Acta Cryst. (1989). A**45**, 4–10

Minimization of Cross Entropy: a Tool for Solving Crystal Structures

By Robert W. Harrison*

Center for Chemical Physics, National Bureau of Standards, Gaithersburg, MD 20899, *USA*

(*Received* 14 *September* 1987; *accepted* 12 *July* 1988)

Abstract

A new algorithm for the solution of crystal structures *via* the principle of minimum cross entropy is described. This method is an extension of direct methods. Conventional direct methods are used to find an initial set of phases which are extended and refined by a search for the minimum-cross-entropy solution. This search is done directly in terms of the phases rather than the Lagrange multipliers used in most other approaches. It was tested by solving two structures, a small molecule and a macromolecule of 610 atoms in the asymmetric unit.

Introduction

The maximum entropy method has stimulated much interest in the expansion of direct methods to phase determination for macromolecules. This paper details the use of the cross entropy to solve two structures. One of these structures is a small molecule, thiolysine, which is readily solved by conventional direct methods. The other structure is a 15-base oligomer of deoxyribonucleic acid (DNA), which would be difficult to solve by conventional direct methods as there are 610 atoms in the asymmetric unit and the crystals diffract only to about 3 Å resolution.

Density modification techniques, including the maximum entropy method, are often treated in isolation from the rest of the process of structure determination. A comparison with direct methods is useful. Direct methods are more complicated than just the use of the tangent rule to refine phases. The choice of the origin-determining reflections, the expansion of these with or without extra symbols to form a starting set, and the use of appropriate statistics to monitor the convergence of the process are all examples of issues which must be properly addressed for direct methods to find a solution. Neglect of the geometric and statistical framework on which density modification and direct methods are implicitly based is a recipe for failure, while with appropriate statistical treatment even a very crude and approximate density modification scheme may work.

When entropy is used to find a solution to a crystal structure similar issues must be addressed. The process is started in the same way as direct methods; an origin set of reflections is chosen and expanded with Sayre's equation. This approximate solution is then refined with an entropy-based algorithm to produce a more interpretable structure. Maximum entropy is used for the same purpose as the tangent rule in direct methods.

Here, entropy is defined similarly, and is given a similar treatment to its use in other entropy-based approaches (Collins, 1982; Wilkins, Varghese & Lehmann, 1983; Bryan, Bansal, Folkhard, Nave & Marvin, 1983; Marvin, Bryan & Nave, 1987; Bricogne, 1984; Wilkins & Stuart, 1986). However, the algorithm for finding this solution is different from the other approaches. The terminology is somewhat different. The cross entropy used here is simply the negative of the relative entropy (Hobson & Cheng, 1973; Shore & Johnson, 1980, 1981). While this may initially cause some confusion, referring to the minimization of cross entropy rather than the maximization of relative entropy better reflects the underlying mathematics.

Entropy and cross entropy

Entropy (H) is defined as the expected value of the information (Shannon, 1948; Yu, 1976). The information is defined as the negative logarithm of the probability distribution.

$$H = \int \mathrm{d}V\rho(-\ln\rho) \qquad (1)$$

where ρ is the electron density and $\mathrm{d}V$ is the differential element of volume. The cross or relative entropy is defined as the expected value of the difference in the information between some initial estimate and the current estimate:

$$H_c = \int \mathrm{d}V\rho[\ln(\rho/\mathrm{prior})] \qquad (1a)$$

where prior is the initial or prior estimate for ρ. Positivity is implicit in these definitions, but both the logarithm and its expected value can be extended to treat negative density values. When the prior is chosen to have a constant value, these equations are equivalent (Hobson & Cheng, 1973). In the absence of further knowledge the maximum entropy distribution is flat. Typically this would correspond to all of

---

* Current address: Crystallography Laboratory, NCI-Frederick Cancer Research Facility, BRI-Basic Research Program, PO Box B, Frederick, MD 21701, USA.

0108-7673/89/010004-07$03.00

the knowledge available from the $F(000)$. This would be the least informative prior that could be chosen. In a Bayesian sense this is the most reliable in that the fewest unwarranted assumptions about ρ are made. More strictly, it can be shown from a limited set of axioms about statistical consistency and uniqueness that cross entropy forms the most general rule for inference (Shore & Johnson, 1980).

Constraints *vs* coordinates

Traditionally additional knowledge not included in a prior is introduced in the form of constraints (Jaynes, 1957, 1968; Frieden, 1972, 1981; Gull & Daniell, 1978; Bricogne, 1984; Bryan & Skilling, 1986; Navaza, 1986; Shore & Johnson, 1980, 1981). The ability to use constraints was one of the original reasons for the choice of entropy as an information measure (Jaynes, 1957). The algebraic form of the constraints could affect the nature of the results. If the results depend on some property of the algebraic form of the constraints, and the form of the constraints is an arbitrary choice, then the advantages of maximum entropy as a method of inference are lost. Rather than choose any specific form of the constraints, we shall use a differentiable function, χ, which could represent any constraint. This results in a modified entropy H',

$$H' = H + \chi. \qquad (2)$$

The solution for ρ of the modified entropy can be accomplished by variational means. As long as ρ and χ are functions of the same variables this results in the differential equation

$$\mathrm{d}\chi - \ln \rho - 1 = 0 \qquad (3)$$

or

$$\rho = \exp(-1 + \mathrm{d}\chi).$$

$\mathrm{d}\chi$ can then be adjusted to find the density ρ that meets the constraints.

This approach, referred to as exponential modeling (Bricogne, 1984; Collins & Mahar, 1983), is not the best algorithm for finding the solution. It does, however, provide the foundation for some useful results. In particular it allows a definition of the 'natural' form of the constraint term χ. Suppose ρ_c is the correct solution. From (3) it is trivial to find $\mathrm{d}\chi$ such that ρ will equal ρ_c. χ is then $\int \mathrm{d}V \rho_c \ln \rho_c$. A further term can be defined from this. The mutual information (M) is the sum of χ and H:

$$M = H' = \int \mathrm{d}V\,(\rho_c \ln \rho_c - \rho \ln \rho). \qquad (4)$$

Since $\rho_c \ln \rho_c$ and $\rho \ln \rho$ share the same functional form, this suggests that the problem of finding the constraints may be transformed from a variational problem into some other form. In particular, this suggests that the dual problem may be more tractable. A dual problem is an equivalent formulation of a problem where coordinates and constraints have been interchanged; for example, reformulation of circular motion in polar rather than rectangular coordinates. The only experimental data available in a standard crystallographic experiment are the magnitudes of the structure factors, $|F|$. There may, of course, be several sets of these values corresponding to different heavy-atom derivatives for isomorphous replacement. The constraints have been introduced primarily to reproduce these magnitudes. Therefore, a logical choice for a dual problem is to replace ρ by $\sum_{\mathbf{h}} |F_{\mathbf{h}}| \exp(i\alpha) \exp(2\pi i \mathbf{h}.\mathbf{x})$ and only allow the phase α to vary. Each constraint for a reflection is replaced by a coordinate α. The minimizer for the cross entropy can then find the solution by searching these coordinates.

A direct minimization of the cross entropy in terms of the free or independent variables of the system simplifies the problem of finding a maximum entropy and minimum cross entropy solution. By avoiding constraints, the convergence problems associated with finding them are also avoided. A dual problem where the constraints are replaced by coordinates is constructed. This is equivalent to examining many maps, all of which have correct $|F|$'s, to find the minimum-cross-entropy map, rather than examining many maps of which only some have correct $|F|$'s. The advantage of this approach is that the minimizer is tuned to a specific problem such as determining values for unknown phases. This is also a disadvantage in that the algorithm may not be easy to transport to a general problem.

Normalization

Before detailing this algorithm, the problems associated with normalization must be defined. Normalization can be handled in two ways with constraint-based algorithms. The simplest approach is to calculate the partition function, $\int \mathrm{d}V\rho$, and use this explicitly to normalize ρ (Jaynes, 1957, 1968; Berger, 1985). The normalization may also be included as a constraint. A simple constraint of the form $\sum \rho$ is sufficient.

A similar choice of approaches is available when the minimization is done directly as a function of the phases without constraints. First, a normalization constraint may be added to formulation, in which case there is implicit normalization. Following this path results in the minimization of the mutual information. The other approach is to normalize explicitly during the minimization; this is equivalent to the use of the partition function in the constrained minimization. The solutions to both approaches are equivalent, but there may be a difference in convergence. Explicit normalization was used in the solution of the structures described here. It tends to be more convergent than implicit normalization.

In general, normalization is not a problem with Fourier data. Parseval's theorem shows that when the magnitudes of the structure factors are correct, the density is properly scaled (Tolstov, 1976). In addition, the $F(000)$ term must be included in the Fourier transform for positivity to be defined. As long as this term and the rest of the Fourier coefficients are finite, and the Fourier series is band limited, then the density is bounded and the normalization is automatically defined.

A practical algorithm

For a crystallographic problem the phase of the reflection is varied to find the minimum-cross-entropy solution. This minimization is performed in a complex space and without simplification can be very difficult. The basic equations for the full-scale maximization of a related entropy, the Burg entropy, are derived by Narayan & Nityananda (1982), and are easy to generalize to any other form of the entropy. Unfortunately they are not easy to apply because the second derivative of the entropy with respect to two independent reflections is non-zero, which means that it is difficult to construct an entropy maximizer, as the matrix of second derivatives will be both large and densely populated.

The general scheme is one of a modified Newton's method with projection (Scales, 1985). Starting from an initial set of structure factors, with phases chosen independently from this minimization, one can find an initial estimate for the electron density with a Fourier transform. In real space a shift towards the minimum cross entropy is taken, and, after another Fourier transform, the new values for the structure factors are used to update the phases. This scheme is outlined in Fig. 1. As the Fourier transform is a complete orthogonal sequence, figures of merit and combinations of $|F_o|$ and $|F_c|$ may be used. Since algorithms for the calculation of the Fourier transform are well known, only the details of the shift in real space will be described.

The real-space step in the minimization is done in the following manner. The first and second derivatives of the unconstrained cross entropy in real space are simply

$$1+\ln(\rho/\text{prior}) \quad \text{and} \quad 1/\rho.$$

A simple Newton–Raphson minimizer would then be

$$\rho' = \rho - \rho[\ln(\rho/\text{prior}) + 1].$$

Completely unconstrained, this would converge to a value of $\rho = e^{-1}$ prior. This simple form is not quite stable enough to be practical. When ρ is larger than prior it will first send ρ to a value smaller than prior and then converge. A simple modification to the second derivative, $1/\rho$, fixes this problem. When $1/\rho$ is small, it is replaced with a larger value. The constant 8 in the modified second derivative was chosen as the smallest number for which stale convergence occurred in three-dimensional tests:

$$\rho' = \rho - \{\min[\rho, e^{-1}\text{prior} + (\rho - e^{-1}\text{prior})/8]\} \times [\ln(\rho/\text{prior}) + 1]. \qquad (5)$$

A similar form can be found for the mutual information,

$$\rho' = \rho - \{\min[\rho, \text{prior} + (\rho - \text{prior})/8][\ln(\rho/\text{prior})]\}, \qquad (5a)$$

the only difference being that the mutual information will converge directly to ρ. For (5) the scale between ρ' and ρ will be e^1. The difference between them is in the treatment of normalization. The iteration defined by (5) appears to be more convergent than that of (5a). In both of these equations the first derivative of the entropy with respect to the phase is used exactly, since by the chain rule for differentiation it is proportional to the Fourier transform of this term, but the second derivative is treated approximately. When ρ is negative or zero then the solution for an approximate bound to the entropy, $\rho = |\rho| + \varepsilon$ (ε is a small constant dependent on the computer representation of a floating-point number, approximately 10^{-7} for a 32-bit machine), is used first and the new value for ρ is refined with (5) or (5a).

After performing this modification, the phases are calculated from the Fourier transform of ρ' and the

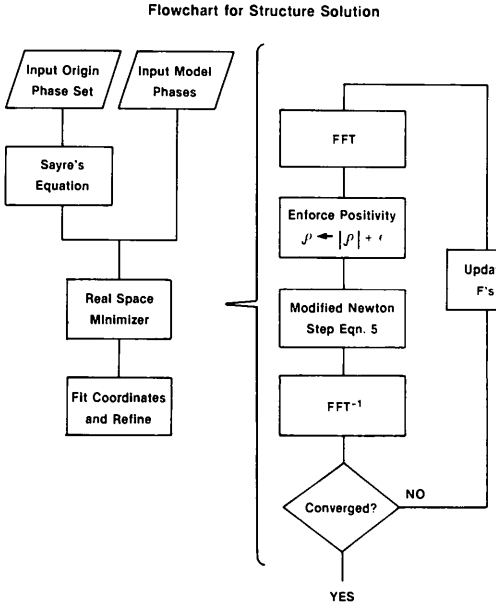

Fig. 1. The algorithm described in this paper is shown in this flow chart. The overall algorithm is described on the left, and the modified Newton's-method minimizer is shown in the expanded section.

phases of the reflections are set to the new values. Either this can be done in $P1$ or the space-group symmetry in reciprocal space may be preserved by only updating the asymmetric unit and then expanding that to the complete $P1$ set. It is useful to calculate figures of merit based on F_o and F_c. The figure of merit is a summary of the probability of $|F_o| \exp(i\alpha)$ given $|F_c| \exp(i\alpha_{\rm calc})$ found by $\int {\rm d}\alpha\, P(\alpha)|F_o| \exp(i\alpha)$. Assuming a delta function at $|F_c| \exp(i\alpha_{\rm calc})$ as a prior distribution, maximum entropy can be used to find the posterior distribution for $|F_o| \exp(i\alpha)$. The general maximum-entropy solution for the Burg entropy, $\int \ln \rho \,{\rm d}V$, in such a case is a polar expansion of the form

$$\sum_n a_{2n}/||F_o| \exp(i\alpha) - |F_c| \exp(i\alpha_{\rm calc})|^{2n}$$

(Press, Flannery, Teukolsky & Vetterling, 1986). This distribution is not sensitive to the values for the constants a_{2n}, and the values $a_{2n} = (2n)!$ are a good, although not normalized, approximation. The only adjustable parameter is then the scale between F_o and F_c which is a constant that depends only on the form of the iteration (the dependence on the average error between F_o and F_c would be in the numerator and would be lost in the normalization used when finding the figure of merit).

The termination criteria that have been most useful with this iteration scheme are the crystallographic R factor and the root mean square phase shift. Convergence has occurred when the r.m.s. phase shift stays small for several consecutive cycles and the R factor is low.

Reciprocal-space maximum entropy

Real-space cross-entropy minimizers or entropy maximizers have a serious problem when starting from a small number of reflections such as the set which defines the crystallographic origin. When there is very little prior information, the map calculated from such a phase set is closer to the featureless global entropy maximum than the final solution. Under such circumstances the derivatives used to find the new phases will be zero or nearly zero, and it is difficult to proceed as the phase shifts will be poorly conditioned and unreliable. One could give all the phases random values, but then the value of the origin and enantiomorph-determining phases would be lost. This would be equivalent to attempting to sort out an equal mixture of all of the origins and enantiomorphs, which is very difficult. What is necessary is a way to sort out those reflections whose phases are highly dependent on the limited origin set, and then to find phase values which are near to the entropy maximum.

Since the Fourier transform of a logarithm is analytically untractable, it is necessary to expand the logarithm in a power series to find the entropy in terms of structure factors. Expanding $\rho \ln \rho$ as a power series and integrating produces a series of bounds to the entropy, the simplest being equivalent to Sayre's equation (Harrison, 1987). These bounds can be directly evaluated in reciprocal space on a limited set of reflections. The lowest-order bound appears to be the most useful because the complexity of calculating these bounds is exponential with the number of terms in the bound (it is of polynomial complexity with the number of reflections). The higher-order bounds usually make better estimates for the phases, but seldom are the estimates worth the extra cost. These expansions can only be used a few times on a four-reflection origin set before the errors in the approximation accumulate to where the result is nearly random.

The use of these polynomial bounds differs from the use of Sayre's equation in classical direct methods. In direct methods a subset of the reflections for which this relationship is 'most accurate' is chosen, and the phases of these are adjusted to find a solution (Ladd & Palmer, 1980). When used with entropy this relationship is simply a convenient short-cut to the solution and, with the possible exception of point-group dependence, no special weighting should be used. There is no probability associated with individual terms in the equation when used in an entropic sense.

Polynomial expansion of the entropy also simplifies the treatment of negative values of ρ. Clearly the algorithm described above (5) for finding the maximum entropy is not defined when ρ is less than or equal to zero. This condition, however, is unavoidable during the process of finding the maximum. Use of a polynomial bound in place of the entropy for small values of ρ defines the appropriate way to proceed when ρ is small or negative. With the simplest bound, ρ is replaced with its absolute value.

Proof of concept

The general approach for solving an unknown structure can now be defined. First, as in direct methods, an origin and enantiomorph set of reflections is chosen. The same geometric constraints on primitivity and unique origin (for example, origin reflections must not be self-seminvariant) should be followed (Rogers, 1980). It is also useful to choose strong reflections, and the enantiomorph-defining reflection should interact strongly with the origin-determining set. An initial starting set of closely linked reflections is phased from this small set by the use of a polynomial bound to the entropy (Sayre's equation). Finally, this somewhat enlarged set of reflections is refined *via* a real-space entropy maximizer. During this process the space-group symmetry is maintained by updating the asymmetric unit and expanding that to the full $P1$ set. The solution of a partially known structure would simply skip the first steps and use the partial structure to provide the initial phases.

A small simple structure was solved in order to show that this approach could work. Thiolysine hydrochloride (*S*-aminoethylcysteine) crystal diffraction data collected by Ammon & Gerlt (1987) were used. The X-ray data to 0·75 Å resolution were utilized. There were 755 independent observations. The crystal belongs to the space group $P2_1$ with a unit cell of $a = 5{\cdot}1275$, $b = 7{\cdot}9809$, $c = 11{\cdot}2079$ Å and $\beta = 104{\cdot}127°$. This structure is readily solved by direct methods, but no coordinate or phase information from the known structure was used in this trial. The reflections $\bar{3}06$, $\bar{2}14$, 101 were chosen as the origin set, and the phase of the reflection $\bar{3}58$ was set to 45° to fix the enantiomorph. Three cycles of Sayre's equation were used to expand this set to 593 phases. The structure was not visible in a map calculated from this phase set. After 23 cycles of cross-entropy minimization the structure was clearly visible. A model fit to the map refined to a 10% crystallographic *R* factor without the addition of H atoms or anisotropic thermal factors.

This approach is *not* restricted to small molecules at high resolution. Crystals of a 15-base pair oligomer of DNA with an inserted unpaired base were obtained as described by Miller, Wlodawer, Appella & Sussman (1987). These crystals were from the orthorhombic space group $I222$ with a unit cell of $a = 36{\cdot}99$, $b = 53{\cdot}7$ and $c = 101{\cdot}6$ Å and diffract only to about 3 Å resolution. This structure could not be solved by multiple isomorphous replacement (MIR) methods because we failed in the attempt to prepare isomorphous heavy-atom derivatives. This problem was also chosen because the DNA molecule has a regular structure which would be recognizable even in a noisy map. The data were collected at the National Bureau of Standards with a two-dimensional area detector with oscillation geometry (Howard, Gilliland, Finzel, Poulos, Ohlendorf & Sálemme, 1987). It was necessary to cool the crystals during the data collection as they are very thermosensitive. The data from one crystal were used in the phasing. Approximately 75% of the reflections were observed [$I > 1{\cdot}5\sigma(I)$] at 3 Å resolution, but the missing data were in the form of a wedge rather than randomly distributed in reciprocal space.

The reflections 031, 303 and 510 were given phases of 0 to determine the origin, and the reflection 114 was given a phase of 90° to set the enantiomorph. This set was expanded in three cycles of Sayre's equation resulting in phases for 1100 out of 1211 reflections. A map was calculated at this stage and while it had encouraging features like large 'solvent' areas it was not interpretable in terms of a realistic molecular structure (Fig. 2). These 1100 reflections were refined for four cycles, and then the rest of the reflections were added. These were refined for 100 cycles (which was more than required for convergence). The initial *R* factor was 0·367 and the r.m.s. phase shift was 40°. The final values were 0·121 and 2·8°. The map calculated at this step clearly showed the location and orientation of the DNA molecule. Fig. 3 shows a 20 Å section of this map, and Fig. 4 shows the equivalent section of a $2F_o - F_c$ map calculated after refinement. Two copies of the molecule are visible. The molecule runs diagonally from the middle of the figure at the bottom of the map to the upper right corner. Part of the molecule is above the

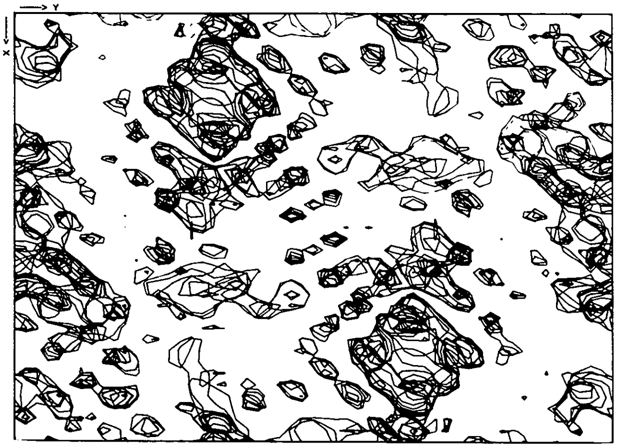

Fig. 2. The map calculated with the phases from the use of Sayre's equation is shown in this 20 Å section. While it shows interesting features like large 'solvent' areas, it is not interpretable.

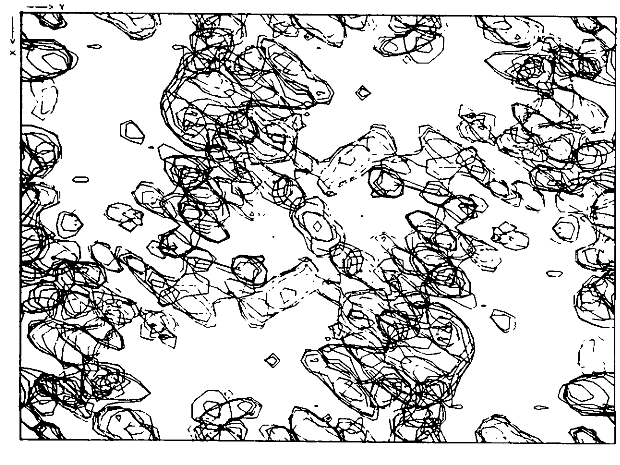

Fig. 3. A 20 Å section of the minimum-cross-entropy map of the DNA. The sections are in the *xy* plane with the crystallographic origin at the middle of the left edge of the map. Two molecules are shown in this map. One molecule runs from the center of the lower edge of the map to the upper right corner, and the other is rotated about the twofold axis through the center of the map. The molecules run between a pair of non-equivalent twofold axes. The upper right-hand part of the map is somewhat 'washed out'. Partially, this is caused by phase errors, but the molecule also runs out of the top of the map in this region. Phase errors are also evident in the density on the twofold axis in the center of the map.

map sections shown, which accounts for the partial washing out of density in the upper right corner. The density on the crystallographic twofold axis in the center of the minimum-cross-entropy map (Fig. 3) shows the presence of some error in the phases. Based on inspection of the diffraction pattern in a precession photograph, it was expected that the helix axis would lie in the xy plane of the unit cell (Miller, Wlodawer, Appella & Sussman, 1987). This indeed was the case. In addition, the 3·3 Å base-pair separation, the location of the major and minor grooves, and the helical pitch were clearly visible in this map. The density was fit by manually inserting a dodecamer structure (Wing, Drew, Takano, Broka, Tanaka, Itakura & Dickerson, 1980) into the map and adjusting it for the difference in the helix pitch. Standard molecular refinement techniques and refitting the map with the correct sequence produced a structure with a crystallographic R factor of 26·3%, against a more complete data set from several crystals, which has been reported elsewhere (Miller, Harrison, Appella, Wlodawer & Sussman, 1988).

Both of these structures were done relative to the uniform prior distribution $F(000)/V_{\text{cell}}$. No advantage was taken of a more detailed prior, and these solutions could be considered as simply 'maximum-entropy' solutions as well.

Towards a global algorithm

The algorithm as presented here is not globally convergent. There are several test cases, including some small molecules which conventional direct methods can solve, which the current algorithm has been unable to solve. The reasons for this are not completely clear, but it appears that this algorithm, when it fails, has a tendency to converge to an equal mixture of enantiomorphs.

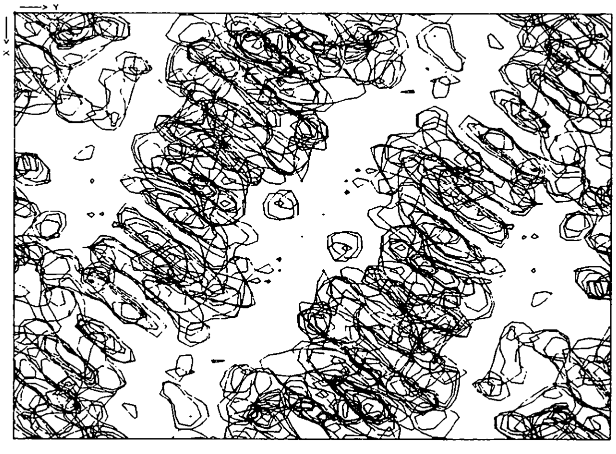

Fig. 4. The same 20 Å section as in Figs. 2 and 3, but of the $2F_o - F_c$ map, is presented for comparison purposes. While the density is more connected than in Fig. 3, the same basic features and molecular envelope remain.

This failing is probably due to a deficiency in the minimizer. While the first derivative of the cross entropy is calculated exactly in (5) and (5a), the second derivative is only approximated. Also, a modified Newton's-method algorithm is used, which is not the best available minimizer for a multivariate problem.

In addition, the use of Sayre's equation to obtain a starting set was done in a naive manner and a more sophisticated approach to finding a starting set should be used.

It should also be noted that only a uniform prior has been used in the current work and it is likely that a properly chosen prior will enhance the convergence. If nothing else, a limited prior can be used to enforce the enantiomorph more strongly.

References

AMMON, H. L. & GERLT, J. (1987). Private communication.

BERGER, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. New York: Springer-Verlag.

BRICOGNE, G. (1984). *Acta Cryst.* **A40**, 410-445.

BRYAN, R. K. BANSAL, M., FOLKHARD, W., NAVE, C. & MARVIN, D. A. (1983). *Proc. Natl Acad. Sci. USA*, **80**, 4728-4731.

BRYAN, R. K. & SKILLING, J. (1986). *Opt. Acta*, **33**, 287-299.

COLLINS, D. M. (1982). *Nature (London)*, **298**, 49-51.

COLLINS, D. M. & MAHAR, M. C. (1983). *Acta Cryst.* **A39**, 252-256.

FRIEDEN, B. R. (1972). *J. Opt. Soc. Am.* **62**, 511-518.

FRIEDEN, B. R. (1981). *J. Opt. Soc. Am.* **71**, 295-303.

GULL, S. F. & DANIELL, G. J. (1978). *Nature (London)*, **272**, 686-690.

HARRISON, R. W. (1987). *Acta Cryst.* **A43**, 428-430.

HOBSON, A. & CHENG, B. K. (1973). *J. Stat. Phys.* **7**, 301-310.

HOWARD, A. J., GILLILAND, G. L., FINZEL, B. C., POULOS, T. L., OHLENDORF, D. H. & SALEMME, F. R. (1987). *J. Appl. Cryst.* **20**, 383-387.

JAYNES, E. T. (1957). *Phys. Rev.* **106**, 620-630.

JAYNES, E. T. (1968). *IEEE Trans. Syst. Sci. Cybern.* **SSC4**, 227-241.

LADD, M. F. C. & PALMER, R. A. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER. New York: Plenum.

MARVIN, D. A., BRYAN, R. K. & NAVE, C. (1987). *J. Mol. Biol.* **193**, 315-343.

MILLER, M., HARRISON, R., APPELLA, E., WLODAWER, A. & SUSSMAN, J. L. (1988). *Nature (London)*, **334**, 85-86.

MILLER, M., WLODAWER, A., APPELLA, E. & SUSSMAN, J. L. (1987). *J. Mol. Biol.* **195**, 967-968.

NARAYAN, R. & NITYANANDA, R. (1982). *Acta Cryst.* **A38**, 122-128.

NAVAZA, J. (1986). *Acta Cryst.* **A42**, 212-223.

PRESS, W. H., FLANNERY, B. P., TEUKOLSKY, S. A. & VETTERLING, W. T. (1986). *Numerical Recipes. The Art of Scientific Computing.* Cambridge Univ. Press.

ROGERS, D. (1980). In *Theory and Practice of Direct Methods in Crystallography*, edited by M. F. C. LADD & R. A. PALMER. New York: Plenum.

SCALES, L. E. (1985). *Introduction to Non-Linear Optimization.* New York: Springer-Verlag.

SHANNON, E. C. (1948). *Bell Syst. Tech. J.* **27**, 379-423.

SHORE, J. E. & JOHNSON, R. W. (1980). *IEEE Trans. Inf. Theory*, **IT-26**, 26-37.

SHORE, J. E. & JOHNSON, R. W. (1981). *IEEE Trans. Inf. Theory*, **IT-27**, 472-482.

TOLSTOV, G. P. (1976). *Fourier Series*. New York: Dover Publications.
WILKINS, S. W. & STUART, D. (1986). *Acta Cryst.* A**42**, 197-202.
WILKINS, S. W., VARGHESE, J. N. & LEHMANN, M. S. (1983). *Acta Cryst.* A**39**, 47-60, 594.
WING, R. M., DREW, H., TAKANO, T., BROKA, C., TANAKA, S., ITAKURA, K. & DICKERSON, R. E. (1980). *Nature (London)*, **287**, 755-758.
YU, F. T. S. (1976). *Optics and Information Theory*. New York: John Wiley.

Acta Cryst. (1989). A**45**, 10-20

Infinite Periodic Minimal Surfaces and their Crystallography in the Hyperbolic Plane

BY J.-F. SADOC AND J. CHARVOLIN

Laboratoire de Physique des Solides, Bâtiment 510, *Université Paris-Sud*, 91405 *Orsay, France*

(*Received* 25 *February* 1988; *accepted* 4 *July* 1988)

Abstract

Infinite periodic minimal surfaces are now being introduced to describe some complex structures with large cells, formed by inorganic and organic materials, which can be considered as crystals of surfaces or films. Among them are the spectacular cubic crystalline structures built by amphiphilic molecules in the presence of water. The crystallographic properties of these surfaces are studied from an intrinsic point of view, using operations of groups of symmetry defined by displacements on their surface. This approach takes advantage of the relation existing between these groups and those characterizing the tilings of the hyperbolic plane. First, the general bases of the particular crystallography of the hyperbolic plane are presented. Then the translation subgroups of the hyperbolic plane are determined in one particular case, that of the tiling involved in the problem of cubic structures of liquid crystals. Finally, it is shown that the infinite periodic minimal surfaces used to describe these structures can be obtained from the hyperbolic plane when some translations are forced to identity. This is indeed formally analogous to the simple process of transformation of a Euclidean plane into a cylinder, when a translation of the plane is forced to identity by rolling the plane onto itself. Thus, this approach transforms the 3D problem of infinite periodic minimal surfaces into a 2D problem and, although the latter is to be treated in a non-Euclidean space, provides a relatively simple formalism for the investigation of infinite periodic surfaces in general and the study of the geometrical transformations relating them.

I. Introduction

Recent studies of some 3D crystalline structures with large cells have pointed out the limitation of the classical aspect of crystallography, as concerned with the study of periodic organizations of topologically zero-dimensional objects such as atoms and molecules, and called for the introduction of more operative concepts, permitting analysis of them as periodic organizations of two-dimensional objects such as surfaces and films. Such structures are often observed in liquid crystals - the 'bicontinuous' cubic phases of lyotropics, the D phases and 'blue' phases of thermotropic smectics and cholesterics - but also in some biological and inorganic materials. The need for new terms to describe them was advocated in some recent papers (Scriven, 1976, 1977; Hyde & Andersson, 1984; Mackay, 1985; Mackay & Klinowski, 1986; Sadoc & Charvolin, 1986). Among these structures we are particularly interested in liquid crystalline ones, formed by amphiphilic molecules in the presence of water, which can be described as periodic entanglements of two fluids media separated by interfaces organized in a symmetric film exhibiting a very rich polymorphism (Luzzati, 1968; Ekwall, 1975). We have recently demonstrated that, in the case of the 'bicontinous' cubic structures of these materials, the film built by the interfaces is supported by surfaces directly related to the F, P and G infinite periodic minimal surfaces (or IPMS) of the mathematicians (Charvolin & Sadoc, 1987). These surfaces can be described as periodic non-intersecting surfaces with zero mean curvature separating space in two identical labyrinths. Thus, the above structures are interesting not only on purely physicochemical grounds but, also, as actual structures modelling surfaces of great mathematical interest.

Our approach to the polymorphism of the structures formed by amphiphilic molecules is presented in Sadoc & Charvolin (1986), and its application to the case of 'bicontinuous' cubic structures is developed in Charvolin & Sadoc (1987). It is based upon the idea that a geometrical frustration, related to local interactions of the molecules and packing constraints, takes place within the film. This frustration is relaxed if the film is transferred into the 3D space with positive

0108-7673/89/010010-11$03.00