



Soviet-era science, translated into English

Reports of the Academy of Sciences of the USSR

Corresponding Member of the Academy of Sciences of the USSR B.
K. Vainshtein,

1963

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196301.52749>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

Reports of the Academy of Sciences of the USSR
1963. Volume 153, No. 1

CRYSTALLOGRAPHY

Corresponding Member of the Academy of Sciences of the USSR B. K. Vainshstein,

Corresponding Member of the Academy of Sciences of the USSR I. M. Gelfand,
R. L. Kaushina, Yu. G. Fedorov

FINDING CRYSTAL STRUCTURES BY THE METHOD OF MINIMIZING THE R -FACTOR

X-ray diffraction analysis of the atomic structure of crystals is based on measuring the intensities of X-ray reflections from crystals, I_{hkl} , from which the squares of the moduli of the structure amplitudes, $|F_{hkl}|^2$, are directly found. The first stage in determining a structure is usually considered to be the finding of a preliminary model in which all or most of the atoms are placed at positions differing from the true ones by no more than 0.1–0.2 Å. The second stage is refinement of the structure.

The criterion for the correctness of the preliminary model found is taken to be the value of the discrepancy factor

$$R = \frac{\sum_{hkl} ||F_{hkl}| - |F_{hkl}|_{\text{expt}}|}{\sum_{hkl} |F_{hkl}|_{\text{expt}}} \quad (1)$$

of the order of 15–20%. The discrepancy factor R is a function of the 3*N* coordinates of the atoms of the cell entering into the expression

$$F_{hkl} = \sum_{j=1}^N f_j \exp 2\pi i(hx_j + ky_j + lz_j). \quad (2)$$

It has long been suggested by various authors (¹, ²) that the function R itself, or other functions similar to it, can be used for finding a preliminary model, since the correct structure corresponds to a minimum of R . However, the application of known minimization methods to the function R , or to other analogous functions, gives results only in solving refinement problems, when the preliminary model has already been found (³, ⁴). Preliminary models from the function R ,

either by these methods or by any others, have not yet been obtained. The reason is that there is no general method for finding the minimum of a function of many variables.

In papers (⁵, ⁶), a method of “ravines,” or nonlocal search, is developed for finding the global minimum of functions of many variables, and it is indicated how it can be applied to problems of structural analysis. The essence of the method consists in rapidly obtaining regions of low values of the function sought and examining these regions. In this, the “proper organization” of the function sought is essential. We note that in the ravine method there are two important numerical parameters—the “gradient trial” and the “step along the ravine” — whose choice depends on the particular function being minimized. After these numerical parameters have been chosen, the calculations are entrusted entirely to the machine.

Instead of the function R itself of $3N$ variables, which has a complicated nature, it is more expedient to use a “parametrized” function \tilde{R} , depending on a smaller number of variables p (⁶). Parametrization is most simply carried out for molecular structures, when the shape of the molecule

known in advance. In this case the parameters are the usual parameters of a solid body: the coordinates of the center of the molecule, which may be chosen as any of its atoms, and the Euler angles of rotation. Expressing the coordinates of the atoms through the introduced parameters and substituting them into (1) and (2), we obtain the “parametrized” function \tilde{R} .

It is essential not only that \tilde{R} depends on a smaller number of variables, but also that it is more “correctly” organized, since it already takes into account the known information on the mutual arrangement of the atoms.

The number of reflections required for the preliminary search is connected with the number of independent variables p of the function \tilde{R} . It turned out that, with a reasonable selection of reflections, it is convenient to take their number at the beginning of the search to be of the order of $7p$ — $10p$. In doing so, the following should be kept in mind. Reflections with small hkl correspond to low harmonics of the Fourier expansion and are less sensitive to inaccuracies in specifying the coordinates of the atoms in the model of the initial molecule. It is therefore advisable to use them at the first stage of the search. Strong reflections, which carry the greatest information, must be included in the search, while weak ones may be excluded from it. However, some strong reflections may be appreciably lowered because of extinction. Therefore the following restriction is imposed on them: if at some point $|F_{hkl}|_{\text{exp}} < |F_{hkl}|$, then the term $||F_{hkl}| - |F_{hkl}|_{\text{exp}}||$ entering the function \tilde{R} is set equal to zero at this point.

Fig. 1. Initial model of *l*-proline. Atom C_4 is in the trans position to $C_1O_1O_2$ relative to the plane of the pyrrolidine ring

When selecting and evaluating minima it is necessary to increase the number of reflections entering the function \tilde{R} in the region of these minima. In addition,

Fig. 1. Initial model of l-proline. Atom C4 is in the trans position to C1O1O2 relative to the plane of the pyrrolidine ring

Figure 1: Fig. 1. Initial model of l-proline. Atom C4 is in the trans position to C1O1O2 relative to the plane of the pyrrolidine ring

for selecting minima one can use the criterion of permissible intermolecular distances (7). A function was constructed which helped to bypass minima of \tilde{R} not satisfying this criterion.

Naphthalene $C_{10}H_8$, space group $P2_1/a$, $z = 2$ (8), was chosen as the first trial structure. The position of the centrosymmetric naphthalene molecule is determined by three angular parameters. The inclusion of 30 reflections clearly revealed 2 symmetrically related minima corresponding to the true arrangement of the naphthalene molecules in the crystal.

The possibilities of the new method were also tested on the already known non-centrosymmetric structure of *l*-oxyproline $C_5H_9O_3N$ (9). The space group is $P2_{12_12}1$, $z = 4$. A search from a certain arbitrary position of the molecule in the cell, first using 40 reflections ($\sin \theta/\lambda < 0.35$) and then 64 reflections ($\sin \theta/\lambda < 0.4$), with allowance for the extinction restriction, led to 4 symmetrically related regions of minima corresponding to the true structure. The function \tilde{R} proved to be organized more complexly than in the case of naphthalene.

It is important to emphasize that, as in the case of naphthalene, when constructing the functions \tilde{R} we proceeded from the true arrangement of the atoms in the molecule, already known with high accuracy, which in the case of an unknown structure is practically impossible to do. However, small distortions of the molecule, leading to deviations of the atoms from their true positions in the molecule by amounts of the order of 0.1–0.2 Å, did not shift the regions of minima of the function \tilde{R} , but led only to an increase in its values. At the same time, large distortions could, generally speaking, lead to a disruption of the “correct” organization of the function.

Finally, the nonlocal search method was applied to determine

previously unknown structure of one of the natural amino acids, *l*-proline, $C_5H_9O_2N$. The unit cell is $a = 11.44$ Å, $b = 9.02$ Å, and $c = 5.20$ Å, space group $P2_{12_12}1$, $z = 4$ (which agrees with the data of work ⁽¹⁰⁾). About 500 moduli of the structure amplitudes were determined with an accuracy of the order of 10–15%. For the parametrization, the molecular model previously obtained in determining the structure of $Cu(C_5H_9O_2N)_2 \cdot 2H_2O$ ⁽¹¹⁾ was adopted (see Fig. 1). In addition to the 6 rigid-body parameters (coordinates of atom C_2 , x, y, z , and Euler angles $\varphi_1, \theta, \varphi_2$), one more angular parameter, χ , was introduced, characterizing the rotation of the carboxyl group $C_1O_1O_2$ about the C_1C_2 axis. The remaining atoms of the molecule were fixed (atom C_4 in the trans position to the $C_1O_1O_2$ group relative to the plane of the pyrrolidine ring).

Fig. 2

Figure 2: Fig. 2

The initial search was carried out on 30 reflections with $\sin \theta / \lambda < 0.2$ and led to many poorly distinguishable minima. After the number of reflections was increased to 60 ($\sin \theta / \lambda < 0.35$), with inclusion of most of the strong reflections, the number of minima decreased, and a broad region with low values of the \widetilde{R} -function (of the order of 30%) was found. This region proved unacceptable for crystal-chemical reasons. Then the intermolecular-distance function mentioned above was introduced into the search. However, the combined use of the distance function and the \widetilde{R} -function did not make it possible to leave this broad region.

Fig. 2. *a*—level lines of the function \widetilde{R}_{200} in the xy plane (x, y —fractions of the unit-cell parameters); *b*—level lines of the function \widetilde{R}_{200} in the $\varphi_1\varphi_2$ plane ($\varphi_1\varphi_2$ —in radians)

This was apparently connected with the incorrect fixing of one or several atoms in the starting molecule (the inaccuracy, as the example of hydroxyproline showed, should have been > 0.2 - 0.3 Å). Therefore one more parameter was introduced—the displacement of atom C_4 relative to the plane of the ring, since for structures containing proline different positions of atom C_4 had been indicated. Ten strong reflections not previously included in the search were introduced, with the extinction restriction described above. This led to the finding of a minimum with $\widetilde{R}_{200} = 32\%$ (with 200 reflections taken into account), satisfactory with respect to intermolecular distances. Atom C_4 proved to be in the cis position. Nevertheless, attempts to refine this structure by the least-squares method did not lead to a decrease in the R -factor. Construction of two-dimensional Fourier series and calculation of R separately for the $hk0$ -, $h0l$ -, and $0kl$ -reflections showed that the orientation of the molecule and the yz -coordinates of its center were apparently determined correctly. Moving the ravine along the x axis led to a deeper minimum corresponding to the correct solution, in which $\widetilde{R}_{200} = 20\%$ with 200 reflections taken into account and $R = 29\%$ for all reflections. The vicinity of $\widetilde{R}_{200} \leq 25\%$ of this point has a width of about 0.3 Å. The level lines $R_{200} = 25\%$ and $\widetilde{R}_{200} = 30\%$ in the xy and $\varphi_1\varphi_2$ planes, characterizing the relief of the function \widetilde{R}_{200} in the region of the minimum, are shown in Fig. 2, *a, b*. In this region we carried out, by the ravine method, a refinement of the positions of all atoms, passing from the function \widetilde{R} to R , which gave $\widetilde{R} = 21\%$ for 420 reflections and $R = 23.5\%$ for all reflections. At this point the search by the ravine method was completed.

Additional criteria that we used in selecting minima and fixing the final solution were: a) fulfillment of crystal-chemical regularities; b) convergence of refinement cycles by the least-squares method; c) analysis of projections of electron density (absence of false peaks, etc.) and separate calculation of R by zones.

The projection of the electron density onto the xy plane for the model with

$R = 20.7\%$ is shown in Fig. 3.

It is interesting that, apparently, there is a possibility of finding the correct orientation of the molecule even when its position in the cell is still incorrect (or, more precisely, regardless of its position). This follows from the following. The quantities $|F_{hkl}|^2$ are the Fourier coefficients of the function of interatomic distances. With the correct orientation of the molecule, all intramolecular interatomic vectors corresponding to the peaks of the Patterson function coincide. This means that in such a case, in the

Fig. 3. Projection of the electron density onto the xy plane (the signs are calculated from the coordinates of the model with $R = 20.7\%$). The contour lines are drawn at intervals of $4 e/\text{\AA}^2$.

calculated value of $|F_{hkl}|$ (2) there is a definitely correct component, whereas with an incorrect orientation this component is also incorrect. Therefore the correct orientation corresponds to minima of the function \hat{R} with respect to the angular variables.

Thus, minimization of the R -factor by the ravine method proved effective for finding a preliminary model of molecular structures. It may be assumed that further improvement of the method will make it possible to apply it to more complex structures.

The authors express their gratitude to A. I. Kitaigorodskii, who suggested using the criterion of permissible intermolecular contacts, and to I. I. Pyatetskii-Shapiro for valuable advice and discussion of the questions raised. The authors thank L. N. Ivanova and S. L. Ginzburg for active assistance in developing the method and in carrying out calculations on oxyproline and *l*-proline.

Received
16 VII 1963

REFERENCES

1. A. D. Booth, *Nature*, **160**, 196 (1947).
2. V. Vand, A. Niggly, R. Pepinsky, *Acta Crystallogr.*, **13**, 12, 1001, 1002 (1960).
3. M. I. Porai-Koshits, *Practical Course of X-ray Structural Analysis*, 1960.
4. B. K. Vainshtein, *Structural Electron Diffraction*, 1956.
5. I. M. Gelfand, M. L. Tsetlin, *Uspekhi Mat. Nauk*, **17**, issue 1 (103) (1962).
6. I. M. Gelfand, I. I. Pyatetskii-Shapiro, Yu. G. Fedorov, *DAN*, **152**, No. 5 (1963).

7. A. I. Kitaigorodskii, *Organic Crystal Chemistry*, 1955.
8. S. C. Abrahams, J. M. Robertson, J. G. White, *Acta Crystallogr.*, **2**, 238 (1949).
9. J. Donohue, K. N. Trueblood, *Acta Crystallogr.*, **5**, 414 (1952).
10. B. A. Wright, P. A. Cole, *Acta Crystallogr.*, **2**, 129 (1949).
11. D. Mch. Matthieson, H. K. Welsh, *Acta Crystallogr.*, **5**, 599 (1952).

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.