PSSP: AN OPEN SOURCE POWDER STRUCTURE SOLUTION PROGRAM FOR DIRECT SPACE SIMULATED ANNEALING

Peter W. Stephens and Ashfia Huq
Department of Physics and Astronomy, State University of New York at Stony Brook
Stony Brook, New York 11794-3800

1 ABSTRACT

This paper describes a moderately new computer program *(PSSP for Powder Structure Solution Program)* for the solution of molecular crystal structures from X-ray powder diffraction data. The procedure is based in the agreement between the integrated intensities arising from trial structural models and the experimental diffraction data (Le Bail fits), and a new agreement factor considering overlapping peaks is introduced. The program *PSSP* is available in the public domain at http://powder.physics.sunysb.edu; complete operating instructions are given in the web site. The use of the program is illustrated with two examples: the drug ranitidine hydrochloride, and a Re complex.

2 INTRODUCTION

Single crystal X-ray diffraction is certainly the most widely used and powerful technique for solution of crystal structures; however many materials are available only as microcrystalline powders. The information contained in a powder diffraction pattern is intrinsically more limited since the three-dimensional intensity information of single crystal diffraction data is compressed to one dimension. The overlap of the diffraction peaks makes uncertain the assignment of individual intensities to the Bragg reflections traditionally used in the determination of crystal structures. This Transactions Symposium illustrates the rapid growth of techniques to deal with powder data. This article does not endeavor to provide a review of the development or achievements in the field; I hope that none of the authors who have made important contributions will feel slighted.

For organic molecular solids, a large amount of molecular geometry, e.g., most of the bond lengths and angles, can be predicted from the previously known chemical structure of the molecule. Accordingly, the crystal structure of a molecular solid can be described by specifying the lattice parameters and space group, the location and orientation of the molecule in the unit cell, and any internal degrees of freedom. For a rigid molecule, the location may be specified as three fractional coordinates of some reference point on the molecule, and the orientation as three Eulerian angles of rotation about that reference point with respect to some chosen axes. In most cases, an organic molecule of low symmetry will be located at a general position of the unit cell, but it is not necessarily so. This may be extended to structures of salts, where the counter-ion must be independently located, and to structures with more than one identical molecule in inequivalent crystallographic sites.

The method used in PSSP (and many other programs, publicly available and otherwise) is to produce candidate structures in direct space and compare their powder diffraction patterns to the experimental data. Nevertheless, in typical problems, there are ten to twenty or more parameters, and so it is not a trivial task to find the global minimum. The simulated annealing algorithm was invented in 1983 as a method to find approximate solutions to global optimization problems in spaces that are too large to explore exhaustively.[1] The principle is simple and this method probably has the potential to be almost routinely

applied to appropriate structure determinations from powder data. However, more experience is needed in testing various strategies for annealing schedules, structure parameterization, and measurement of quality of agreement between calculated and experimental data. Accordingly, we have developed a simple but powerful *Powder Structure Solution Program* and made it available in the public domain. The source code for *PSSP* is written in C, and is available, with examples, at http://powder.physics.sunysb.edu.[2]

3 METHOD

Simulated annealing is essentially a series of searches applying the Metropolis criterion at decreasing values of the control parameter (temperature). If carried out with enough trial structures at each temperature, it is likely to reach a solution with low value of the cost function, i.e., good agreement between experimental and calculated diffraction pattern. Of course, there is no guarantee that any particular solution found will be the true optimum. To apply simulated annealing to any problem, it is necessary to have:

A. a generation mechanism that postulates trial solutions by modifying a given configuration,
B. a cost function that ranks the quality of each configuration (analogous to the energy of a given configuration of a physical system), and control parameter (analogous to the temperature), that determines the fraction of unfavorable local steps accepted in the search.
C. a strategy for evolution of the pseudo-temperature optimized to obtain solutions as quickly as possible.
D. a plan for when to stop.

Each of these four aspects of simulated annealing is a target for further improvement. We have written PSSP with the intention that it will be flexible enough to serve as a test bed for such research, by ourselves and others. In the following paragraphs, we describe the current approach taken, and point out different possibilities.

3.1 The Generation Mechanism

In this work, structure is defined by taking a prototype of one or more molecules and placing them into the crystallographic unit cell. In most of the cases, organic molecules are not rigid, so a variable number of intramolecular degrees of freedom (molecular torsions) have to be explored to determine the conformation adopted in the solid. In many cases of interest, the internal degrees of freedom of the molecule may be described as rotations of one or more atoms about a particular interatomic axis. For example, in the ranitidine molecule sketched in Fig. 1 and discussed in section III.A., one can define how the molecule is twisted up by stating the eleven torsion angles illustrated.

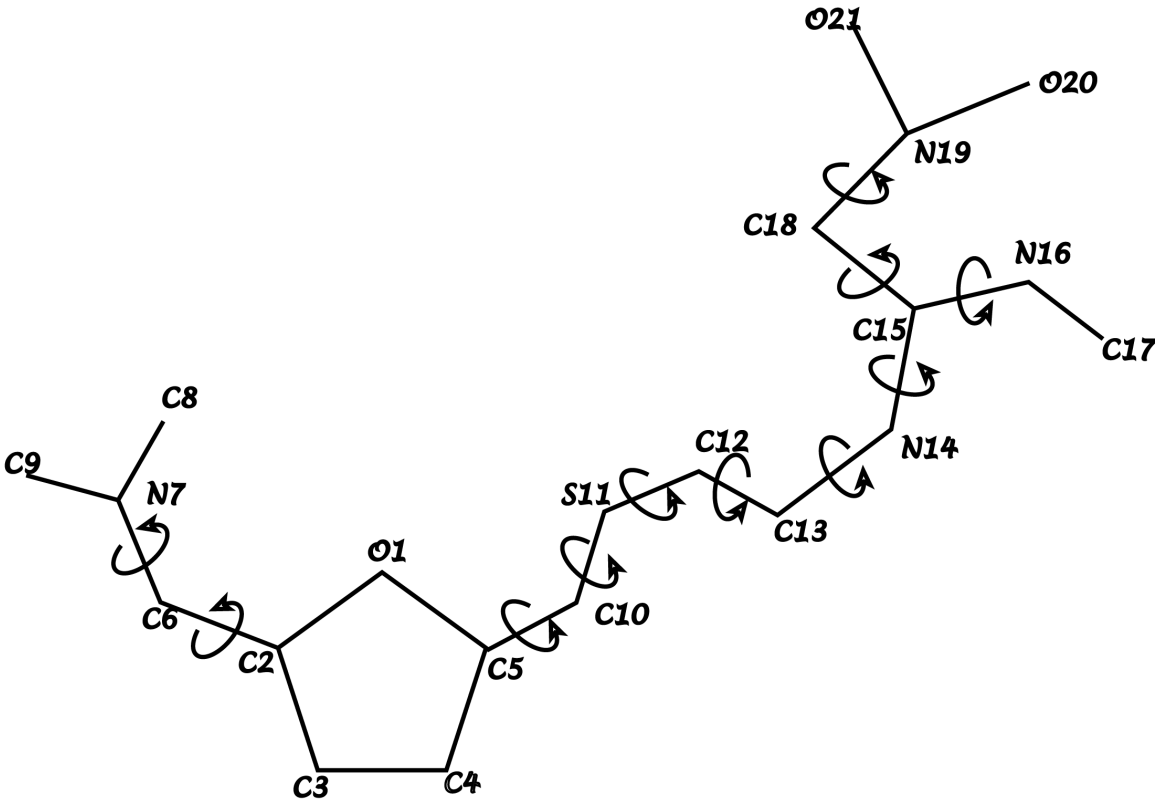

Figure 1. The ranitidine molecule, showing single bonds which are presumably torsionally flexible.

Any strategy of generating new trial configurations from such mechanism must be a compromise between using the partial success from the previous trial and a willingness to jump elsewhere in configuration space to get out of false minima. In this work, each trial configuration is described as a set of parameters $\{P_i\}$, and a trial step is defined by a series of increments $\{\Delta_i\}$. Each parameter P_i has a natural range R_i, generally 0° to 360° for Eulerian angles and unrestricted bond rotations, and 0 to 1 for fractional coordinates, with periodic boundary conditions. If a parameter P_i is restricted to a smaller range, periodic boundary conditions are not used, and the allowed Δ_i values are restricted to insure that P_i remains within its allowed range. Small steps allow the improvement of a nearly-correct structure, but room must be left for major reconfigurations. We have tried three algorithms:

1. In each trial solution, one parameter is chosen at random, and it is changed to a random value distributed uniformly throughout its range. We have found this to be the most effective.
2. In each trial solution, all parameters are changed together, with a preset bias towards small steps. We were surprised to discover that this is significantly less effective than #1 alone, and largely independent of the degree of bias towards smaller steps.
3. A hybrid of #1 and #2, in which alternate steps are chosen by each method. In practice, this runs at about half the speed of #1

Consequently, after considerable testing, we have abandoned algorithms #2 and #3 and use only #1.

The set of adjusted parameters$\{P_i\}$code for the generation of atomic coordinates through a script in the control file. An example is given below; for details, the reader should refer to a full description of the program.[2]

3.2 The Cost Function

The overlap of nearby powder diffraction peaks is a universal problem in the extraction of intensities for any sort of analysis. Until the late 1960's, powder data analysis was generally based on extracting integrated intensities for individual peaks. One of the most important advances in the technique of powder diffraction was the introduction of the Rietveld method,[3] in which the raw data is compared with a model profile, which includes a description both of the diffracted intensities (atomic information) and of the lineshapes (instrumental broadening, crystallite size and strain, etc.) It is widely regarded that the Rietveld method is superior to extracted intensities, because no assumptions are made about the information contained in the raw data that are not supported by the refined structural model. Many of the currently used programs for structure solution from powder data compute profiles and compare them to the raw data, *a la* Rietveld, a rather computationally cumbersome task. In attempting to design an economical method for computing the Rietveld profile, we hit upon the following shortcut.

The premise of the Rietveld method is that all of the peaks in a powder pattern can be described by a few parameters, so that, *e.g.*, a peak at angle 2θ is a Voigt lineshape with $\Gamma^2_{Gaussian} = U\tan^2\theta + V\tan\theta + W$ and/or $\Gamma_{Lorentzian} = X/\cos\theta + Y\tan\theta$, and the parameters *U, V, W, X, Y* are adjusted to fit the entire pattern. Assume that the lineshape and lattice parameters, as well as estimates A_{hkl} of each reflection intensity, have been determined by a profile (Le Bail [4]) fit to the raw data, so that the experimental profile is well fitted by

$$I_{LeBail}(2\theta) = Background(2\theta) + \sum_{hkl} A_{hkl} f_{hkl}(2\theta) . \quad (1)$$

Here, $f_{hkl}(2\theta)$ is the fitted functional form of the (*hkl*) peak, normalized to unit integrated intensity. Then, at any given stage of structure solution or refinement, the model profile would be

$$I_{Model}(2\theta) = Background(2\theta) + \sum_{hkl} B_{hkl} f_{hkl}(2\theta) , \quad (2)$$

where B_{hkl} are the calculated integrated intensities.

Disagreement between the raw data and the model *profile* may be assessed through the reduced χ^2, and if we grant that the profile fit is a good one, why not calculate agreement between the fitted (Le Bail) and model profiles?

$$\chi^2 = \frac{1}{N-R-1}\sum_{2\theta}^{N}\left[\frac{I_{LeBail}(2\theta) - I_{Model}(2\theta)}{\sigma(2\theta)}\right]^2 , \quad (3)$$

with *N* the number of points in the profile, *R* the number of parameters refined in the structural model, and σ(2θ) the statistical error expected of the profile point at 2θ.

Substituting the profiles in terms of individual peak functions,

$$\chi^2 = \frac{1}{N-R-1}\sum_{2\theta}^{N}\left[\frac{\sum_{hkl} A_{hkl} f_{hkl}(2\theta) - \sum_{h'k'l'} B_{h'k'l'} f_{h'k'l'}(2\theta)}{\sigma(2\theta)}\right]^2 \quad (4)$$

To optimize efficiency, one might think of calculating once and storing the fitted functional form of each peak $f_{hkl}(2\theta)$. Better yet, the order of the sums could be exchanged if one could get factor out the term $\sigma(2\theta)$. Let us define a new agreement factor, based on the naïve argument that a good solution is a good solution whatever the weighting factor.

$$S = \frac{\sum_{2\theta}[I_{LeBail}(2\theta) - I_{Model}(2\theta)]^2}{\sum_{2\theta}[I_{LeBail}(2\theta)]^2}, \tag{5}$$

where the denominator is chosen to make it look something like an *R* factor, and to give a sensible normalization. This can be rearranged to

$$S = \frac{\sum_{hkl,h'k'l'}(A_{hkl} - B_{hkl})F_{hkl,h'k'l'}(A_{h'k'l'} - B_{h'k'l'})}{\left(\sum_{hkl} A_{hkl}\right)^2}, \tag{6}$$

with

$$F_{hkl,h'k'l'} = \sum_{2\theta} f_{hkl}(2\theta) f_{h'k'l'}(2\theta). \tag{7}$$

There are several noteworthy points about this formulation. The overlap coefficient $F_{hkl,h'k'l'}$ is large only for closely separated peaks, and so the sum in the numerator contains only a few times more terms than the number of reflections under consideration. $F_{hkl,h'k'l'}$ can be calculated once and tabulated at the beginning of a search for a structure solution, to speed the calculation of the *S* factor for a candidate solution. In our work, we have limited the consideration of overlapping peaks to a band of five distinct (*h'k'l'*) values above each (*hkl*). For simplicity, we have used an analytical approximation based on treating the overlapping peaks as if they were pure Gaussians, so that

$$F_{hkl,h'k'l'} = \frac{1}{\sqrt{2\pi}\gamma}\exp-\left(\frac{(2\theta_{hkl} - 2\theta_{h'k'l'})^2}{2\gamma^2}\right), \text{where } \gamma^2 = \frac{FWHM_{hkl}^2 + FWHM_{h'k'l'}^2}{8\ln 2}. \tag{8}$$

(Here γ is used for the standard deviation of a Gaussian lineshape.) The approximation of Eq. 8 may shift the estimate of *S* from its defined value, but it does not appear to degrade the ability of the algorithm to find a satisfactory minimum of *S*.

Equation 6 is similar to the formulation by David *et al*.,[5] used in DASH, based on a least-squares (Pawley [6]) fit of all integrated intensities, in which the place of the overlap coefficients is taken by elements of the correlation matrix. That formulation retains the normalization of statistical standard errors from the Pawley fit, including the influence of the background intensity on the counting statistics. Le Bail fits are available in widely used Rietveld packages such as Fullprof [7] and GSAS,[8] whereas neither of these programs supports Pawley fits.

In both DASH and PSSP, the formulation makes use of all of the information in a profile, even as it has been reduced to estimates of the integrated intensities and the overlap or correlation coefficients. Several authors apparently perform simulated annealing or other techniques, computing an entire profile and comparing it with the experimental data, [9-11] but that involves significantly more computational overhead, and does not retain any more information from the original profile.

3.3 The Evolution of the Control Parameter

Recall that simulated annealing consists of running the Metropolis algorithm at a given temperature, then lowering the temperature and repeating. In PSSP, we have taken the simplest approach of fixed cooling rate. In PSSP, there are four parameters used to specify a cooling schedule: the initial temperature T_1, cooling ratio alpha, the final temperature T_f, and the number I_{MC} of Monte-Carlo cycles to be run at each temperature. At the end of the N_{MC} Monte Carlo cycle at a given temperature T_n, the temperature is decreased, and the next cycle is begun at $T_{n+1} = \alpha T_n$; the process stops when $T_n < T_f$. We have done most of our work with α=0.8. There is doubtless significant improvement in performance to be gained by choosing a better cooling contour, spending more time when it is needed. One could cascade cooling cycles with different rates between different intermediate temperatures, based on empirical evaluations of performance. Alternatively, David *et al.* have used an algorithm that senses energy fluctuations to determine the temperature at which the annealing process is most productive.[5]

One issue is whether the Monte Carlo cycles at T_{n+1} should start at the best or at the last achieved configuration $\{P_i\}$ at T_n. It would seem that starting at the last solution is closest to the philosophy of simulated annealing, as that is more of a "typical" member of the ensemble than is the lowest energy. However, we have found that the performance is much better if we start from the best candidate. This may be a consequence of the aggressive ratio by which we lower alpha. Within this paradigm of taking the best candidate before lowering the temperature, a ratio $\alpha \sim 0.8$ turns out to be roughly optimal.

3.4 When to Stop

At some point, a given annealing cycle condenses into some local minimum, which is either the desired solution or one of a myriad of local minima. Our general experience is that candidates with $S > 0.1$ are not worth considering, but that the temperature T has to be below 0.01 for valid solutions to be separated from invalid ones. However, again, experience is needed in a wider variety of problems in order to establish heuristics that can be used to solve truly unknown structures. It is likely that time could be saved by switching to a straightforward least squares refinement on S at the later stages of searching; we have not yet implemented that idea.

We always run repeated cycles of annealing from random starting points in order to ensure that we are exploring the solution space sufficiently well. The final values of S generally fall into two groups that allow us to distinguish candidate solutions to be taken seriously. Finally, if a given problem cannot be solved with a given number of Monte Carlo cycles N_{MC}, we increase the number by factors of two, five, or ten and let it run some more, in the hope that sufficient computation will find a good solution.

4 EXAMPLES AND RESULTS

The high resolution X-ray powder diffraction data of all samples was collected at the X3B1 powder diffraction beamline, National Synchrotron Light Source, Brookhaven National Laboratory. The wavelength is selected by a monochromator consisting of two parallel Si(111) crystals. The incident beam on the sample is measured using an ion chamber. The horizontal resolution in the diffracted beam is given by slits whereas the vertical resolution is determined by a Ge(111) analyzer crystal. The diffracted beam is measured with a NaI(Tl) scintillation detector.

Several previously unknown structures solved with PSSP have been published: an assortment of small molecules that we used to develop the program,[2] synthetic malaria pigment,[12] and three color polymorphs of a bicyclic system.[13] We present here two accounts of new work, partially completed.

4.1 Ranitidine HCl

The structure of this histamine H_2 receptor antagonist, commonly used as an anti-ulcer medication and originally known as Zantac®, has been previously solved from single crystal data.[14] It has been found that part of the molecule is orientationally disordered within the crystal structure. For this reason, it provides an interesting opportunity to test an issue which arises in the general field of direct space structure solution: what if the starting molecular model is not accurate? In particular, the usual assumption that there is a unique configuration of the molecule would not be correct, and it is of interest to see if this can be seen in the outcome of naively pursuing a solution by simulated annealing and direct space modeling.

The sample was obtained from USP, annealed for one hour at 60ºC under vacuum, and loaded into a 1.5mm diameter Lindemann capillary. X-ray data were collected at 1.1508 Å wavelength in steps of 0.005 degrees, with a counting time that increased from 2 seconds at low angles to 22 seconds at 48 degrees (minimum *d* spacing = 1.4Å). The cell was indexed to a monoclinic cell of dimensions 18.8086(4)Å x 12.9808(2)Å x 7.2113(1)Å, β = 95.048(1)º, space group $P2_1/n$, in agreement with the single crystal results. Intensities of the first hundred reflections were extracted by the Le Bail algorithm in the Rietveld program FULLPROF. For the purposes of illustration, we present the PSSP input file, suitably annotated.

```
title Ranitidine HCl
lebailf ranit.hkl 100
lambda 1.1508
lattice 18.808594  12.980787   7.211260  90. 95.048 90.
spacegroup P 21/n
symm  -2
1 0 0 0 1 0 0 0 1 0 0 0
-1 0 0 0 1 0 0 0 -1 .5 .5 .5
```

The first few lines of the file give a descriptive title, the file name for the reflection list and intensities. When the program reads in the Le Bail file, it calculates and stores the $F_{hkl,h'k'l'}$ overlap coefficients for later use. The next lines contain the experimental wavelength, unit cell parameters, the name of the spacegroup (unused in PSSP, passed to drawing programs), and the symmetry operators used to compute diffraction patterns. "symm -2" indicates that there are two symmetry operations (the identity and the glide plane), plus inversion symmetry; a positive number would indicate no inversion center. The following lines

contain the symmetry elements as nine matrix elements and three translations, most easily defined by example; here they are the identity (x, y, z), and (1/2-x, 1/2+y, 1/2-z).

```
molecule 0 21
8    O1   0.0000      0.00000     0.00000    1
6    C2  -1.156967    0.765869    0.001663   1
6    C3  -0.843246    2.104629    0.006378   1
6    C4   0.592957    2.190292    0.006302   1
6    C5   1.063111    0.896835    0.006302   1
6    C6  -2.45253     0.037842    0.025055   1
7    N7  -3.27665     0.388519   -1.170624   1
6    C8  -4.7043      0.050155   -0.976349   1
6    C9  -2.731109   -0.170868   -2.426651   1
6    C10  2.424942    0.338089   -0.014756   1
16   S11  3.181534    0.427368    1.65509    1
6    C12  4.630539   -0.652742    1.467636   1
6    C13  4.297364   -2.130921    1.634826   1
7    N14  5.458893   -2.989136    1.288315   1
6    C15  6.440186   -3.282685    2.271194   1
7    N16  7.088654   -4.533066    2.204178   1
6    C17  7.192581   -5.258347    0.927566   1
6    C18  6.736206   -2.373673    3.268066   1
7    N19  7.829178   -2.508866    4.20253    1
8    O20  8.028412   -1.64946     5.038528   1
8    O21  8.56987    -3.493912    4.138626   1
```

This defines a reference state of the ranitidine molecule. It was produced with the standard molecular mechanics program MOPAC,[15] and is used as the starting geometry, providing all bond lengths and angles. The first line labels this as molecule number zero, containing 21 atoms. Columns in the following lines are the atomic number, a site label, Cartesian coordinates in Å, and the occupancy factors (useful for molecules that are located on crystallographic special positions). The origin is here taken as the oxygen atom within the 5-cycle; this will be the center of rotation for the Euler angles, and the reference point of attachment within the crystallographic cell.

```
molecule 1 1
17 Cl 0.0 0.0 0.0 1
```

The chloride ion is defined as molecule number one.

```
params 20
0 360
0 360
0 360
0 1
0 1
0 1
0 1
0 1
0 1
0 360
0 360
0 360
0 360
0 360
0 360
0 360
0 360
0 360
0 360
0 360
```

This says that there will be twenty parameters to define the structure, and gives the range of each parameter. In this case, the Euler angles and bond torsions can lie anywhere in the range of 0 to 360º, and the fractional coordinates can range throughout the unit cell. Which parameter is which is defined in the next lines.

```
structure 22
getmol 0
rot_axis  6  7  10  8 9
rot_axis  2  6  11  7 8 9
rot_axis  5 10  12  11 12 13 14 15 16 17 18 19 20 21
rot_axis 10 11  13  12 13 14 15 16 17 18 19 20 21
rot_axis 11 12  14  13 14 15 16 17 18 19 20 21
rot_axis 12 13  15  14 15 16 17 18 19 20 21
rot_axis 13 14  16  15 16 17 18 19 20 21
rot_axis 14 15  17  16 17 18 19 20 21
rot_axis 15 16  18  17
rot_axis 15 18  19  19 20 21
rot_axis 18 19  20  20 21
rot_body_var z 1
rot_body_var x 2
rot_body_var z 3
putmol 0 0 0 4 5 6
getmol 1
putmol 0 0 0 7 8 9
endstructure
```

The text between control words "structure" and "endstructure" constitute a script that tells the program how to generate a candidate structure from the prototype and the adjustable parameters. It is useful to imagine that the prototype molecule is placed onto Procrustes' bed, where it is rearranged, and then moved into the unit cell. This is first done with the ranitidine molecule (`getmol 0`). Parameters 10 through 20 control the eleven torsions illustrated in Fig. 1. For example, the third line of this script says that atoms 6 and 7 define an axis, and atoms 8 and 9 will be rotated about that axis according to parameter number 10. This will preserve the bond distances C8-N7 and C9-N7 and angle C8-N7-C9, but vary the torsion C8-N7-C6-C2. Next, the assembly {N7, C8, C9} is rotated about the axis defined by C2 and C6; this leaves all bond distances and the C8-N7-C6-C2 torsion undisturbed, and varies only the torsion N7-C6-C2-O1. Torsion operations defined in this way commute, and so it does not matter which order they are defined. Note that the parameters take the values of the torsion relative to the reference configuration, not the actual torsion angle.

Once the internal structure of the molecule is defined through torsions, the entire molecule is oriented through three Euler angles, parameters 1, 2, and 3, conventionally taken about the *z*, *x*, and *z* axes respectively. The center of these rotations is the origin of the Cartesian coordinates read in with the reference structure, in this case, the oxygen atom O1. Three parameters (4-6) locate the twisted and oriented molecule into the crystallographic unit cell, as the fractional coordinates of the origin. Finally, the chloride ion must be located; it has no orientation, and so it is described only by three fractional coordinates, 7, 8, and 9.

```
random
anneal 50 .8 1000000 .001
pattern
fullprofpcr ranit.pcr
cif ranit.cif
```

Now the program starts simulated annealing cycles, trying to solve the structure. This series of commands says to start with a random set of parameters, and then perform a series of simulated annealing cycles, with initial temperature T_1 = 50, cooling ratio α = 0.8, final temperature T_f = 0.001, and I_{MC} = 1000000 Monte-Carlo cycles to be run at each temperature; in this case there will be $ln(0.001 / 50) / ln(0.8) = 49$ temperatures, so 4.9×10^7 structures must be computed and compared – approximately 50 hours for the 667 MHz Linux system used for this work. The program then writes the computed diffraction pattern into its output file, and writes two files useful for diagnostics: ranit.pcr is a control file for FULLPROF, to make sure that the diffraction pattern is computed correctly or to proceed to crystallographic refinements, ranit.cif is a fragment of a crystallographic information file useable in MERCURY or other imaging programs. Conventionally, we repeat the sequence of random, anneal, pattern commands many times to see if the program is reproducing the same solution.

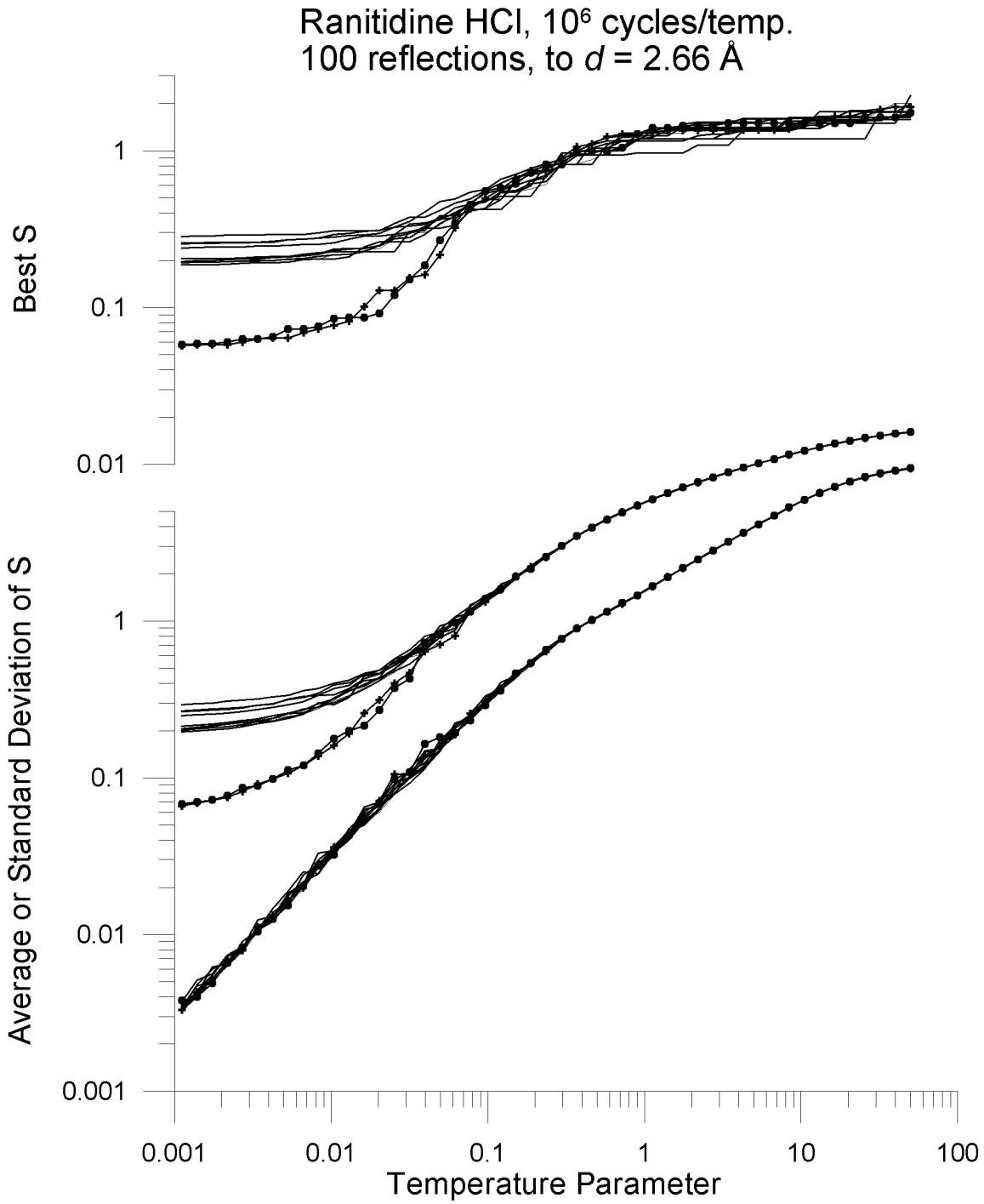

Figure 2. Progress of twelve independent runs of PSSP to solve the structure of Ranitidine HCl from powder data. From the top, the traces show the lowest value of *S*, the average value, and the standard deviation of a sample of values. The two successful solutions are indicated by solid circles and crosses along with the lines.

Performance of the simulated annealing algorithm for this problem is illustrated in Fig. 2. The program proceeds from right to left along the temperature axis, and the plots show the minimum value of *S* reached, the average *S*, and the standard deviation of *S* from the sample of accepted moves. Twelve runs are shown, of which two gave sufficiently low values of *S* that we could refine the structure. Note that the good solutions break away from the unsatisfactory ones at a temperature of ~0.07, where the average ± standard deviation of *S* is ~1.0±0.2, and the best match has *S* ~ 0.4. There is probably significant room for improvement in tuning up the annealing protocol.

Starting from the PSSP output, the structure may be Rietveld refined as illustrated in Fig. 3a. The Debye-Waller factor is separately refined for each atom. Any experienced crystallographer would reject the structure shown in Fig. 3a, for the reason that such widely discrepant thermal vibrations are unlikely for adjacently bonded atoms.

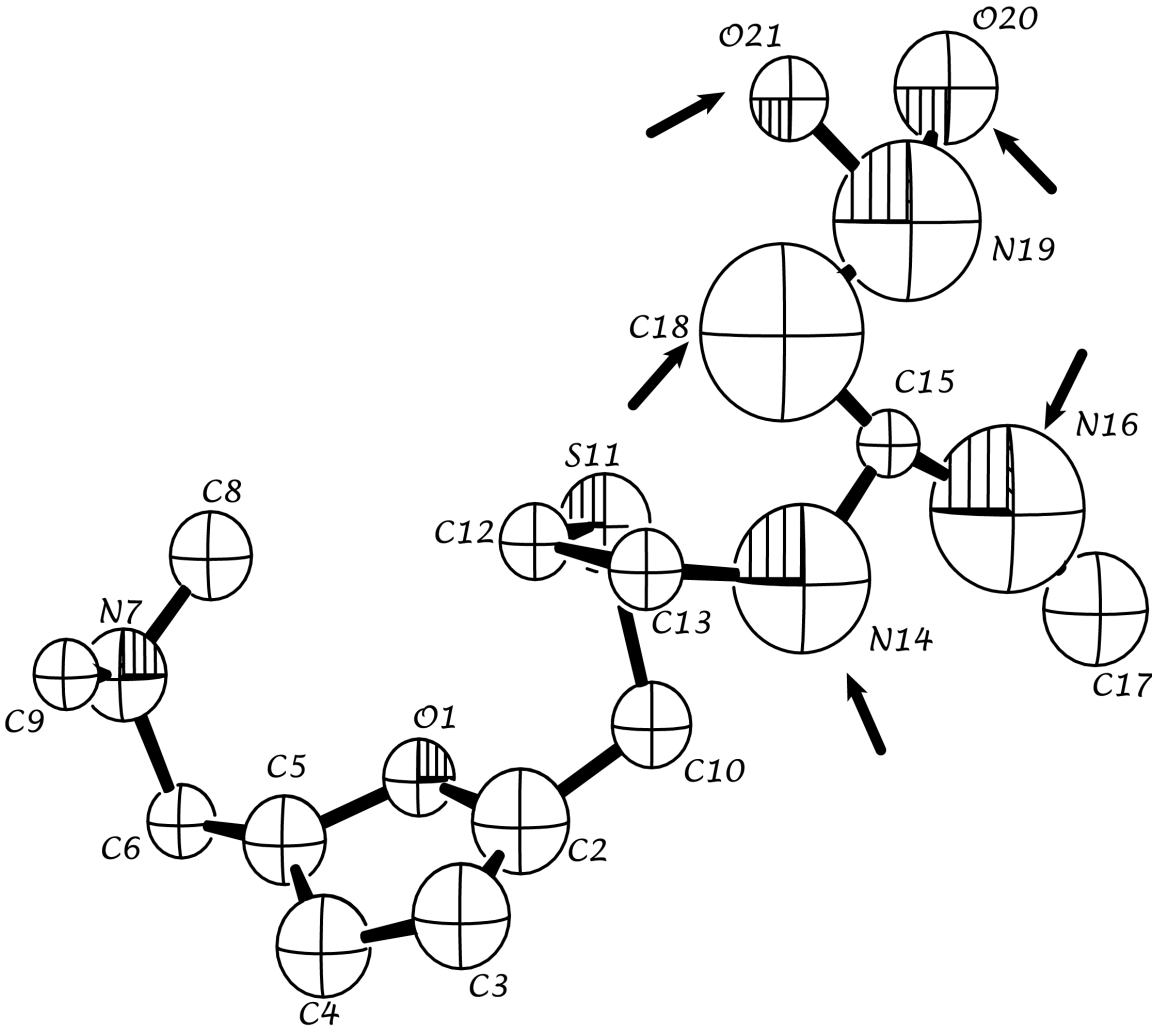

Figure 3a. One candidate structure of ranitidine hydrochloride determined *ab initio* from powder data. Spheres denote contours of 50% occupancy from independently refined thermal parameters.

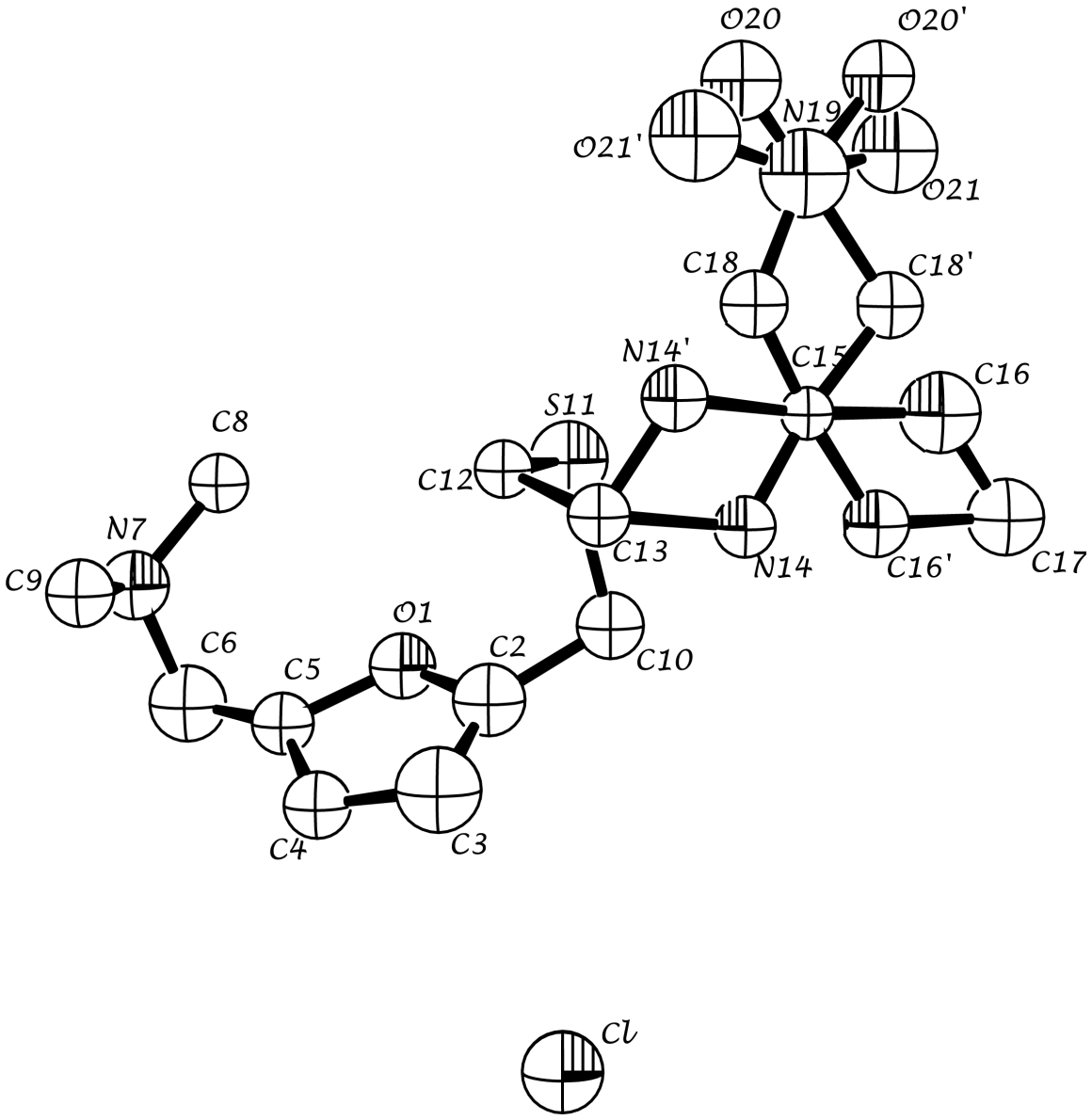

Figure 3b. Structure of ranitidine hydrochloride including the occupational disorder seen by comparing the four independent solutions from PSSP.

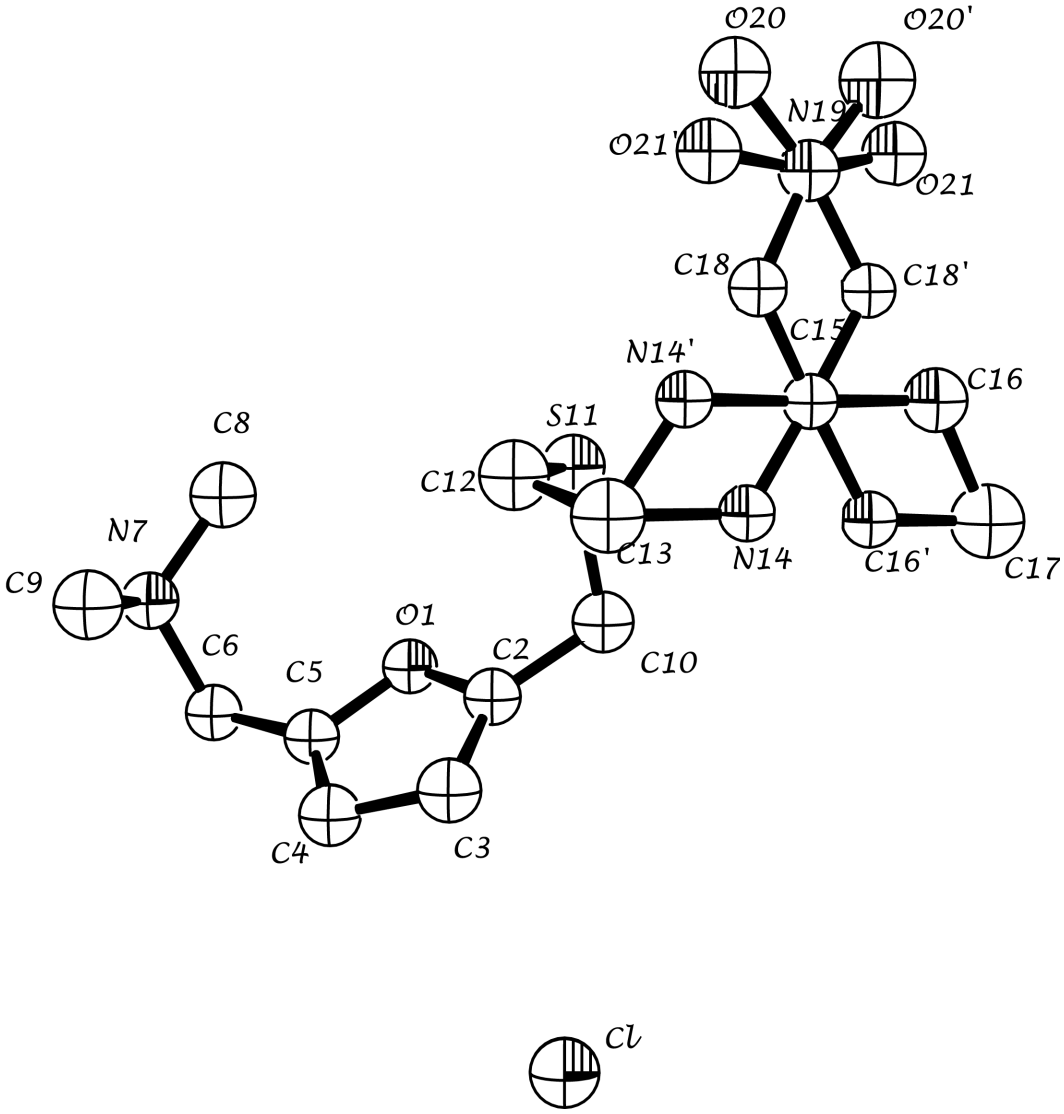

Figure 3c. Structure of ranitidine hydrochloride from single crystal experiment, taken from reference 15.

In a total of twenty runs, we obtained four candidate solutions with $S < 0.06$, but they were not all identical. These four solutions can be superimposed, with the backbone from C8 and C9 to C13 as well as C15, C17, and N19 in essentially identical positions, and two sets of positions for N14, C16, C18, O20, and O21. We accept this as evidence that PSSP has found the two distinct conformations already known from the single crystal structure. Figure 3b shows a refinement of the same powder data with 50% fractional occupancies in each of two sites for atoms N14, N16, C18, O20, and O21, as determined form the two different sets of PSSP solutions. It is clear that simulated annealing, through the program PSSP, has produced a good candidate structure, and that the powder data are of sufficient quality to find which atoms are disordered. Finally, in Fig. 3c, we present the known single crystal structure, which compares favorably with our solution shown in Fig. 3b.

4.2 $(\eta^5\text{-}C_5H_4SiMe_3)Re(CO)_2(NO)] \cdot BF_4$

The compound listed above was synthesized by the research group of David White at University of North Carolina at Wilmington as part of a research program in transfer of chirality from reagents to organometallic complexes. Single crystals are grown at room temperature and the structure has been solved in the orthorhombic spacegroup $P2_12_12_1$ with a volume of 401 Å^3 per formula unit. When heated above 100ºC, the crystals fracture into a good powder, and it is of interest to know if this irreversible transformation is merely a rearrangement of the molecular units, or if there is a solid-state ligand rearrangement reaction. We indexed the resulting powder to an orthorhombic cell with lattice parameters 29.294(1)Å x 12.910(1)Å x 13.228(1)Å, volume 4993 Å^3 implying Z=12 formula units per cell. Systematic absences imply spacegroup *Ama2* or $Cmc2_1$ which implies that there must be inequivalent molecules in the unit cell, because there are no sites in any orthorhombic spacegroup with multiplicity twelve. The resolution comes in part from noting that the Re complex could have a mirror symmetry. Inspecting the pictures in the spacegroup table, it is possible that the Re complex could be distributed between a general position and a position on a mirror plane. Likewise, the BF_4^- counterions could lie in general positions, mirror planes, or two-fold axes.

PSSP has been designed to be able to deal with problems in which molecules are located on special positions. For example, if a mirror-symmetric molecule is hypothesized to lie on a mirror plane, there would be two parameters to specify its position in the plane, and one to specify a rotation about an axis perpendicular to the plane. A molecule on a two-fold axis would have a single rotation and a single translation (as well as a possible flip parallel or antiparallel to the axis). This is best illustrated by example. Between the two candidate spacegroups and the various possibilities for the BF_4, there are eight possible cases, and we ran PSSP eight times. The only successful solutions came in $Cmc2_1$, which has a mirror plane at (0, *y*, *z*), with one copy of the Re complex and the BF_4 in a general position, and one of each in the mirror plane. That is a somewhat obscure spacegroup for organic and organometallic structures, accounting for only 0.2% of the entries in the Cambridge database.

The use of PSSP for this problem is illustrated in the following script.

```
structure 42
getmol 0
rot_body_var z 1
rot_body_var x  2
rot_body_var z  3
putmol 0 0 0 4 5 6
getmol 1
rot_body_var x 7
putmol 0 0 0 0 8 9
getmol 2
rot_body_var z 10
rot_body_var x 11
rot_body_var z 12
putmol 0 0 0 13 14 15
getmol 3
rot_body_var x 16
putmol 0 0 0 0 17 18
endstructure
```

This says that molecule 0 (the Re complex) is in the general position, and we see that it is subjected to three Euler rotations, and put into a general position by parameters 1 through 6. Molecule 1 is defined the same set of Cartesian (Ångstroem) coordinates, but with occupancies of ½ for each atom, because it will be doubled by the space group operations. Its reference structure starts out with the mirror symmetry that an atom at (*x*, *y*, *z*) is echoed by one at (*x*, -*y*, -*z*), and so it should be rotated only about the *x* axis (parameter 7) and put onto the *y*-*z* plane (parameters 8 and 9). The same operations are performed on the BF_4 molecule, although it is not clear that its orientation is important – it may well be orientationally disordered. (In this case, the mirror plane is at $x = 0$; if it had been at $x = ¼$, this could have been encoded by putmol 0.25 0 0 0 8 9.) Figure 4 shows the two independent Re complex molecules and the two independent BF_4 molecules at a particular position within the unit cell – symmetry equivalent molecules are not shown for clarity.

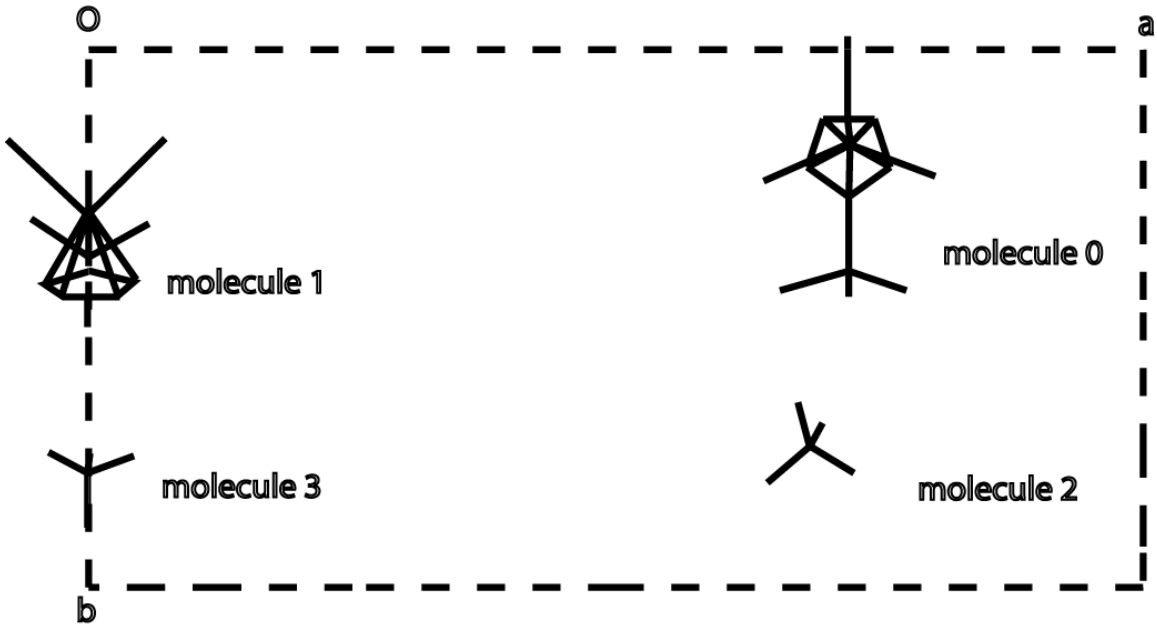

Figure 4. Unit cell of $(\eta^5\text{-}C_5H_4SiMe_3)Re(CO)_2(NO) \cdot BF_4$, viewed along the crystallographic *c* axis, with *a* axis horizontal. Two molecules of the Re complex and two BF_4 molecules are shown; the molecules on the left are in the mirror plane, and the molecules on the right are in general positions and orientations.

Candidate solutions for this structure have values of *S* in the range of 0.037 to 0.055. A partial refinement of this structure is shown in Fig. 5, and the structure itself in Fig. 6. At this point, we have only refined the position and orientation of the four independent molecules as rigid bodies. Since the volume per molecule has expanded by 4% from the room temperature state, it is likely that it is molecular disorder of the BF_4^- ions, the $SiMe_3$ moieties, and/or the $(CO)_2NO$ moieties which drive the phase transition. We hope to resolve that issue with Fourier maps or other techniques in the near future.

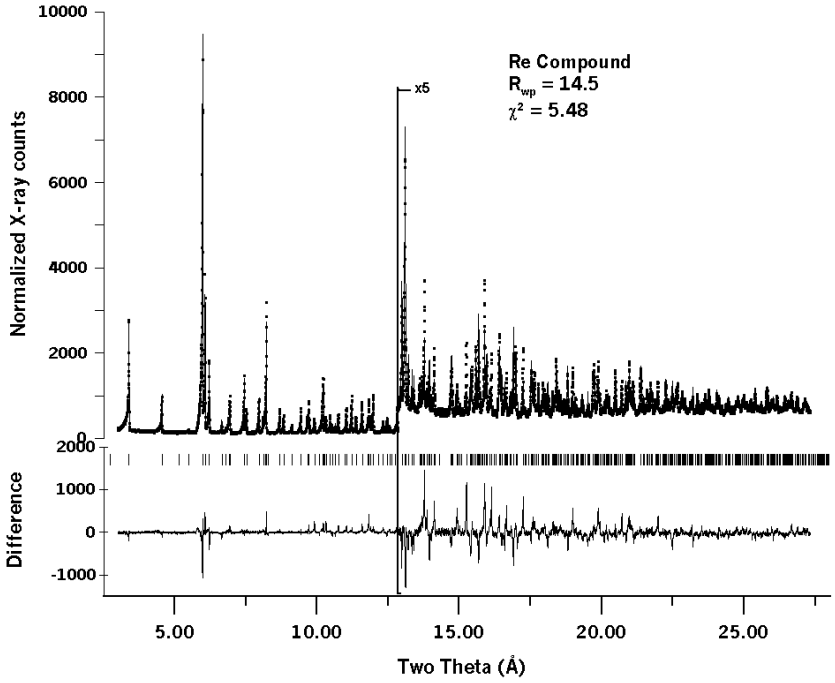

Figure 5. Rietveld refinement of $(\eta^5\text{-}C_5H_4SiMe_3)Re(CO)_2(NO) \cdot BF_4$ at room temperature, in the metastable high temperature phase. The refinement is not yet completed; the molecules have been refined only as rigid bodies based on the room temperature structure.

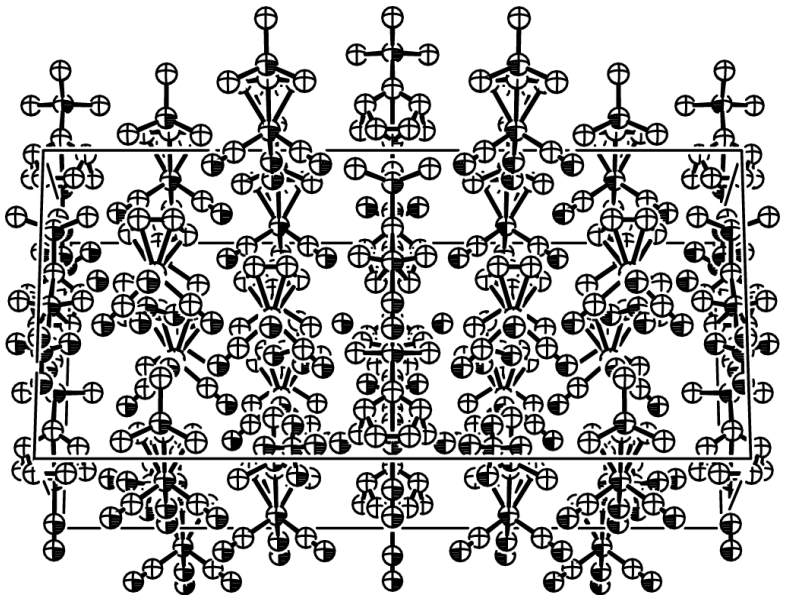

Figure 6. Structure of high temperature phase of $(\eta^5\text{-}C_5H_4SiMe_3)Re(CO)_2(NO) \cdot BF_4$, viewed along b axis. The mirror planes are vertical, at the left and right sides and center of the unit cell.

5 ACKNOWLEDGMENTS

We are grateful to Bill David and Ken Shankland for inspiration and numerous useful conversations. Their program DASH seems to be the gold standard of simulated annealing, with significantly better performance than the current incarnation of PSSP. Silvina Pagola has made important contributions to PSSP, and so this work also reflects her efforts. The Re complex was provided, and brought to our attention, by David White of the University of North Carolina at Wilmington.

Research performed at the National Synchrotron Light Source, Brookhaven National Laboratory, which is supported by the US Department of Energy, Division of Chemical Sciences and Division of Materials Sciences. The SUNY X3 beamline at the National Synchrotron Light Source was supported by the Division of Basic Energy Sciences of the US Department of Energy (DE-FG02-86ER45231).

6 REFERENCES

[1] S. Kirkpatrick, C.D. Gellatt Jr., M.P. Vecchi *Science* **220,** (1983) 671-680.

[2] S. Pagola and P.W. Stephens, *Materials Science Forum* **321-324**, (2002) 40-45; program and instructions available at http://powder.physics.sunysb.edu.

[3] H.M. Rietveld, *J. Appl. Cryst.* **2**, (1969) 65-71.

[4] A. Le Bail, H. Duroy, and J.L. Fourquet, *Mater. Res. Bull.* **23**, (1988) 447-452.

[5] W.I.F. David, K. Shankland, and N. Shankland, *Chem. Comm.* (1998) 931-932.

[6] G.S. Pawley, *J. Appl. Crystallog.* **14,** (1981) 357-61.

[7] J. Rodríguez Carvajal. (1990). *Program FULLPROF*, *Abstracts of the Satellite Meeting on Powder Diffraction of the XV Congress of the IUCr, Toulouse, France*; anonymous ftp from charybde.saclay.cea.fr, in directory pub/divers/fullp.

[8] A.C. Larson and R.B. Von Dreele, (1987). *Los Alamos Laboratory Report No. LA-UR-86-748*; available by anonymous ftp from ftp://ftp.lanl.gov/public/gsas.

[9] Y.G. Andreev and P.G. Bruce, *J. Chem. Soc., Dalton Trans.,* (1998) 4071-4080.

[10] G.E. Engel, S. Wilke, O. Konig, K.D.M. Harris, F.J.J. Leusen, *J. Appl. Cryst.* **32,** (1999) 1169-1179.

[11] M. Tremayne, B.M. Kariuki, K.D.M. Harris, *Angew. Chem. Int. Ed. Engl.* **36,** (1997) 770-772.

[12] S. Pagola, P.W. Stephens, D.S. Bohle, A.D. Kosar, and S.K. Madsen, *Nature* **404** (2000) 307-310.

[13] S. Pagola, P.W. Stephens, X. He, and S.R. Byrn, *Materials Science Forum* **378-3**, (2001) 789-794.

[14] T. Ishida, Y. In, and M. Inoue, *Acta. Cryst.* **C 46**, (1990) 1893-96.

[15] Stewart J. J. P. (1993). *MOPAC Manual (Seventh Edition)*. Available from Computational Chemistry List, www.ccl.net.