MOREY, A. V. & JUNI, E. (1970). *Methods Enzymol.* **18A**, 238–245.
Picker FACS-1 Disk Operating System (1972). Picker Corporation, Cleveland, Ohio.
PLETCHER, J. & SAX, M. (1966). *Science*, **154**, 1331–1333.
PLETCHER, J. & SAX, M. (1972). *J. Amer. Chem. Soc.* **94**, 3998–4005.
PLETCHER, J., SAX, M., SENGUPTA, S., CHU, J. & YOO, C. S. (1972). *Acta Cryst.* B**28**, 2928–2935.
PLETCHER, J., SAX, M., YOO, C. S., CHU, J. & POWER, L. (1974). *Acta Cryst.* B**30**, 496–503.
POWER, L., PLETCHER, J. & SAX, M. (1970). *Acta Cryst.* B**26**, 143–148.
PULSINELLI, P. D. (1970). Thesis, *Conformational and Stereoelectronic Aspects of the Thiamine Catalysis Mechanism: The Crystal and Molecular Structures of Two 'Active Acetaldehyde'–Thiamine Pyrophosphate Analogues*, pp. 56–59, Univ. of Pittsburgh, Pittsburgh, Pa.
RICHARDSON, M. F., FRANKLIN, K. & THOMPSON, D. M. (1975). *J. Amer. Chem. Soc.* **97**, 3204–3209.
SAX, M., PULSINELLI, P. & PLETCHER, J. (1974). *J. Amer. Chem. Soc.* **96**, 155–165.
SCHELLENBERGER, A. (1967). *Angew. Chem. Int. Ed.* **6**, 1024–1034.
SCHELLENBERGER, A. & HÜBNER, G. (1967). *Z. physiol. Chem.* **348**, 491.
SHIONO, R. (1971). *Crystallographic Computing Programs for the IBM* 1130 *System*. Tech. Rep. No. 49, Department of Crystallography, Univ. of Pittsburgh, Pittsburgh, Pa.
WHITAKER, A. & JEFFREY, J. W. (1970). *Acta Cryst.* B**26**, 1429–1440.
WITTORF, J. H. & GUBLER, C. J. (1970). *Eur. J. Biochem.* **14**, 53–60.
ZIEVE, L. (1969). *Ann. N. Y. Acad. Sci.* **162**, 732–743.
ZIEVE, L., DOIZAKI, W. M. & STENROOS, L. E. (1968). *J. Lab. Clin. Med.* **72**, 268–277.

Acta Cryst. (1976). B**32**, 2975

A Method to Determine Heavy-Atom Positions for Virus Structures

BY PATRICK ARGOS AND MICHAEL G. ROSSMANN

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, *U.S.A.*

(*Received* 6 *January* 1976; *accepted* 10 *May* 1976)

Small spherical virus particles consist of multiples of 60 chemically identical protein subunits related by icosahedral symmetry. If the orientation of the virus has been previously established, then Patterson self-vectors can be calculated between heavy atoms attached specifically to equivalent points in each subunit. A systematic search of all reasonable positions within the subunit and a comparison with the known difference Patterson map would thus establish heavy-atom positions relative to the particle center. This method was used successfully on two heavy-atom derivatives of satellite tobacco necrosis virus [Lentz, Strandberg, Unge, Vaara, Borell, Fridborg & Petef, *Acta Cryst.* (1976), B**32**, 2979–2983].

Introduction

Low-resolution X-ray diffraction data are now available for a number of small spherical viruses, such as turnip yellow mosaic virus (Klug, Longley & Leberman, 1966), tomato bushy stunt virus (TBSV; Harrison, 1971), satellite tobacco necrosis virus (STNV; Åkervall *et al.*, 1971), and southern bean mosaic virus (SBMV; Johnson, Rossmann, Smiley & Wagner, 1974). Data extension to atomic resolution is presently in progress. The solution of the phase problem will depend upon the molecular replacement technique (Rossmann, 1972; Bricogne, 1974; Argos, Ford & Rossmann, 1975), the isomorphous replacement technique, or both. Heavy-atom derivatives have been prepared and analyzed for TBSV (Harrison & Jack, 1975), STNV (Lentz *et al.*, 1976), and SBMV (Wagner, 1974).

The presence of several heavy-atom sites within the crystallographic asymmetric unit has, in general, been a limitation to the isomorphous replacement technique. The interpretation of the TBSV heavy-atom difference Patterson synthesis was a complex task (Harrison, 1971; Harrison & Jack, 1975). Yet when heavy-atom sites are related by non-crystallographic symmetry as in virus structures, the problems of interpreting difference Patterson syntheses can be lessened as only those sites in the non-crystallographic asymmetric unit need be determined. Argos & Rossmann (1974) have shown the feasibility of solving systematically a difference Patterson synthesis for a tetrameric protein with 222 molecular symmetry. Their methods have now been extended for use in virus structures with 532 symmetry in the asymmetric unit, resulting in 60 non-crystallographic symmetry-related subunits in the crystal unit cell. Such viruses have $T=1$ symmetry (Caspar & Klug, 1962), the simplest of all situations. One virus with such simplicity is STNV and the subsequent paper (Lentz *et al.*, 1976) describes the success of the method for the STNV data.

The search procedure

This paper discusses a procedure to determine heavy-atom sites relative to the particle center, but not with respect to the cell origin. The method corresponds to case (*a*) discussed by Argos & Rossmann (1974). Procedures such as those discussed by Tollin (1966), Crowther & Blow (1967) and Rossmann, Ford, Watson & Banaszak (1972) can be used to position the particle in the cell given the heavy-atom distribution relative to the particle. Alternatively the position of the virus particle within the cell is frequently known in advance from other sources such as packing considerations, low-resolution structure amplitudes, or space-group symmetry (Johnson *et al.*, 1974).

For a particle with 532 non-crystallographic symmetry, two basic problems arise in generating heavy-atom vectors and comparing them with a difference Patterson synthesis; these are not present in the simpler situation of molecular 222 symmetry (Argos & Rossmann, 1974). The first is the magnitude of the computing operation. One heavy atom per non-crystallographic asymmetric unit generates 7080 self-vectors for a $T=1$ virus as opposed to only 24 for a tetrameric molecule, given the four particles per cell as in the STNV case. There are also 21600 cross-vectors between heavy-atom sites in different particles. The cross-vectors have been considered here to provide an essentially smooth carpet on which the self-vectors are erected. The second problem is created by the absence of orthogonality between some of the symmetry axes. While Cartesian coordinates are suitable for considering the simpler 222 symmetry, polar coordinates are more amenable for treating icosahedral symmetry.

The procedure here follows that of Argos & Rossmann (1974). An atom is first considered at an assumed position within the icosahedral non-crystallographic asymmetric unit. The position is then successively multiplied by rotations about a three-, a two-, a five-, and again a twofold axis (Table 1). The sequence of selection of these axes and their relative orientation is critical if all sixty independent sites are to be generated. The first three axes need to be adjacent axes on a great circle while the final twofold axis is perpendicular to the plane of this circle. The position of the initial atom is expressed both in terms of spherical coordinates using the conventions of Rossmann & Blow (1962) and in terms of Cartesian coordinates defined by an orthogonal 222 set with respect to the icosahedral virus particle. The rotation matrices about each axis can be generated with respect to the icosahedral axes (*International Tables for X-ray Crystallography*, 1967) and then referred to the cell (Argos & Rossmann, 1974). Alternatively the orientation of the four axes can be directly stated in terms of polar coordinates with respect to the cell and then transformed into rotation matrices [see Table 1(*b*) of Rossmann & Blow (1962)]. The latter approach is more directly coupled with results from the rotation function.

Table 1. *Polar coordinates of STNV axes used in generating equivalent icosahedral positions and to define the asymmetric unit*

Definitions of ψ and φ are given in Fig. 1 and correspond to the convention of Rossmann & Blow (1962). The axial orientations for STNV were determined by Rossmann, Åkervall, Lentz & Strandberg (1973).

Rotation axes	Order used in generating 60 equivalent positions	Axes used to define vertices of asymmetric spherical triangle	ψ	φ
3	1	√	65·7°	15·0°
2	2	√	45·0	15·0
5	3		103·3	15·0
2	4		90·0	105·0
5		√	53·02	56·15

When the Patterson vectors between the 60 generated sites have been computed, they are sorted and allocated to the nearest grid points in the asymmetric unit of the Patterson synthesis. The value of the Patterson synthesis P_i was taken only once at each of the N grid points where one or more vectors appeared. The search criterion C was then taken as

$$C=\sum_{i=1}^{N} P_i - NK \qquad (1)$$

where K is a constant. The latter accounts for the arbitrary zero level within the Patterson function caused by the omission of an $F(000)$ term and in particular for the carpet (or background) of cross-vectors. In (1) it has been assumed that all vectors near a given grid point are coincident. This avoids tedious multiplicity corrections (Argos & Rossmann, 1974) and takes into consideration (at least approximately) the large number of cross-vectors.

The value of the search criterion is dependent upon the grid interval on which the Patterson synthesis is sampled. Coarse grids will result in smaller values of N while fine grids will cause the search criterion to approach the reciprocal-space procedure (see Appendix). In practice, it has been found that the appearance of the function is not altered significantly by changing the size of the grid, provided that the grid interval is smaller than one-third of the resolution of the data used to compute the Patterson synthesis. A starting choice for K should be near the average value of the Patterson synthesis, excluding the origin peak. A poor choice will generate trenches or ridges along symmetry planes in the criterion function as there is a greater coincidence of peaks at grid points for a heavy atom on or near a special position.

Computing time can be lessened in the initial generation of the 60 sites by expressing these in terms of their polar coordinates r,ψ,φ. This allows positions along a radial line with constant ψ and φ values to be tested systematically merely by altering the length of the sorted Patterson vectors in proportion to r. How-

ever, the transformation of the isotropically expanded (or contracted) vectors into the asymmetric unit of Patterson space is still unique for each heavy-atom position along the radial line.

Care must be taken with vectors within the origin peak. Three options are available in dealing with these vectors: the origin can be subtracted; special consideration can be given to those vectors within a fixed radius of the origin (Argos & Rossmann, 1974); or those vectors on grid points with Patterson values above a fixed height can be omitted. The last option was used for the STNV case (Lentz *et al.*, 1976).

Limits of asymmetric unit of search function

If the symmetry axes of an icosahedron are plotted on a stereographic projection, an asymmetric unit can be delineated by a spherical triangle with two fivefold axes and one threefold axis at its vertices (Fig. 1). However, since the Patterson vector distribution contains a center of symmetry, the asymmetric unit of the search function will only be half the true asymmetric unit of an icosahedron, delineated by a Napierian triangle with five-, three-, and twofold axes at its vertices. The planes of the great circles whose intersections form the vertices of the Napierian triangle are thus mirror planes in the search function.

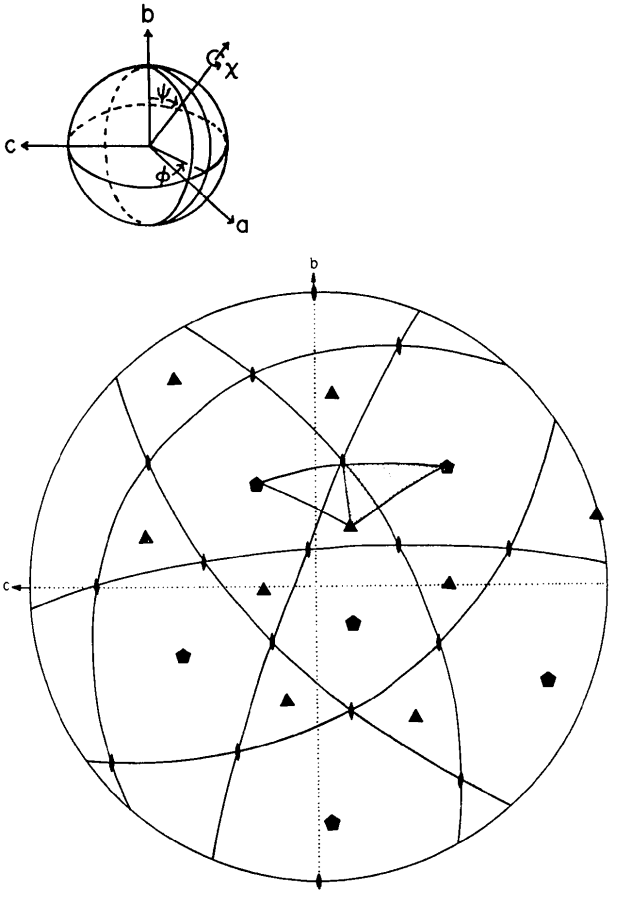

Fig. 1. Distribution of symmetry elements for one STNV particle in its monoclinic cell. The chosen search asymmetric unit has been shaded. Insert shows ψ and φ angles with respect to the selected orthogonal crystal system following the convention of Rossmann & Blow (1962).

A grid of possible heavy-atom positions within the asymmetric Napierian triangle was generated in the following manner. The ψ,φ coordinates of the rotation axes defining the selected spherical triangle (Fig. 1 and Table 1) were converted to Cartesian coordinates for a given value of the radius, r. The distance between the three- and twofold axes was divided into an integral number of intervals which were less than or equal to the desired grid spacing. The same number of intervals was used to divide the threefold and fivefold distance. This procedure provided corresponding markers along the two lines. The various distances between the markers were also divided into an integral number of spacings less than the desired grid interval. The Cartesian coordinates of the grid points were then re-converted to polar coordinates, thus permitting the sampling of points on a sphere with constant radius, rather than on a plane. In order to maintain the computational saving of the isotropic vector expansion, a grid at the mean radius of the protein coat can be chosen.

Discussion

The self-vector search procedure was successfully applied first to various model data sets and then to the heavy-atom difference Patterson syntheses of the STNV $T=1$ virus, as described in Lentz *et al.* (1976). Only one heavy-atom site was found in the icosahedral asymmetric unit for a platinum and a mercury derivative of STNV. The difference Patterson maps were calculated from 10 Å resolution data on an approximate 3 Å grid. The positions determined from the search criterion were confirmed through cross difference Fourier syntheses and by the molecular replacement technique (Rossmann, 1972).

The present search procedure only considers self-vectors between chemically identical sites in each protein subunit. However, for a multiple-site derivative (*e.g.* a $T=3$ virus or multiply substituted $T=1$ virus) self-vectors between chemically non-identical heavy atoms would result in additional difference Patterson peaks erected on the carpet of cross-vectors. This further background could lessen the clarity of the search results; yet the magnitude of the computing operation to determine and search all multiple-site self-vectors would be prohibitive. However, the recognition of even only one heavy-atom site, which is likely to be that with greatest occupancy, could be sufficient to verify or find any other major or minor sites. For instance, these might be related by quasi-symmetry which cannot be checked precisely in the vector search method. Alternatively the postulated heavy-atom positions, along with the particle-center position, can be

refined and used to compute a difference electron density map to determine minor sites (Rossmann, 1976; Lentz *et al.*, 1976). In any event the present results show that it is quite satisfactory to treat the cross-vectors as a roughly constant carpet in the Patterson map by a suitable choice of the constant K in equation (1). Further improvement might also be made by using higher-resolution data.

Prior averaging of the original difference Patterson synthesis among the non-crystallographic icosahedral asymmetric units could also aid in the solution of a particularly intransigent problem. This process would enhance all the self-vectors associated with a particular virus particle, removing both cross-vectors and the self-Patterson vectors of other crystallographically related particles. This procedure would be advantageous when cross-vectors dominate the self-Patterson map. On the other hand, if the particle center is known and the majority of vectors can be computed, an initial Patterson averaging procedure could be more harmful than helpful.

Gilbert & Klug (1974) note that if there are non-crystallographically related heavy atoms lying on the locus of a circle, piling up of vectors will occur in the Patterson map at a radius equal to the diameter of the circle in real space. Similarly, for a spherical virus with equivalently substituted heavy atoms in each protein subunit, Patterson space will contain a spherical shell of higher density at a radius equal to the diameter of the heavy-atom sphere in the real virus. Prior icosahedral averaging of the Patterson synthesis would remove cross-vectors between different crystallographically oriented viruses and between sets of different sites within one virus (which would appear as spheres centered at different origins). Summation of the Patterson values in a series of shells would produce large sums at those radii where there is a heavy-atom substitution. This procedure would eliminate the need to search at radii other than those near the heavy-atom radius, although a complete radial search must be more sensitive than the spherical averaging procedure. However, when there is more than one substitution per protein subunit, an initial spherical analysis might give useful information.

In conclusion it is apparent that the utilization of abundant non-crystallographic symmetry can provide a systematic, rigorous and fast procedure for the determination of heavy-atom positions in a virus crystal structure. Computer programs and associated write-ups are available upon request from the authors.

We have greatly appreciated the inventive guidance of Dr Geoffrey Ford and the many helpful discussions with Drs Jack Johnson, Bror Strandberg and Paul Lentz. We thank Sharon Wilder for technical help in the preparation of the manuscript. The work was supported by National Institutes of Health grant No. AI 11219, National Science Foundation grant No. BMS74-23537, and a supply grant from Eli Lilly Co.

APPENDIX

The reciprocal-space approach to the vector search method

The derivation below provides an alternative analysis of the search procedure to give greater insight into the significance of the method.

Let F_o be the observed structure factors of the heavy-atom difference Patterson synthesis. Thus

$$F_o \simeq (|F_{VH}| - |F_V|)^2.$$

Let F_c be the corresponding calculated structure factors based on an assumed distribution of heavy atoms. A reasonable search function might be defined as

$$S(\psi, \varphi, R) = \sum_h F_o^2 F_c^2 . \qquad (A1)$$

It will have a maximum if there is good agreement between the observed and calculated structure factors. The search function, S, is dependent only on the spherical coordinates ψ, φ, R of the trial heavy atom in the non-crystallographic asymmetric unit. Now

$$F_c^2 = [\sum_{i=1}^{N} f \exp(2\pi i \mathbf{h} \,.\, \mathbf{x}_i)] \, [\sum_{j=1}^{N} f \exp(-2\pi i \mathbf{h} \,.\, \mathbf{x}_j)]$$

where N is the number of non-crystallographic asymmetric units in the unit cell, and f is the effective scattering factor of the heavy atom.

It follows that

$$F_c^2 = f^2[N + 2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \cos 2\pi \mathbf{h} \,.\, (\mathbf{x}_i - \mathbf{x}_j)] .$$

Hence from (*A*1)

$$S(\psi, \varphi, R) = N \sum_h f F_o^2 + 2 \sum_h F_o^2 f^2 \sum_{i=1}^{N} \sum_{j=i+1}^{N} \cos 2\pi \mathbf{h} \,.\, (\mathbf{x}_i - \mathbf{x}_j) .$$

By omitting the constant first term on the right-hand side and by changing the order of summation of the second term, a modified search function may be written as

$$S'(\psi, \varphi, R) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} [\sum_h F_o^2 f^2 \cos 2\pi \mathbf{h} \,.\, (\mathbf{x}_i - \mathbf{x}_j)] .$$

If point atoms are considered, then f is constant and

$$S'(\psi, \varphi, R) = \sum_{i=1}^{N} \sum_{j=i+1}^{N} [\sum_h F_o^2 \cos 2\pi \mathbf{h} \,.\, (\mathbf{x}_i - \mathbf{x}_j)] .$$

The term in the square brackets now represents the value of the observed Patterson map sampled at the end of the interatomic vectors $(\mathbf{x}_i - \mathbf{x}_j)$. Hence the search function S' can be seen as the sum of the Patterson values over all interatomic vectors. This shows the equivalence of the search concept in Patterson space with that in reciprocal space as defined in (*A*1).

There are, however, two differences between the Patterson and reciprocal-space results. The first is the treatment of multiple vectors at a single grid point of the Patterson map. The Patterson approach suggests that the value at each Patterson grid point should be used only once in the search function sum. The reciprocal lattice approach states that the Patterson map must be sampled at each vector regardless of other vectors in the same position. It will, however, be seen that as the chosen grid interval is reduced the Patterson and reciprocal-space criteria converge. The second difference is that the reciprocal-space search function in no way suggests a positive contribution from the unknown terms corresponding to the cross-vectors between virus particles. Differences between these criteria are to be expected as the chosen functions are themselves different.

References

ÅKERVALL, K., STRANDBERG, B., ROSSMANN, M. G., BENGTSSON, U., FRIDBORG, K., JOHANNISEN, H., KANNAN, K. K., LÖVGREN, S., PETEF, G., ÖBERG, B., EAKER, D., HJERTÉN, S., RYDÉN, L. & MORING, I. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 469–483.

ARGOS, P., FORD, G. C. & ROSSMANN, M. G. (1975). *Acta Cryst.* A**31**, 499–506.

ARGOS, P. & ROSSMANN, M. G. (1974). *Acta Cryst.* A**30**, 672–677.

BRICOGNE, G. (1974). *Acta Cryst.* A**30**, 395–405.

CASPAR, D. L. D. & KLUG, A. (1962). *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.

CROWTHER, R. A. & BLOW, D. M. (1967). *Acta Cryst.* **23**, 544–548.

GILBERT, P. F. C. & KLUG, A. (1974). *J. Mol. Biol.* **86**, 193–207.

HARRISON, S. C. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 495–501.

HARRISON, S. C. & JACK, A. (1975). *J. Mol. Biol.* **97**, 173–191.

International Tables for X-ray Crystallography (1967). Vol. II, 2nd ed., pp. 30–35. Birmingham: Kynoch Press.

JOHNSON, J. E., ROSSMANN, M. G., SMILEY, I. E. & WAGNER, M. A. (1974). *J. Ultrastruct. Res.* **46**, 441–451.

KLUG, A., LONGLEY, W. & LEBERMAN, R. (1966). *J. Mol. Biol.* **15**, 315–343.

LENTZ, P. J. JR, STRANDBERG, B., UNGE, T. VAARA, I., BORELL, A., FRIDBORG, K. & PETEF, G. (1976). *Acta Cryst.* B**32**, 2979–2983.

ROSSMANN, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon and Breach.

ROSSMANN, M. G. (1976). *Acta Cryst.* A**32**, 774–777.

ROSSMANN, M. G., ÅKERVALL, K., LENTZ, P. J. JR & STRANDBERG, B. (1973). *J. Mol. Biol.* **79**, 197–204.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.

ROSSMANN, M. G., FORD, G. C., WATSON, H. C. & BANASZAK, L. J. (1972). *J. Mol. Biol.* **64**, 237–249.

TOLLIN, P. (1966). *Acta Cryst.* **21**, 613–614.

WAGNER, M. A. (1974). Private communication.

Acta Cryst. (1976). B**32**, 2979

The Determination of the Heavy-Atom Substitution Sites in the Satellite Tobacco Necrosis Virus

BY PAUL J. LENTZ JR,* BROR STRANDBERG, TORSTEN UNGE, IVAR VAARA, AGNETA BORELL, KERSTIN FRIDBORG AND GEORGI PETEF

Wallenberg Laboratory, University of Uppsala, Box 562, *S*-751 22, *Uppsala, Sweden*

(*Received* 6 *January* 1976; *accepted* 19 *April* 1976)

The heavy-atom binding sites in two derivatives of crystalline satellite tobacco necrosis virus were found by a vector search method [Argos & Rossman, *Acta Cryst.* (1976), B**32**, 2975–2979]. For each derivative there is a single major site in each of the 60 icosahedrally related subunits in the virus coat. The derivative scale factor, the occupancy, and positional and thermal parameters of the sites were refined by structure factor calculations. Single isomorphous phases calculated from either refined derivative were sufficiently accurate to locate the substitution site in the difference Fourier map of the other derivative. The refinement statistics indicate that the major sites have been found.

Introduction

The satellite tobacco necrosis virus (STNV) is one of the smallest viruses known, so small that it lacks some functions necessary for replication and must co-infect with the larger, fully competent tobacco necrosis virus in order to propagate (Kassanis & Nixon, 1961). The virus particle is an aggregate of 60 copies of the single coat-protein molecule, M.W. 23 000, arranged in a $T=1$ icosahedral surface lattice (Åkervall *et al.*, 1971; Klug, 1971) and a single-stranded ribonucleic acid molecule, M.W. 400 000, which codes for that coat protein (Reichmann, 1964).

* Present address: Department of Chemistry, University of Michigan, Ann Arbor, MI 48109, U.S.A.