetric cases are not symmetrical about the mean (Ramachandran and Srinivasan, 1960) and hence perfect anti-correlation cannot be realised for them. However, the maximum possible anti-correlation for any type of distribution can be worked out and they are given in Appendix I for the centrosymmetric and non-centrosymmetric cases. The theoretical maximum of $- 0·89$ for the non-centrosymmetric case agrees fairly well with the value of $- 0·75$ as calculated from the actual data used for the anti-correlated synthesis. The difference between the two is to be expected since the procedure adopted is subject to the same types of errors as are encountered in the calculation of a Wilson plot. The synthesis was calculated only for the non-centrosymmetric projection and is shown in Fig. 1 (*d*). It is very surprising indeed to find that even this diagram shows the structure in good detail.

3. Explanation of the Syntheses

(*a*) *From the Fourier Point of View.*—The normalised phase synthesis confirms the most essential property of the phase synthesis deduced theo-retically in Parts II and IV, namely, that it resembles closely the Fourier synthe-sis. The consequence of taking $\sqrt{\Sigma f_j^2}$ for the coefficients instead of unit amplitudes as in $e^{i\alpha}$ will be that peaks will not be sharp but will have a finite size and shape corresponding to an average atom in the structure. The normalisation also makes it possible to compare directly this synthesis with the actual Fourier synthesis. It is interesting to see that *even the relative strengths of carbon, oxygen and chlorine have come out correctly*, even though the amplitudes fed in correspond to those of a single average atom in the unit cell.

It will be noticed that the random synthesis is also almost as good as the normalised phase synthesis though it contains some spurious peaks as compared with the practically uniform background in the latter. That the random synthesis would reveal the structure in such good detail could not have been anticipated to start with. In fact, it was this rather unexpected result that suggested the next test that was made, namely, to use a set of amplitudes completely anti-correlated with the correct ones. The resultant synthesis, Fig. 1 (*d*), is of course worse than the random synthesis, but still

the tyrosine molecule is clearly discernible. What is more interesting is that the relative strengths of the peaks at the positions of chlorine, carbon, etc., are practically unaffected. However, the number of spurious peaks is now somewhat larger as compared with the random synthesis and their strengths are also larger.

The real reason why even under these diverse conditions the phase synthesis is able to reveal the structure can be understood by making a closer study of the problem. The explanation follows essentially from the general idea of the deconvolution of the Patterson function and of the principle of modulation of one structure by another (Parts I and II). All the above syntheses can be interpreted as the modulation of the appropriate modulus synthesis with the phase synthesis, $e^{i\alpha}$. The latter is practically same as the Fourier synthesis and contains first order peaks of strengths $0 \cdot 90 f_i/S_N$ at the same positions as in Fourier where $S_N = \sqrt{\sum_j f_j^2}$ (Part IV). Now,

because of the fact that the structure amplitudes used in the case of random and the anti-correlated syntheses do not correspond to any actual "structure", their modulus synthesis would not contain only positive peaks as in a modulus synthesis of any real structure (Part IV). However, at the origin alone both
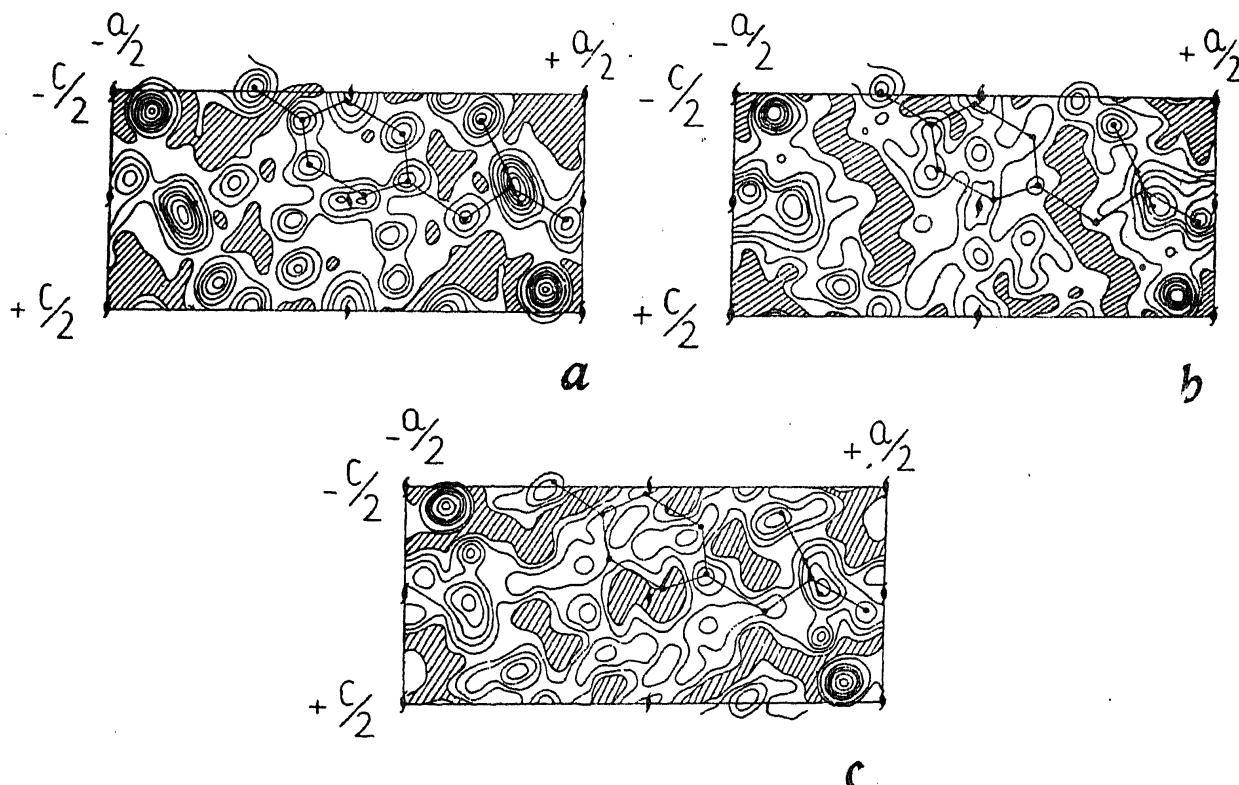

FIG. 2. (a) The fourier synthesis, (b) the normalised phase synthesis and (c) the random phase synthesis of L-tyrosine hydrochloride projected down the b-axis. Contours are at intervals of $2e/A^2$ at chlorine and $1e/A^2$ elsewhere, starting from $3e/A^2$. Shaded area corresponds to less than $3e/A^2$.

these will have a strength equal to that of the actual modulus synthesis. Elsewhere, one would expect a fluctuating background. The only strong interactions which arise in the modulation of this synthesis with the phase synthesis will therefore correspond to those of an origin peak with the first order peaks of the phase synthesis, and will thus lead to the structure contained in the phase synthesis, i.e., at the correct atomic positions.

Another general feature that might be observed in Figs. 1 and 2 is that the phase synthesis for the non-centrosymmetric case is distinctly better than that for the centrosymmetric case. This result is best understood after we consider another interesting aspect of the problem, namely, from the statistical point of view.

(b) From the Statistical Point of View. The terms random and anti-correlated have been used in the previous sections in relation to the mean of the distribution. These would have conveyed the idea that the distributions chosen are not correlated or anti-correlated to the distribution of the structure amplitudes for the correct structure. However, it should be remembered that in a Fourier synthesis it is the actual structure amplitudes that is employed so that the correlation about the mean (*k*) is not so important as the direct correlation (*c*). The terms correlation about the mean and the direct correlation are explained in Appendix.

If Table I of Appendix is examined it will be seen that the direct correlation between the structure amplitudes of the assumed distribution and the actual distribution is as high as 0.79 for the non-centrosymmetric case and 0.63 for the centrosymmetric case when these are randomly correlated about the mean. Consequently, it is not at all surprising that the random synthesis reveals much of the details of the actual structure. Even when the two sets of structure amplitudes are anti-correlated (or as much anti-correlated as possible) the direct correlation coefficient is still as high as 0.60 for the non-centrosymmetric case. In this case, even the anti-correlated synthesis does reveal the structure quite well and this becomes understandable from the magnitude of the direct correlation coefficient. The correlation is poorer in the centrosymmetric case (about 0.40) but it is believed it is good enough to reveal some details of the structure.

The large value of the direct correlation arises essentially because the quantity *F* is always positive. Consequently, even if there are variations about the mean there is always a positive correlation between the absolute magnitudes in the two sets of the data because both of them must necessarily be positive.

It will be noticed from Table I that the direct correlation for the non-centrosymmetric case is systematically larger than that for the centro-symmetric case for all the different types of syntheses. This explains the fact already mentioned, namely that the phase synthesis for the non-centrosymmetric case is distinctly better than that for the centrosymmetric case, as a general feature.

4. TESTS ON A LIGHT-ATOM STRUCTURE

(a) *The Normalised Phase Synthesis.* It will be observed that all the above syntheses have been calculated for the compound L-tyrosine hydrochloride which is not a purely light-atom structure. Hence the theoretical considerations given above are not strictly applicable to this case. The effect of the presence of heavy atom on the distribution function $P(z)$ has already been worked out (Srinivasan, 1960 b). In the present case, in which there are two heavy atoms in the unit cell, the general tendency is to alter the distribution curve for both centrosymmetric and non-centrosymmetric cases such that there are a large number of reflections having their value near the mean. The effect of this in both cases, is to make the value of ϵ larger (*see* Appendix I) than if the heavy atoms were not present. In view of this fact, it was felt that the test should be carried out on a pure light-atom structure. This was actually done and the normalised phase synthesis for both the non-centrosymmetric and centrosymmetric cases were calculated for the structure " tyrosine " alone obtained by removing the two chlorine atoms from the tyrosine hydrochloride structure. The resultant syntheses [Figs. 3 (a) and 3 (b)] will be found to be practically as good as Figs. 1 (b) and 2 (b). It was therefore felt not necessary to calculate the other syntheses.
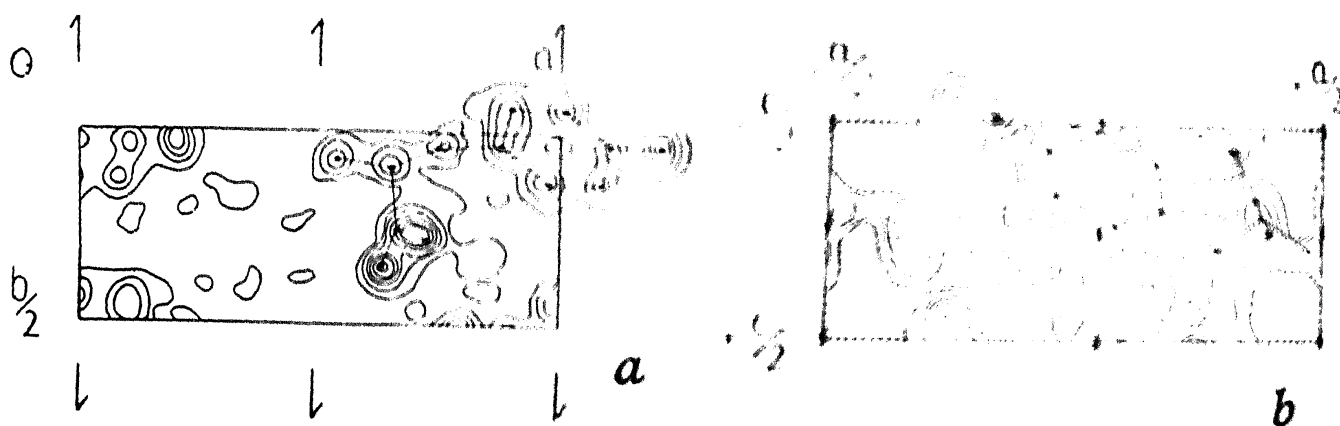

FIG. 3. The normalised phase synthesis of L-tyrosine projected down (a) the c-axis and (b) the b-axis. Contours are same as in Figs. 1 and 2 respectively.

(b) *Use of the Structure Amplitudes of an Entirely Different Structure.*— The results of carrying out the different types of syntheses described earlier

suggested one other interesting possibility, namely, that of combining the structure amplitudes of an actual structure with the phase angles calculated from an entirely different structure. For this purpose, six atoms were assumed at the positions marked by crosses in Fig. 4 (a) and these atoms were assumed to have a suitable scattering factor (actually 1·7 times the 'f' value of carbon) falling off with θ as for a carbon atom. This was done to make the value of $\sum_i f_j^2$ of this structure (which we may call as the correct structure A, and denote its structure amplitudes by $|F_A|$) the same as that for the assumed structure. Now, the assumed structure (B) was taken to be the structure of pure tyrosine. The phase angles α_B for tyrosine were calculated and were used along with the structure amplitudes $|F_A|$ to calculate a synthesis using $|F_A|$ exp. $i\alpha_B$ as Fourier coefficients. The synthesis for the non-centrosymmetric projection is shown in Fig. 4 (a). Figure 4 (b) shows the difference Fourier synthesis $(|F_A| - |F_B|)$ exp. $i\alpha_B$.
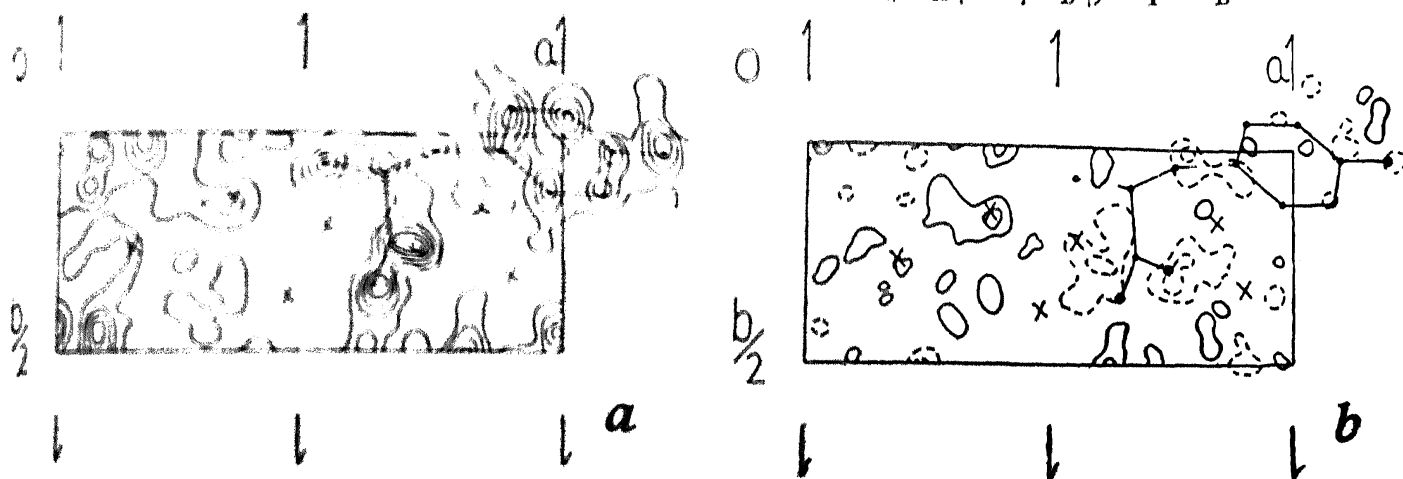

Fig. 4. (a) The synthesis using $|F_A|$ exp. α_B as fourier coefficients where A corresponde to an arbitrary six-atom structure (marked by crosses) and B to tyrosine. Contours are sams as in Fig. 1. (b) The difference synthesis, $(|F_A| - |F_B|)$ exp. $i\alpha_B$. Contours are at intervals of 1c A². Positive and negative contours are shown by continuous and broken lines respectively. The zero contour is omitted.

5. DISCUSSION

The synthesis using the structure amplitudes of an arbitrary structure along with phases calculated from a different structure represents probably the extreme possible case in the present series of tests. The agreement between this synthesis [Fig. 4 (a)] and the Fourier [Fig. 1 (a) with the chlorine removed] is very good and it will be seen that all the atoms of tyrosine have come out in it. Particularly it is noteworthy that, excepting at one place, there are practically no strong peaks at the atomic positions of the arbitrary structure A [marked by crosses in Fig. 4 (a)], whose structure amplitudes were used in the synthesis. The one peak mentioned above is, however, **believed to be spurious.**

From this, the following important conclusion is obvious, namely, that if a Fourier synthesis is made using the amplitudes of the correct structure (A) along with the phases calculated for an assumed wrong structure (B) then the resultant synthesis would contain peaks only at the assumed positions and not at the correct positions.

This seems to be in apparent contradiction with the well-known fact in structure analysis, namely, that the correct structure is usually arrived at by a series of successive Fourier syntheses, assuming initially atoms to be present at all the prominent peaks in a trial Fourier synthesis. The resolution of this paradox follows from the general idea of modulation (Part I).

Now, the modulus synthesis, $|F_A|$, contains an origin peak of strength $0 \cdot 92 \, S_N$ and non-origin peaks of strengths $0 \cdot 38 \, f_i f_j / S_N$ at $r_{Ai} - r_{Aj}$, while the phase synthesis, $e^{i\alpha_B}$, contains peaks of strengths $0 \cdot 90 \, f_i / S_N$ at r_{Bi}. The modulation of the two should obviously lead to peaks at r_{Bi} arising out of the interactions between the origin peak of the former and the first order peaks of $e^{i\alpha_B}$ at r_{Bi}. The other interactions between $r_{Ai} - r_{Aj}$ and r_{Bi} would, in general, lead only to a background and in particular there will not be any concentration of electron density at r_{Ai}. If, on the other hand, the structures A and B have some atoms in common, say P of them, then P of the peaks in r_{Bi} would be of the type r_{Aj} so that the interactions between $r_{Ai} - r_{Aj}$ and r_{Bi} will lead to concentration at r_{Ai}. The strengths of A will, of course, depend on the number of coincident atoms (P) and will, in general, be weaker than those at B.

It follows from the above that *the necessary condition for peaks to occur at the correct atomic positions, A, is that at least some of the atoms used for calculating the phases must coincide with those in the structure A.*

This again shows that the usual method of successive Fourier syntheses would successfully converge to the correct structure only if the assumed structure or at least part of it is near the correct one. The method would obviously fail if completely wrong positions are used, to start with. Of course, there is no reason why the phases fed in from a completely arbitrary structure should lead to the correct one since, if it were to be true, it would be an absurdly simple solution to the phase problem.

However, the fact, that the assumed structure is wrong, would easily be revealed by the R value. For a completely wrong structure it would be 83% for the centrosymmetric and 59% for the non-centrosymmetric case respectively (Wilson, 1950). Further, the difference Fourier synthesis might

also help in detecting the falsity of the structure. Even this is not so very helpful as is usually supposed. Though there is a small concentration of negative density in Fig. 4 (*b*) at some of the tyrosine atoms, both positive and negative peaks occur at random in the synthesis the maximum values corresponding approximately only to a helium atom. Probably the error synthesis would have been more useful in detecting the wrong atomic positions.

REFERENCES

Ramachandran, G. N. and Raman, S.   *Acta Cryst.*, 1959, **12**, 957.

———— and Srinivasan, R.   ..   *Ibid.*, 1959, **12**, 410.

Raman, S.   ..   *Ibid.*, 1959, **12**, 964.

————   ..   *Ibid.*, 1960, in press.

Srinivasan, R.   ..   *Proc. Ind. Acad. Sci.*, 1959, **50 A**, 19.

————   ..   Sent for publication. 1960 *a*.

————   ..   *Acta Cryst.*, 1960 *b*, **13**, 388.

———— and Aravindakshan, C.   ..   *Ibid.*, 1960, sent for publication.

APPENDIX I

Correlation between Structure Amplitudes

We define the correlation between the structure factors F_1 and F_2 of two structures 1 and 2 as

$$c(F) = \frac{1}{2} \frac{\Sigma(F_1 F_2^* + F_2 F_1^*)}{\sqrt{\Sigma|F_1|^2} \cdot \sqrt{\Sigma|F_2|^2}} = \frac{\Sigma|F_1||F_2|\cos(\alpha_1 - \alpha_2)}{\sqrt{\Sigma|F_1|^2} \cdot \sqrt{\Sigma|F_2|^2}} \quad (1)$$

where the notation is self-explanatory. We would be mainly interested in the case when the two phase angles α_1 and α_2 are the same for any reflection so that the cosine factor in (1) becomes unity giving the following quantity which would also be called the correlation between the structure amplitudes $|F_1|$ and $|F_2|$. Thus

$$c(|F|) = \frac{\Sigma|F_1||F_2|}{\sqrt{\Sigma|F_1|^2} \sqrt{\Sigma|F_2|^2}}. \quad (2)$$

The term coefficient of linear correlation (k) between two statistical variates is usually used to denote the correlation about the mean. Applying this to our case, we may write k as

$$k = \frac{\Sigma(|F_1| - \overline{|F_1|})(|F_2| - \overline{|F_2|})}{\sqrt{\Sigma|F_1|^2} \sqrt{\Sigma|F_2|^2}}. \quad (3)$$

We shall hereafter use the term "direct correlation" to denote $c(|F|)$ as defined by (2) and the term "correlation coefficient" or simply "correlation" to denote k as defined by (3). (If it desired to be explicit we may use "correlation about the mean" to denote k.)

It is possible to express c in terms of k by expanding (3) and substituting from (2). This gives

$$c = \frac{\Sigma m_1(|F_1|) m_1(|F_2|) - k\sigma(|F_1|)\sigma(|F_2|)}{\sqrt{m_2(|F_1|)} \sqrt{m_2(|F_2|)}} \quad (4)$$

where σ is the standard deviation and m_1 and m_2 are the first and second moments of the variates.

In the discussion to follow we would be mainly interested in the case when the two variated have the same probability distributions so that we may take $m_1(|F_1|) = m_1(|F_2|) = m_1$ and $\sigma(|F_1|) = \sigma(|F_2|) = \sigma$, etc. Equation (4) then reduces to

$$c = k + \frac{m_1^2}{m_2}(1 - k). \quad (5)$$

When the two variates are one and the same, both k and c should be equal to unity which also follows from (5).

The values of m_1 and m_2 for the normalised structure amplitudes are well known in the case of both non-centrosymmetric and centrosymmetric distributions. These are $m_1 = (\sqrt{\pi}/2)$, $m_2 = 1$ for the former and $m_1 = \sqrt{(2/\pi)}$, $m_2 = 1$ for the latter. Substituting these values in (5) the values of c are obtained to be as in Table I. The values in brackets (where given) correspond to the actual values found in the calculation of the corresponding syntheses described in the text.

It will be seen from Table I that even for perfect anti-correlation* the value of c is quite appreciable and particularly for the non-centrosymmetric case it is quite high. Also the value of c for the non-centrosymmetric case is systematically higher than the corresponding value for the centrosymmetric case. This essentially arises because of the fact that the probability distribution function in the former case has a large number of reflections with their values clustered round the mean while in the centrosymmetric case they have a comparatively wider dispersion. Thus if we could have an ideal case of a distribution which is a delta function at the mean, the value of c would always be unity. Table I also tells us that the worst possible direct

TABLE I

Values of the direct correlation c for different values of k
Actually observed values are given in brackets

Distri-bution k	1·0	0·5	0·0	−0·5	−0·80*	−0·89*	−1·0
Non-centro-symmetric ..	1·0	0·89	0·79 (0·74)	0·68	..	0·59 (0·55)	0·57
Centrosymmetric	1·0	0·82	0·63 (0.69)	0·45	0·34	..	0·26

* These values correspond to the maximum possible values of the anti-correlation coefficient for the centrosymmetric and non-centrosymmetric cases respectively.

* The terms "correlated" and "anti-correlated" (about the mean) are used in general, where k is positive and negative respectively, while "perfect correlation" and "perfect anti-correlation" are used to denote the extreme cases of $k = +1$ and -1 respectively.

correlation is for a centrosymmetric structure when the structure amplitudes are perfectly anti-correlated.

However, the above discussion on the relationship between c and k is purely from a theoretical point of view and it remains a separate issue whether the case of perfect anti-correlation is at all realisable in practice for any type of distribution. In fact, it may be stated here that, unless the distribution is symmetrical about the mean, perfect anti-correlation about the mean is not possible. (The details will be presented elsewhere.) This is also equivalent to the condition that the mean and the median of the distribution should be the same. Further, it is possible to calculate the maximum value of anti-correlation attainable in practice for any type of distribution. The values for the non-centrosymmetric and the centrosymmetric distributions are mentioned in Table I. It will be noticed that this maximum value is larger for the non-centrosymmetric than the centrosymmetric case. This is understandable since the distribution for the former case is fairly symmetrical about the mean while in the latter case it is rather asymmetrical.