

Critere de fin de l'Affinement d'une Structure Atomique

PAR S. CATICHA-ELLIS

Instituto de Física. Facultad de Ingeniería, Montevideo, Uruguay

ET A. RIMSKY

Laboratoire de Minéralogie et de Cristallographie, Faculté des Sciences, Paris, France

(Reçu le 2 Octobre 1957)

A criterion for the termination of the process of refinement of an atomic structure is given.

It is assumed that for any structure in a definite state there exists a set of ideal values of structure factors, i.e. without errors, which are not, of course, accessible to measurement. Both the calculated and observed sets of structure factors should be two equally good approximations to the ideal set. This condition provides a criterion for the termination of the process of refinement. In fact it is simpler to compare the discrepancy between the calculated and the ideal set with the experimental error. The first one can be expressed in a simple way in terms of the linear correlation factor r .

Independently of the main argument it is shown that the accidental errors are unimportant provided that the number of data collected is high. In this case the limit to the process of refinement is given by the systematic errors, for whose estimation a method is proposed.

An outcome of our reasoning is an argument favouring the adoption of the linear correlation factor r for the evaluation of a crystal structure, since it is simply related to the error in the structure.

Introduction

Au cours du processus de l'affinement d'une structure atomique cristalline, on cherche à obtenir des facteurs de structure calculés aussi voisins que possible des facteurs de structure observés. A notre connaissance aucun critère n'a été donné pour fixer la limite de l'affinement. Cette limite doit être donnée par l'erreur sur l'observation. Autrement dit, il est illusoire de déterminer une structure telle que l'erreur sur l'ensemble des facteurs de structure calculés soit plus petite que l'erreur sur l'ensemble observé.

En effet, l'observation peut être entachée d'erreurs systématiques (absorption, extinction, etc.) et d'erreurs accidentelles ou fortuites. Les erreurs systématiques pour une même mesure effectuée plusieurs fois de suite et dans des conditions définies ont toujours lieu dans le même sens. Par contre les erreurs fortuites se groupent dans les deux sens autour de zero (Borel & Deltheil, 1940).

Nous avons essayé de déterminer la part respective de ces deux types d'erreurs et l'ordre de grandeur de celles ci. Il nous a été possible alors d'établir une relation caractéristique de la valeur de l'affinement effectuée. Le critère d'arrêt est donné par la comparaison entre cette valeur et l'erreur expérimentale constatée.

I. — Soit F_{j_o} un facteur de structure observé et F_{j_c} la valeur calculée correspondante. Nous définissons un facteur de structure idéal F_{j_i} dont la valeur n'est entachée d'aucune erreur. Il est bien entendu que ce facteur de structure idéal est inaccessible.

Exprimons la variance résiduelle $\sigma_r(F_o^i)$ de l'ensemble des n valeurs observées par rapport à l'ensemble idéal des facteurs de structure (Morice & Chartier, 1954). On a :

$$\sigma_r^2(F_o^i) = \frac{1}{n-2} \sum_j (F_{j_o} - F_{j_i})^2. \quad (1)$$

De la même façon, la variance résiduelle de l'ensemble calculé par rapport à la valeur idéale $\sigma_r(F_c^i)$ sera :

$$\sigma_r^2(F_c^i) = \frac{1}{n-2} \sum_j (F_{j_c} - F_{j_i})^2. \quad (2)$$

Ainsi que la variance résiduelle de l'ensemble observé par rapport à l'ensemble de valeurs calculées est :

$$\sigma_r^2(F_o^c) = \frac{1}{n-2} \sum_j (F_{j_o} - F_{j_c})^2. \quad (3)$$

Ceci étant défini, nous pouvons écrire pour un facteur de structure de rang j que :

$$F_{j_i} = F_{j_o} + \Delta F_{j_o}^i, \quad (4)$$

$$F_{j_i} = F_{j_c} + \Delta F_{j_c}^i, \quad (5)$$

$\Delta F_{j_o}^i$ et $\Delta F_{j_c}^i$ étant les erreurs absolues respectives de l'observation et du calcul par rapport à la valeur idéale F_{j_i} du facteur de structure.

Les relations (4) et (5) nous permettent d'écrire :

$$F_{j_o} - F_{j_c} = \Delta F_{j_c}^i - \Delta F_{j_o}^i. \quad (6)$$

Pour l'ensemble de tous les facteurs de structure,

et en appliquant le théorème de l'addition des variances, il vient:

$$\sigma_r^2(F_o^c) = \sigma_r^2(F_o^i) + \sigma_r^2(F_c^i). \quad (7)$$

Tout ce que vient d'être dit implique que l'observation et le calcul sont à la même échelle (absolue), c'est à dire:

$$\begin{aligned} \Sigma F_i^2 &= \Sigma F_o^2 \\ &= \Sigma F_c^2 \cong \Sigma(F_o \cdot F_c) \cong \Sigma(F_o \cdot F_i) \cong \Sigma(F_c \cdot F_i). \end{aligned}$$

Nous admettons que la structure est terminée lorsque les deux ensembles observé et calculé sont deux approximations également bonnes par rapport à l'ensemble des valeurs idéales des facteurs de structure. C'est à dire que les deux écarts considérés sont du même ordre de grandeur. Soit

$$\sigma_r(F_o^i) = \sigma_r(F_c^i), \quad (8)$$

ou bien

$$\sigma_r^2(F_o^c) = 2\sigma_r^2(F_c^i). \quad (9)$$

Designons par ε_j l'erreur relative d'un facteur de structure calculé F_{jc} . On a:

$$F_{ji} - F_{jc} = \Delta F_{jc}^i = \varepsilon_j \cdot F_{jc}. \quad (10)$$

On définit l'erreur moyenne pondérée ε_c du calcul par

$$\varepsilon_c^2 = \Sigma_j (\varepsilon_j F_{jc})^2 / \Sigma F_{jc}^2. \quad (11)$$

Il vient

$$\varepsilon_c^2 = \sigma_r^2(F_c^i) \left/ \frac{1}{n-2} \Sigma (F_{jc})^2 \right. . \quad (12)$$

Mais pour un nombre suffisamment grand de termes de l'ensemble calculé nous pouvons faire l'approximation suivante:

$$\varepsilon_c^2 \cong \sigma_r^2(F_c^i) / *F_c^2, \quad (13)$$

* F_c étant la valeur quadratique moyenne de l'ensemble des facteurs de structure calculés.

Supposons la structure terminée, la relation (9) se trouve alors vérifiée. Nous pouvons écrire pour l'erreur ultime ε_U :

$$\varepsilon_U = \frac{1}{\sqrt{2}} \cdot \frac{\sigma_r(F_o^c)}{*F_c}. \quad (14)$$

Cette valeur doit être égale à l'erreur moyenne ε_o sur l'observation.

Bien entendu, l'expression (14) nous donne l'erreur sur l'ensemble calculé dans le dernier affinement, mais elle ne sera pas valable pour les essais antérieures étant donné que la relation (9) n'est pas encore vérifiée.

Par contre la relation (13) nous donne l'erreur de l'ensemble calculé dans chaque essais.

C'est alors la valeur définie par (13) qu'on doit comparer avec l'erreur expérimentale ε_o .

On peut avoir les trois cas du Tableau 1.

Tableau 1

| | |
|-----------------------------------|------------------------------|
| $\varepsilon_c > \varepsilon_o$: | Structure non terminée |
| $\varepsilon_c = \varepsilon_o$: | Structure terminée |
| $\varepsilon_c < \varepsilon_o$: | Structure affinée avec excès |

II. — Nous allons montrer maintenant que le critère que nous venons d'établir peut être exprimé d'une façon simple en fonction du coefficient de corrélation linéaire r de Bravais-Pearson existant entre les ensembles de valeurs observées et calculées des facteurs de structure.

Rappelons que:

$$r = \frac{\Sigma_j (F_{jo} \cdot F_{jc}) - n\bar{F}_o \cdot \bar{F}_c}{(\Sigma_j F_{jo}^2 - n\bar{F}_o^2)^{\frac{1}{2}} (\Sigma_j F_{jc}^2 - n\bar{F}_c^2)^{\frac{1}{2}}}, \quad (15)$$

avec

$$\left. \begin{aligned} \bar{F}_o &= \frac{1}{n} \Sigma_j F_{jo} \\ \bar{F}_c &= \frac{1}{n} \Sigma_j F_{jc} \end{aligned} \right\} \quad (16)$$

La relation (15) est vraie en grandeur et en signe, et la valeur de r tends vers 1 quand l'accord tend à devenir parfait entre les deux ensembles, et ceci quelle que soit l'échelle de l'observation par rapport au calcul. La pente b de la droite de régression de l'observation par rapport au calcul, c'est à dire le coefficient par lequel il faut multiplier l'observation pour obtenir des valeurs à la même échelle que l'ensemble calculé, est

$$b = \frac{\Sigma_j (F_{jo} \cdot F_{jc}) - n\bar{F}_o \cdot \bar{F}_c}{\Sigma_j F_{jo}^2 - n\bar{F}_o^2}. \quad (17)$$

La variance résiduelle de l'ensemble calculé par rapport à la valeur ajustée du facteur de structure sur la droite de régression sera:

$$\sigma_r^2(F_c^i) = \frac{1-r^2}{n-2} (\Sigma_j F_{jc}^2 - n\bar{F}_c^2). \quad (18)$$

Cette quantité est certainement une bonne approximation de la variance de l'ensemble calculé à la valeur idéale des facteurs de structure, et si nous la substituons dans la relation (12) on a:

$$\varepsilon_c^2 = \frac{1-r^2}{n-2} \frac{\Sigma_j F_{jc}^2 - n\bar{F}_c^2}{\frac{1}{n-2} \Sigma_j F_{jc}^2}$$

ou bien

$$\varepsilon_c^2 = (1-r^2) \left(1 - \frac{n\bar{F}_c^2}{\Sigma_j F_{jc}^2} \right). \quad (19)$$

Remarquons que \bar{F}_c est nul si nous excluons de l'ensemble le terme constant $F(000) = \Sigma_i Z_i$; Z_i étant le nombre atomique de l'atome i contenu dans la maille. Nous devons considérer le cas particulier où il y a un atome situé à l'origine. Dans ce cas il nous

suffira de choisir un autre origine extérieure à tout atome. D'ailleurs cet atome placé à l'origine peut être exclu de la statistique, étant donné que sur sa position il n'y a aucune erreur.

La relation (19) se réduit à :

$$\varepsilon_c^2 = 1 - r^2. \quad (20)$$

Cette expression de ε_c lui doit être substituée dans le Tableau 1.

III. — Mais, quelle est l'erreur ε_o que l'on doit considérer sur l'observation ?

Cette erreur, nous l'avons vu, est composée des erreurs fortuites ε_f et des erreurs systématiques ε_s faites au cours des mesures.

Soit m le nombre minimum d'informations pour que le problème soit déterminé. Compte tenu du nombre total n d'informations dont nous disposons, l'erreur fortuite ε_f' que nous proposons de considérer sur l'ensemble est

$$\varepsilon_f' = \frac{1}{\sqrt{(n/m)}} \cdot \varepsilon_f. \quad (21)$$

L'erreur systématique ε_s n'est pas modifiée par le surplus d'informations. L'erreur moyenne globale est alors

$$\varepsilon_o = \varepsilon_f' + \varepsilon_s. \quad (22)$$

Le nombre n d'observations étant grand, seules subsistent pratiquement les erreurs systématiques puisque les erreurs fortuites sont à peu près éliminées. C'est à dire que leur effet sur le résultat de la structure est négligeable. Nous retrouvons ici la même conclusion vérifiée empiriquement par Robertson & White (1947). Autrement dit, dans les conditions données ($n \gg m$) la limite de la précision de la structure est imposée par les erreurs systématiques.

Nous pouvons estimer celles ci en étudiant la variation du coefficient b de la droite de régression de l'observation par rapport au calcul pour chaque domaine de l'espace réciproque. En particulier pour des domaines correspondants à une même valeur moyenne

du vecteur de diffusion. Cette méthode n'est au fond que celle utilisée pour la correction des erreurs systématiques proposée par Rimsky (1957), que nous résumons ci dessous.

Considérons une structure au stade de détermination dans lequel nous sommes assurés du signe ou de la phase de la plupart des facteurs de structure importants. La valeur du coefficient de corrélation correspondante est d'environ de 0,8–0,9. Pour évaluer les erreurs systématiques on calcule le coefficient b de la droite de régression pour des domaines de l'espace réciproque comprenant un nombre suffisant de facteurs de structure pour que le traitement statistique soit valable, soit au minimum 10 nœuds. On attribue au centre de gravité de chacun de ces domaines un poids égal à la valeur de b correspondante. Les variations de b seront figurées par des courbes d'isovaleur. L'exploitation de cette carte permet de déterminer la fluctuation de b autour de sa valeur moyenne pour chaque valeur du vecteur de diffusion. La fluctuation relative moyenne de b étendue à tout l'espace réciproque donne l'ordre de grandeur de l'erreur relative systématique.

Nous considérons la variance de b :

$$\sigma^2(b) = \frac{1}{\eta} (\Sigma b^2 - \eta \bar{b}^2) \quad (23)$$

avec

$$\bar{b} = \frac{1}{\eta} \Sigma b;$$

On peut alors écrire

$$\varepsilon_s^2 = \sigma^2(b) \frac{1}{\eta} \Sigma b^2, \quad (24)$$

η étant le nombre de domaines explorés de l'espace réciproque.

Cette carte d'isovaleur de b est indépendante des erreurs sur les coordonnées atomiques attribuées au calcul. En effet, la valeur moyenne du module du facteur de structure pris dans un domaine restreint de l'espace réciproque ne dépend que du contenu du cristal comme l'a montré Wilson (1942).

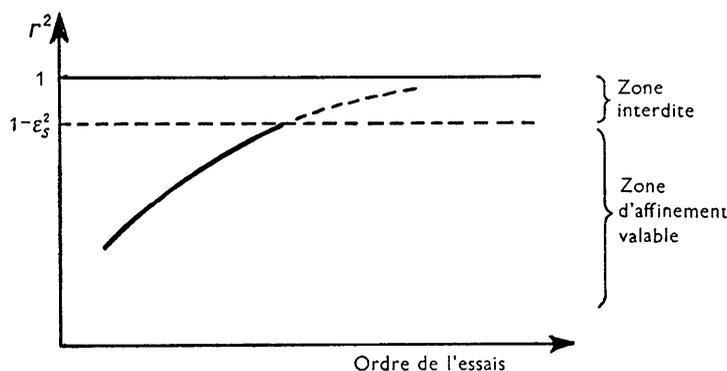


Fig. 1. Illustration de la limite d'affinement.

Dans quatre cas analysés par cette méthode, le rapport de la valeur moyenne minimum de b à sa valeur maximum a été de 2. Ce qui montre que la valeur quadratique moyenne de l'erreur relative systématique est de l'ordre de 20%.

On peut illustrer le sens de la limite du processus d'affinement en portant en abscisses les ordres successifs des essais au cours de ce processus que nous considérons convergent, et en ordonnées, la valeur du carré du coefficient de corrélation linéaire r (Fig. 1). Le coefficient r^2 tend vers 1 quand l'ordre de l'essai croît, c'est à dire que l'accord entre le calcul et l'observation tend à devenir parfait. Mais pour toutes valeurs de r^2 supérieures à $(1 - \varepsilon_s^2)$ la précision physique de la structure reste inchangée. Nous sommes limités par la valeur de l'erreur systématique sur l'observation si le nombre d'observations est grand, et par l'erreur totale (systématique et fortuite) si ce nombre n'est pas très grand par rapport à m .

Toute interprétation de la structure faite au delà de cette limite n'a aucune signification physique. Cette interprétation ne pourra que décrire des modifications de la densité électronique qui ne sont dues qu'aux erreurs de mesure.

Remarque I.—Observons que l'erreur ultime ε_U peut s'écrire:

$$\varepsilon_U^2 = \frac{1}{2} \frac{\sum_j (F_{j0} - F_{jc})^2}{\sum_j F_{jc}^2} \quad (14')$$

et qu'il a une forme semblable au coefficient R_2 défini par Booth (1945):

$$R_2 = \frac{\sum_j (F_{j0} - F_{jc})^2}{\sum_j F_{j0}^2}$$

mais la relation (14) ou (14') a une signification physique parce qu'elle est liée à l'erreur expérimentale ε_o .

Acta Cryst. (1958). **11**, 484

The Crystal Structure of Acetic Acid*

BY ROBERT E. JONES AND DAVID H. TEMPLETON

Department of Chemistry and Radiation Laboratory, University of California, Berkeley, California, U.S.A.

(Received 6 November 1957)

Acetic acid crystals are orthorhombic, space group $Pna2_1$, with $a = 13.32 \pm 0.02$, $b = 4.08 \pm 0.01$, $c = 5.77 \pm 0.01$ Å from single-crystal X-ray diffraction. The structure is similar to that of formic acid with molecules linked into infinite chains by hydrogen bonds. The bond distances are C—C = 1.54 Å, C=O = 1.24 Å, C—O = 1.29 Å, and O—H...O = 2.61 Å, each ± 0.02 Å, by least-squares refinement. The four heavy atoms of each molecule lie in a plane. The adhesion between chains is due to van der Waals forces.

Introduction

The structure of acetic acid (CH_3COOH) has been investigated in the gas by electron diffraction by

* This research was performed under the auspices of the U.S. Atomic Energy Commission.

Remarque II.—Le développement ci-dessus donne des arguments importants pour l'adoption du coefficient de corrélation linéaire de Bravais-Pearson r pour l'évaluation de la précision d'une structure en remplacement du facteur de 'reliability' R utilisé par la plupart des cristallographes ($R = \frac{\sum |F_o - F_c|}{\sum |F_o|}$).

En effet, seul le coefficient r de Bravais-Pearson a un sens statistique et sa grandeur, comme nous l'avons montré, est liée simplement à l'erreur du modèle proposé d'une structure.

Nous préconisons l'usage du coefficient de corrélation linéaire r comme indice de la précision d'une structure.

Nous tenons à remercier tout particulièrement Monsieur le Professeur J. Wyart, ainsi que Messieurs H. Curien, R. Kern et G. von Eller pour l'intérêt qu'ils ont apporté à notre travail et par leurs critiques.

S. C. E. a eu une bourse d'étude de l'UNESCO pendant la réalisation de ce travail.

Bibliographie

- BOOTH, A. D. (1945). *Phil. Mag.* **36**, 609.
 BOREL, E. & DELTHEIL, R. (1940). *Probabilités, Erreurs*. Paris: Colin.
 MORICE, E. & CHARTIER, A. F. (1954). *Méthode Statistique*. Paris: I. N. S. E. E.
 RIMSKY, A. (1957). *Bull. Soc. franç. Minér. Crist.* **53**, 48.
 ROBERTSON, J. M. & WHITE, J. G. (1947). *Proc. Roy. Soc. A*, **190**, 329.
 WILSON, A. J. C. (1942). *Nature, Lond.* **150**, 152.

Karle & Brockway (1944) and in solution by infra-red spectroscopy by Davies & Sutherland (1938). The present work was undertaken to give further information on the structure of the molecule and on the forces between molecules in the solid state.

The normal fatty acids are thought to form dimers