Multi-Solution Genetic Algorithm Approach to Surface Structure Determination Using Direct Methods

E. Landree,* C. Collazo-Davila and L. D. Marks

Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208, USA. E-mail: landree@merle.acns.nwu.edu

Abstract

We show that it is possible to use a multi-solution genetic algorithm search method utilizing direct methods to solve surface structures from surface diffraction data. We suggest that the method is generally applicable and able to replace random searches of the solution space.

1. Introduction

Over the last 20 years, a large number of different methods have been explored for phasing diffraction patterns *via* direct methods. Among these are the well known *MULTAN* (Germain & Woolfson, 1968), *RANTAN* (Yao, 1981), *MITHRIL* (Gilmore, 1984) and *SHELX* (Sheldrick, 1990) codes as well as more recent introductions such as *MICE* (Gilmore, Bricogne & Bannister 1990), which exploits a maximum entropy approach. In rather different areas the last few years have seen the popularization of approaches such as simulated annealing (Sheldrick, 1990; Bhat, 1990) and more recently genetic algorithms [GA's (Goldberg, 1989; Davis, 1987; Gutowski, 1994)] for global optimization problems, as well as the problem of solving atomic and molecular structures (Xiao & Williams, 1993; Chang & Lewis, 1994; Miller, Hogel & Filman, 1996). GA's are interesting since both conceptually and in terms of practical coding they are very simple to implement. In essence one allows a set of possible solutions to evolve using an artificial natural selection to yield close to optimized solutions after a number of generations.

An area which we have recently been exploring is the viability of direct methods for two-dimensional surface diffraction data, either transmission electron diffraction (TED) or grazing incidence X-ray diffraction [GIXD (Collazo-Davila, Marks, Nishii & Tanishiro, 1997; Gilmore, Marks, Grozea, Collazo, Landree & Twesten, 1997; Marks, Plass & Dorset, 1997; Landree, Marks, Zschack & Gilmore, 1997)]. There is good evidence that transmission electron diffraction from surfaces is almost kinematical, with systematic errors due to dynamical effects smaller than those associated with the phase reconstruction (Tanishiro & Takayanagi, 1989; Marks, Savage, Zhang & Ai 1991). For ideal noise-free data one can formulate various figures of merit (FOM's) and then seek to optimize these, confident that the best FOM value is the correct solution. (Hereafter we will always assume that the best FOM value is zero, consistent with minimization notation.) However, with surface diffraction data noise must be included, and even with the best collection methods to date the error in the strongest beams is at least 5%, ranging to >100% for the weaker intensities (2 to perhaps 3 orders of magnitude smaller) due to both Poisson statistics and background noise. Furthermore, when surface transmission electron diffraction spots and bulk diffraction peaks overlap it is difficult, if not impossible, to extract the much weaker surface data, resulting in incomplete data sets. Given these limitations, the best that can be achieved is a FOM that gives a viable solution. A small FOM should be a necessary condition for the correct solution, but may not be sufficient. One therefore wants to find all possible minima within the multi-dimensional solution hyperspace and then, based upon chemical and χ^2 refinement, decide which of these is the correct solution.

The focus of this paper is to explore the viability of a multi-solution genetic algorithm for phasing, particularly for the noisy and often incomplete data available from surfaces using either transmission electron or X-ray diffraction.

2. Genetic algorithm as a search algorithm

While the basic structure of GA's is well established (Holland, 1975; Davis, 1987; Goldberg, 1989; Koza, 1992), a number of points about how we have implemented it for multi-solution phasing problems merit description. In the simplest case we start with a set of structure factors S, then encode the phases of a subset s, consisting of the strongest structure factors in most cases, into a set of bits (gene) using conventional binary coding or Gray codes. Each gene would then represent the phase for a single reflection. The number of phases controlled by the GA, s, is situation dependent, however, typical values range from 10 to 20% of the total number of reflections. This digitization need not be too precise and in general steps of 2–5° for general reflections is adequate, *i.e.* 6–7 bits. For reflections which are either 180 or 360°, a single bit is assigned. All

the genes are stored end-to-end as a single string of bits, otherwise known as a chromosome. Starting with a population N of chromosomes, we calculate the remaining $(S - s)$ phases and evaluate the FOM. Then based upon some 'natural selection' criteria (which favors good FOM values), choose 'parents' for the next generation. From pairs of parents, 'children' are produced by crosslinking, *i.e.* a location along the parent chromosomes is selected at random and the bits on one side of the location are interchanged. Some level of 'mutation' is then introduced into each new population by randomly switching some of the bit values from 0 to 1 or *vice versa*. The FOM's for these children are then evaluated and the process iterated with these being the new parents for the next generation. In addition to this basic structure various modifications of the process can be incorporated, some of which are mentioned below (see also the *Discussion*).

What makes a GA powerful as a global search method is how it handles schemata (Holland, 1968, 1975). A schemata is a similarity template describing a subset of chromosomes with similarities at certain positions, similar to phase combinations for different structure factors. For the simple case where the number of parents and children are the same, even though only N new FOM values are calculated for each generation, the algorithm effectively processes on the order of N^3 schemata (Goldberg, 1989). Crossover disrupts these schemata if it occurs within their length, so shorter schemata are better preserved across generations; short, effective schemata increase roughly exponentially with the number of generations used. Since favorable phase combinations are the target of a direct methods analysis, GA's provide an almost ideal search process.

How crosslinking is performed with respect to the order of phases can be important. The standard method is to crosslink randomly along the chromosome, disrupting longer schemata and disturbing individual phases. Forcing crosslinking to occur only between the genes was considered and appeared to have a small favorable influence (Fig. 1). Using an order to maximize linkage between phases will improve the convergence, although in practice ordering in terms of absolute structure factor, unitary values or using strong low-order reflections (to enforce phasing out to higher angles) were all rather similar.

There are some modifications to classical GA's which are appropriate for direct methods. No FOM is a perfect measure, so there is no point in refining to a high degree of accuracy. More significant is to be able to achieve some ranking of a set of plausible solutions, then search among them for one which corresponds to realistic atom positions. While GA's are better suited toward exploring than refining specific solutions, one needs to avoid strategies which reinforce refinement. One modification involved how parents are chosen. Typically a 'roulette wheel' method is used, where the probability is

proportional to the value of the FOM. Instead we implemented a simple ranking scheme (Baker, 1985) weighted by

$$\text{num} = [\exp(-1.0.\text{ran1}.\mathbf{T}) - \exp(-1.0.\mathbf{T})]/ [1.0 - \exp(-1.0.\mathbf{T})], \tag{1}$$

where $\mathbf{T}$ is a variable controlling how strongly to favor lower values and num is the weighted number between 0 and 1. For all cases herein $\mathbf{T} = 1$ was used. By increasing the value of $\mathbf{T}$, lower numbers become more strongly weighted. The value of num is then combined with

$$p = 1 + \text{integer}(\text{num}.N) \tag{2}$$

to choose the pth chromosome as a parent. The second change to prevent convergence of the algorithm to only a single solution (enforcing a multi-solution search) was to employ a 'uniqueness' operation, removing from the pool of possible parents at each generation copies of any given chromosome.

Another modification which was introduced to enhance convergence was to maintain a running set of the best solutions at the end of each iteration. Then some of these solutions were included in the pool for the N parents in every other generation (elitism).

Three fundamental control parameters are the population size, number of children and mutation rate. Population size has been discussed in the literature (Grefenstette, 1986; Schaffer, Caruana, Eshelman & Das, 1989; Davis, 1989; De Jong & Spears, 1990; Nakano, Davidor & Yamada, 1994) with little definitive conclu-
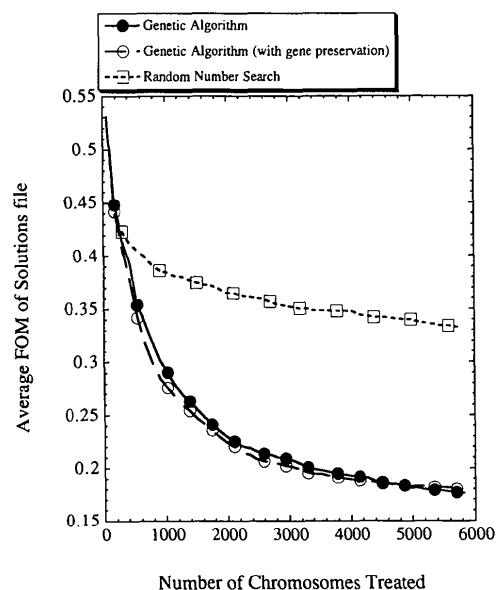

Fig. 1. Comparison of the efficiency of the GA for a noncentrosymmetric structure when crossover is performed anywhere along the chromosome (●) and breaking chromosomes only in between individual genes (○) against a standard random number search.

sion. One needs a balance between schemata exploration which favors larger sizes and schemata reinforcement which will favor smaller numbers. We found (see below) that modest population sizes 1–5 times the number of bits worked well. We also did not use the simple methodology of producing N children from N parents. Instead the number of children was treated as a separate input parameter, choosing only the top N children to act as parents for the next generation. The optimal mutation rates were typically low, however, this again varies. Each of these parameters are so problem dependent that optimization for any given class of problems is required.

A final point is the effect of the size of the hyperspace explored on the algorithm. By reducing the number of beams allowed to vary, either by origin definition or inequality relations between phases, the effective size of the hyperspace can be decreased, thereby increasing performance. In the noncentrosymmetric case, limiting the phase of a strong beam to a smaller subset of possible phases effectively defined a unique origin, which was sufficient for the solutions to converge. For the centrosymmetric model, an origin was defined by assigning specific phases to two beams, which remained unchanged throughout the calculation. In this case it was sometimes necessary to include additional phases to achieve a realistic solution. Such can also be the case when several of the strong beams are absent from the data set.

3. Numerical method, figures of merit and noise

The evaluation function is based upon a Sayre–Tangent formula (Woolfson & Hai-fu, 1995; Woolfson, 1991) with unitary structure factors in two dimensions. For a set of unitary structure factors, $U(\mathbf{g})$, with $\mathbf{g}$ a reciprocal lattice vector, it is possible to write

$$U_m(\mathbf{g}) = \sum_{\mathbf{h}} U_{m-1}(\mathbf{g} - \mathbf{h})U_{m-1}(\mathbf{h}) \qquad (3)$$

to determine an estimate of the mth structure factor from the $m - 1$ estimate, the latter corrected to have the experimental moduli. Keeping fixed those phases assigned by origin definition or phase inequality relationships, and allowing the rest to vary according to the Tangent formula, at each stage the FOM is calculated based upon

$$R^n = \{\Sigma'|[F_{m-1}(\mathbf{g}) - \alpha F_m(\mathbf{g})]/\sigma(\mathbf{g})|^n\}/$$
$$\Sigma'|[F_{m-1}(\mathbf{g})]/\sigma(\mathbf{g})|^n, \qquad (4)$$

where Σ' refers to a summation which includes all terms except the $F(0)$ term; $F_{m-1}(\mathbf{g})$ is the (known) amplitude and (variable) phase of the structure factor at the beginning of each cycle for diffraction spot $\mathbf{g}$; $F_m(\mathbf{g})$ is the amplitude and new (variable) phase calculated at the end of the cycle; α is a scalar chosen to give the smallest

Table 1. *Atomic positions for Model 1 and Model 2*

In both cases the structures consisted of only a single atom species, silicon. The a and b spacing are given along with the corresponding angle, γ. The values of x and y correspond to the atomic positions along a and b.

Model 1: noncentrosymmetric trimer model: $a = 11.52$, $b = 11.52$ Å, $\gamma = 120°$

x	y
0.17	0.0
0.0	0.17
−0.17	−0.17

Model 2: centrosymmetric 4 × 2 model: $a = 7.68$, $b = 15.36$ Å, $\gamma = 90°$

0.19	0.5
0.81	0.5
0.29	0.5
0.71	0.5
0.5	0.15
0.5	0.85
0.5	0.28
0.5	0.72
0.0	0.117
0.0	0.307
0.0	0.693
0.0	0.883

value of R; $\sigma(\mathbf{g})$, when used, is the error; n can be either 1 [similar to the Karle FOM (Woolfson, 1987)] or 2 (similar to a least-squares analysis). The process of calculating phases and determining the FOM is iterated until all the missing phases are defined and then continues refining the phases so long as the value of R reduces. Equation (4) operates as a penalty function for constraining possible solutions. Note that the above uses the complex structure factors not the moduli. While the modulus form of the equation is more commonly used (Woolfson & Hai-fu, 1995), it did not give as good results and will not be discussed further.

One important point concerns the mean value $F(0)$ and the variable α. For surface data it is often difficult to impossible to approximate $F(0)$ from a priori data since the unit-cell contents are rarely well known. (In the case of an adsorbate on a surface, one may have a good idea of the number of adsorbed atoms, but no information on how many relaxed substrate atoms contribute to the surface scattering.) Empirically, we have found that a normalization such that

$$F(0) = 1; \Sigma'F(\mathbf{g})^2 = 1 \qquad (5)$$

is satisfactory. Allowing α to be a free variable compensates for some of the uncertainty in $F(0)$. For the test cases considered herein a better approach was to use the more conventional form, dividing $F(\mathbf{g})$ by an effective atomic scattering factor

$$f(\mathbf{g}) = (\Sigma f_i(\mathbf{g})^2)^{1/2}, \qquad (6)$$

then (for the normalized scattering factors) using the

standard normalization $< E^2 > = 1$ and

$$E(0) = \{[\Sigma f_i(0)]^2 / \Sigma f_i(0)^2\}^{1/2} \qquad (7)$$

with

$$U(\mathbf{g}) = E(\mathbf{g})/E(0) \qquad (8)$$

and the sums in (6) and (7) taken over atoms i.

To quantify the results, we used two structures consisting of the atomic positions shown in Table 1, which will be referred to later as Model 1 and Model 2.

To represent experimental errors within each of the models, we added a random term proportional to

$$\eta(\mathbf{g}) = \beta[r^*(F(\mathbf{g}))] + \gamma, \qquad (9)$$

where r^* is a random number between $-\infty$ and $+\infty$, with a Gaussian distribution having a standard deviation
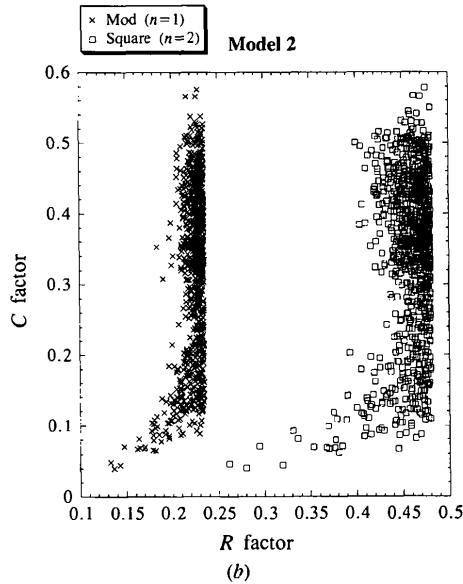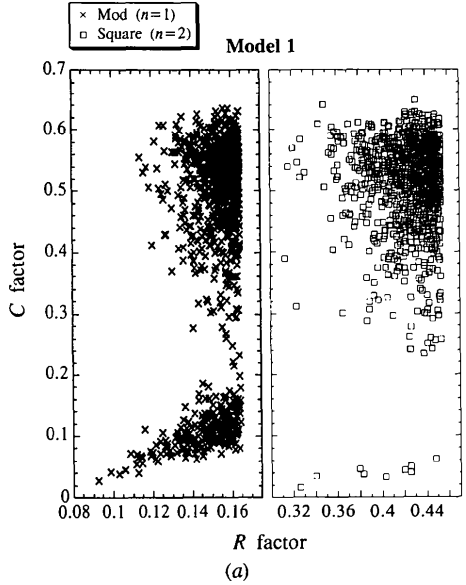

Fig. 2. (a) Comparison of the C factor *versus* FOM for the noncentrosymmetric model (Model 1) with both the $n = 2$ (square) and the $n = 1$ (mod) case. (b) A similar comparison for a centrosymmetric model (Model 2).
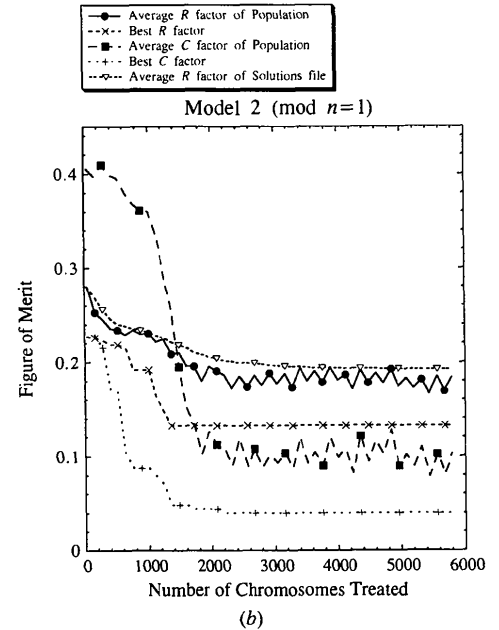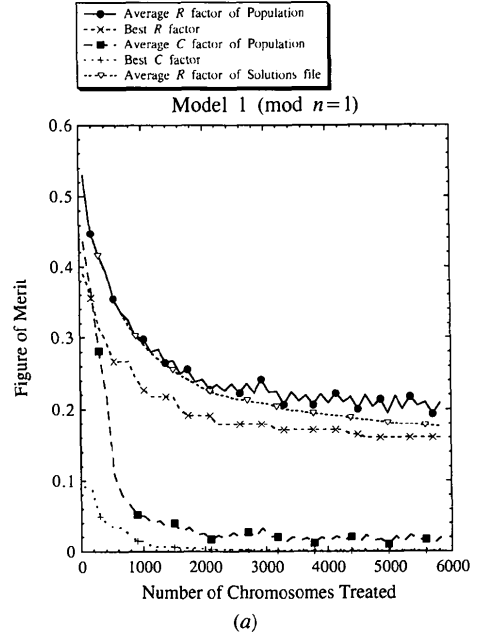

Fig. 3. Evolution of the average R factor for a given population, the best R factor, the average C factor of the population, the best C factor and the average R factor for the solutions file calculated at the end of every iteration for (a) the noncentrosymmetric model and (b) the centrosymmetric model as a function of the number of chromosomes treated.

of 1. β scales the first term in (9) to be roughly 10% for the average of the three strongest beams and γ is set to be 1% of the average of the three strongest beams. The result is a final data set whose errors ranged from roughly 10% for the strong beams to >100% for the weaker beams, modeling a Gaussian error distribution with a constant background.

The diffraction data for both models are calculated from a 2-D arrangement of Si atoms with beams up to 1 Å resolution. Model 1 has $p31m$ symmetry and a total of 49 beams. An origin can be defined by restricting the range of the (4,3) beam to within 60 and 180° and establishing an enantiomorph by limiting the range of the (2,2) beam to values between 0 and 180°. For Model 2, $p2mm$ symmetry, a total of 105 beams were used and the (0,5) and (3,5) beams were set at 180° to define a unique origin. In addition, the (4,0) and (6,0) beams were set at 360° based upon Σ_1 relationships. We then used a consistency check given by

$$C = \Sigma'|F_e(\mathbf{g})|\{1 - \cos(\theta_c - \theta_t)\}/2\Sigma'|F_e(\mathbf{g})|, \quad (10)$$

where F_e is the experimentally measured structure factor, θ_c is the calculated phase and θ_t the true phase for every beam. The correctness value (C factor) was calculated for every chromosome in a given population and provided a means of monitoring the progress of the code during a calculation.

4. Calibration of the FOM's

Representative examples of the behavior of the FOM's *versus* the consistency of the solutions (C factor) are shown in Fig. 2. Choosing $n = 1$ rather than $n = 2$ in (4) when errors are not included was as good as or superior in both cases. While the square ($n = 2$) did succeed in isolating the correct solution for the centrosymmetric model (Model 2), it was unsuccessful at finding the correct solution in the noncentrosymmetric case. In both cases when errors were included in (4), the performance

of the algorithm for $n = 1$ and $n = 2$ decreased. This is a useful general result, indicating that the errors in the direct method algorithm itself are larger than the measurement errors for typical surface diffraction data. Hereafter, we will report results only for the $n = 1$ case.

5. Performance of the algorithm

Fig. 3 illustrates the behavior of the individual FOM's as a function of the number of populations treated. Several permutations of the mutation rate, population size and number of children were explored to determine the optimal parameters. While the exact values were found to be problem-dependent, trends were observed which suggested a useful set of starting values to use. A parent-to-child ratio of 1:2 or 1:3 appeared best, with typical values being between 60 parents to 120 children and 80 parents to 240 children, roughly of the order of the number of bits per chromosome. Mutating roughly every fifth gene appeared to provide the ideal amount of random switching without hampering the convergence of the solution set for centrosymmetric structures. For centrosymmetric data sets, this equated to every fifth phase switching its phase 180°; for noncentrosymmetric structures, roughly five times the number of bits used to define a single gene. (A mutation rate of X is defined as switching the bit value of one in every X bits for an entire population.)

6. Determination of multiple solutions

In order to check that the algorithm would find multiple solutions, a test of Model 1 without restricting the (2,2) beam was performed. Figs. 4(a) and 4(b) are images of two of the solutions for $p31m$ symmetry from the final solution set. Both solutions have essentially the same FOM and, except for the inversion, are identical with minor differences in the background. Comparing the solutions calculated by the GA without limits on the
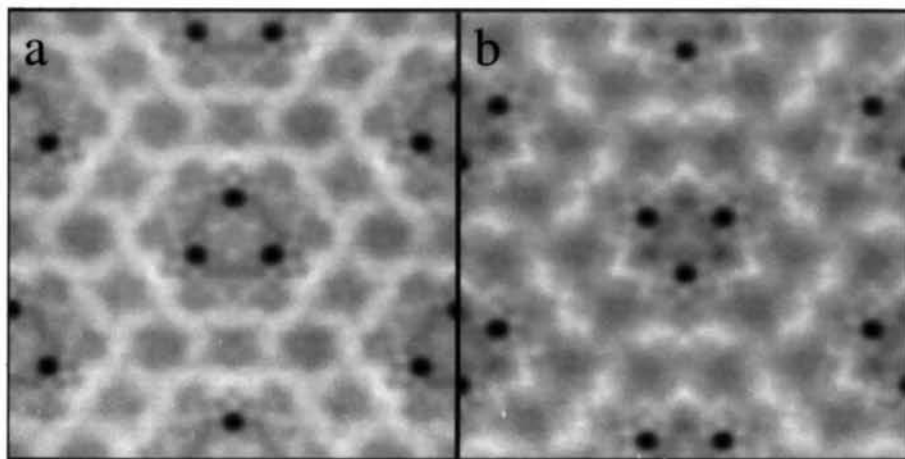

Fig. 4. Images of two solutions calculated for Model 1 without restricting the range of any beams. (a) and (b) have essentially identical FOM's.

possible phases, and those calculated with a limit of 0–180° placed upon the (2,2) beam (Fig. 5), one sees a larger spread of solutions in the set where no limits were imposed. This indicates the algorithm has explored multiple redundant solutions which differ only in their origin.

7. Solution hyperspace

An important issue is the shape of the solution hyperspace and how choosing different FOM's can alter how one explores this 'surface'. In reality, the solution hyperspace occupies $n + 1$ dimensions, where n corresponds to the number of variables, but for simplicity it is instructive to limit this to a two-dimensional surface. One can imagine at either end of the spectrum two types of features, very large shallow pits and very sharp deep spikes. If a very coarse search mesh is projected onto the surface, as in Fig. 6, a search will not necessarily detect the sharp deep minima. What will be found are large shallow minima. Alternatively, for a fine search mesh, all the solutions will occupy the bottom of the sharp deep minima, while solutions from the shallow minima evolve into extinction. In either case, potentially correct solutions are being lost. The FOM also influences the morphology of this surface. This is demonstrated by Fig. 2(a), where, depending on the FOM, the algorithm discovered different solutions. This is also illustrated in Fig. 7, which displays the ratio of the average C factor of the entire solution set for two different figures of merit used *versus* the number of chromosomes treated (see caption). For 60 parents and 180 children, minimization using the normalized structure factors converged to the correct solution more effectively than the unitary

structure factors. The opposite trend was observed for the case of 80 parents and 240 children.

8. Discussion

The general conclusion is that GA's are a viable method of handling direct phasing problems, determining a reasonable set of possible solutions much better than a simple random search. As such, they can readily be
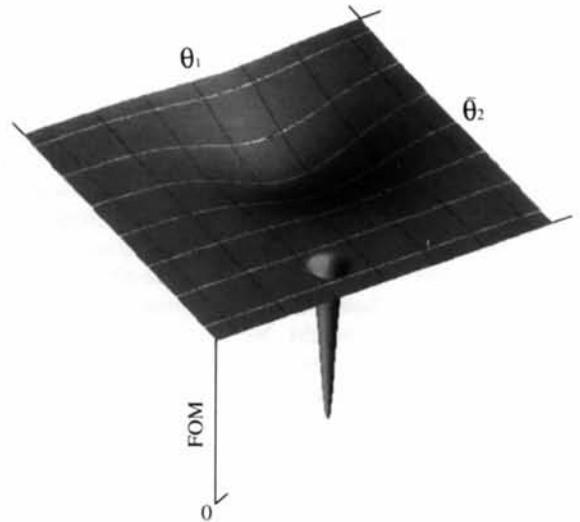
Fig. 6. Representation of a solution hyperspace surface for two different arbitrary phases, θ_1 and θ_2.
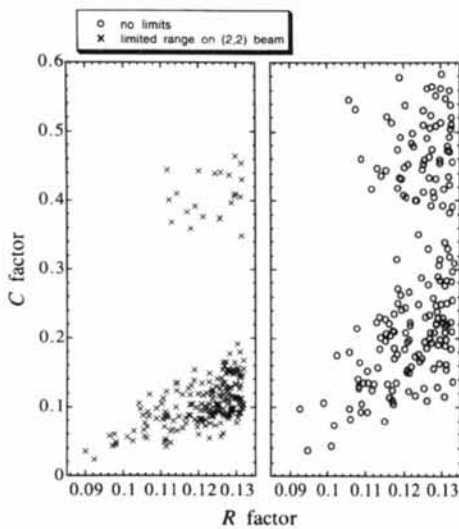
Fig. 5. Comparison of the different solutions found without placing limits on the phase of any of the beams being defined and limiting the (2,2) beam to a phase between 0 and 180°.
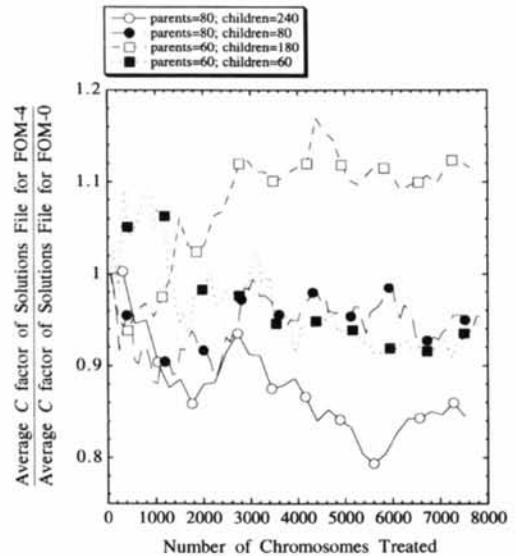
Fig. 7. Plot of the ratio of the average C factor for a solution file per number of chromosomes treated for two different FOM's. FOM-0 used the structure factors in (4), while FOM-4 used the unitary structure factors.

introduced into current methods such as *RANTAN* (Yao, 1981), *Shake-and-Bake* (Miller, Gallo, Khalak & Weeks, 1994) or even *Maximum Entropy* (Gilmore, Bricogne & Bannister, 1990) to control the choice of phases. We have found no problems simultaneously assigning phases for 20 beams with steps of 5–10° and this could be increased without major problems. The rate-limiting step is the phase extension.

It is evident that the ability of the algorithm to find multiple unique solutions is necessarily a combination of the number of parents and children, the mutation rate and the FOM used to define the problem. Unlike other applications of the GA, which look for a single best solution and apply a large amount of refinement, we ideally want a suitable FOM to explore all possible unique solutions without prejudice and then apply some refinement. From this point, chemical criteria and χ^2 analysis can be applied to the set of possible solutions to arrive at a final refined solution. Various modifications may be appropriate for larger problems and might give better convergence. For instance, 'niche specialization' (Goldberg, 1989) could be used to reinforce multi-solution behavior. Some effort to optimize the phase (schemata) ordering would be useful based upon a modification of the standard convergence analysis (Woolfson, 1987). Alternatively, methods exist for incorporating reordering operators to handle this dynamically (Goldberg, 1989). One can also devise selection criteria for the parents using multiple FOM's in a nonadditive sense, which may be useful.

This general FOM methodology has been applied to and solved several known surface reconstructions including the Au on Si(111) $3^{1/2} \times 3^{1/2}$ and 5×2 (Marks, Plass & Dorset, 1997), and the Si(111) 7×7 (Gilmore, Marks, Grozea, Collazo, Landree & Twesten, 1997) structures. The GA has since also been applied to several unknown structures such as In on Si(111) 4×1 (Collazo-Davila, Marks, Nishii & Tanishiro, 1997), $TiO_2(100)$ 1×3 (Landree, Marks, Zschack & Gilmore, 1997) and Au on Si(111) 6×6 (Marks, Grozea, Feidenhans'l, Nielsen & Johnson, 1997) with equally encouraging results.

References

Baker, J. E. (1985). In *Proceedings of an International Conference on Genetic Algorithms and Their Applications.* Hillsdale, NJ: Erlbaum.

Bhat, N. T. (1990). *Acta Cryst.* A46, 735–742.

Chang, G. & Lewis, M. (1994). *Acta Cryst.* D50, 667–674.

Collazo-Davila, C., Marks, L. D., Nishii, K. & Tanishiro, Y. (1997). *Surf. Rev. Lett.* 4, 65–70.

Davis, L. (1987). Editor. *Genetic Algorithms and Simulated Annealing.* London: Pitman Publishing.

Davis, L. (1989). In *Proceedings of the Third International Conference on Genetic Algorithms*, edited by J. D. Schaffer, pp. 61–69. San Mateo, CA: Morgan Kaufman Publishers, Inc.

De Jong, K. A. & Spears, W. M. (1990). In *Proceedings of First Workshop on Parallel Problem Solving from Nature*, edited by G. Goos & J. Hartmanis, pp. 38–47. Berlin: Springer-Verlag.

Germain, G. & Woolfson, M. M. (1968). *Acta Cryst.* B24, 91–96.

Gilmore, C. J. (1984). *J. Appl. Cryst.* 17, 42–46.

Gilmore, C. J., Bricogne, G. & Bannister, C. (1990). *Acta Cryst.* A46, 297–308.

Gilmore, C. J., Marks, L. D., Grozea, D., Collazo, C., Landree, E. & Twesten, R. D. (1997). *Surf. Sci.* 381, 77–91.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Language.* New York: Addison-Wesley Publishing Company, Inc.

Grefenstette, J. J. (1986). *IEEE Trans. Syst. Man Cybernet.* SMC-16, 122–128.

Gutowski, M. W. (1994). *J. Phys. A*, 27, 7893–7904.

Holland, J. H. (1968). *Descriptions of Universal Spaces and Adaptive Systems.* Technical Report. ORA Projects 01252 and 08226. Department of Computer and Communication Sciences, University of Michigan.

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems.* University of Michigan Press.

Koza, J. R. (1992). *Genetic Programming.* The MIT Press.

Landree, E., Marks, L. D., Zschack, P. & Gilmore, C. J. (1997). *Surf. Sci.* Submitted.

Marks, L. D., Grozea, D., Feidenhans'l, R., Nielsen, M. & Johnson, R. L. (1997). *Surf. Rev. Lett.* Submitted.

Marks, L. D., Plass, R. & Dorset, D. L. (1997). *Surf. Rev. Lett.* 4, 1–8.

Marks, L. D., Savage, T. S., Zhang, J. P. & Ai, R. (1991). *Ultramicroscopy*, 38, 343–347.

Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* 27, 613–621.

Miller, S. T., Hogel, J. M. & Filman, D. J. (1996). *Acta Cryst.* D52, 235–251.

Nakano, R., Davidor, Y. & Yamada, T. (1994). In *Proceedings of the Third Conference on Parallel Problem Solving from Nature*, edited by Y. Davidor, H.-P. Schwefel & R. Männer, pp. 130–138. Berlin: Springer-Verlag.

Schaffer, J. D., Caruana, R. A., Eshelman, L. J. & Das, R. (1989). In *Proceedings of the Third International Conference on Genetic Algorithms*, edited by J. D. Schaffer, pp. 51–60. San Mateo, CA: Morgan Kaufman Publishers, Inc.

Sheldrick, G. M. (1990). *Acta Cryst.* A46, 467–473.

Tanishiro, Y. & Takayanagi, K. (1989). *Ultramicroscopy*, 27, 1–8.

Woolfson, M. M. (1987). *Acta Cryst.* A43, 593–612.

Woolfson, M. M. (1991). *Direct Methods of Solving Crystal Structures*, edited by H. Schenk, pp. 41–51. New York: Plenum Press.

Woolfson M. M. & Hai-fu, F. (1995). *Physical and Non-Physical Methods of Solving Crystal Structures.* Cambridge University Press.

Xiao, Y. L. & Williams, D. E. (1993). *Chem. Phys. Lett.* 215, 17–24.

Yao, J. X. (1981). *Acta Cryst.* A37, 642–644.