Acta Crystallographica Section A
Foundations of Crystallography
ISSN 0108-7673

Received 23 August 2001
Accepted 13 February 2002

Energy landscape paving for X-ray structure determination of organic molecules

Hsiao-Ping Hsu,[a] Simon C. Lin[a,b] and Ulrich H. E. Hansmann[c]*

[a]Computing Centre, Academia Sinica, Taipei, Taiwan, [b]Institute of Physics, Academia Sinica, Taipei, Taiwan, and [c]Department of Physics, Michigan Technological University, Houghton, Michigan, USA. Correspondence e-mail: hansmann@mtu.edu

The efficiency of a recently proposed novel global optimization method, energy landscape paving (ELP), is evaluated with regard to the problem of crystal structure determination from simulated X-ray diffraction data comprising integrated diffraction intensities. The new approach has been tested using the example of 9-(methylamino)-1*H*-phenalen-1-one 1,4-dioxan-2-y1 hydroperoxide solvate ($C_{14}H_{11}NO \cdot C_4H_8O_4$). The results indicate that, for this example, ELP outperforms standard techniques such as simulated annealing.

1. Introduction

Solving unknown crystal structures is a common and important problem in condensed-matter physics (both hard and soft) and materials science. In the early years, most researchers used traditional direct methods (Ladd & Palmer, 1980; Hauptman, 1986; Woolfson, 1987; Miller *et al.*, 1993; Weeks *et al.*, 1995) and the Patterson method (Clearfield *et al.*, 1984; Attfield *et al.*, 1986; Lightfoot *et al.*, 1987, 1992; Louër & Louër, 1987; Louër *et al.*, 1995), which try to reconstruct the experimentally non-direct accessible phase information to determine small molecular structures from X-ray single-crystal and powder diffraction data. More recently, the problem has also been considered in coordinate space. In such approaches, Monte Carlo methods, simulated annealing *etc.* (Reck *et al.*, 1988; Harris *et al.*, 1994; Cirujeda *et al.*, 1995; Su, 1995*a*,*b*; Dinnebier *et al.*, 1995; Harris & Tremayne, 1996; Kariuki *et al.*, 1996; Tremayne *et al.*, 1996, 1997; Andreev, Lightfoot & Bruce, 1997; Andreev, MacGlashan & Bruce, 1997; Andreev & Bruce, 1998; Zimmer & Su, 1998; Chen & Su, 2000) are employed and the structures are determined by minimizing the difference between the calculated and experimental diffraction patterns.

The usefulness of optimization techniques such as simulated annealing for crystal structure determination of large macromolecules, however, seems to be restricted (Harris *et al.*, 1994; Su, 1995*a*,*b*; Tremayne *et al.*, 1996; Harris & Tremayne, 1996; Kariuki *et al.*, 1996; Andreev, Lightfoot & Bruce, 1997; Andreev, MacGlashan & Bruce, 1997; Zimmer & Su, 1998; Chen & Su, 2000). This is because their efficiency depends strongly on the careful and system-dependent tuning of such algorithms, and decreases rapidly with the size of the molecule (Hsu *et al.*, 2001). Hence, for crystal structure determination there is still a need for the development and testing of new and more suitable global minimizers.

In this paper, we intend to test one such novel optimization technique, namely energy landscape paving (ELP) (Hansmann & Wille, 2002), for the problem of crystal structure determination from diffraction data. For this non-trivial problem, we wish to compare the efficiency of ELP in finding correct crystal structures with that of an established optimization technique, namely simulated annealing (SA). As a test system we choose a known compound, 9-(methylamino)-1*H*-phenalen-1-one–1,4-dioxan-2-y1 hydroperoxide solvate ($C_{14}H_{11}NO \cdot C_4H_8O_4$) (Hsu *et al.*, 2001), in order to evaluate the efficiency of ELP compared with that of SA. This compound has space group $P\bar{1}$ (triclinic), with cell constants $a = 6.9520$, $b = 9.6900$, $c = 12.5410$ Å, $\alpha = 77.11$, $\beta = 73.78$, $\gamma = 80.62^\circ$. The number of formula units per cell is $Z = 2$. Its structure, represented in Fig. 1, was solved by direct methods (*DIRDIF94*) (Beurskens *et al.*, 1994) and refined (on F^2) by using *SHELXL93* (Sheldrick, 1993) with anisotropic displacement parameters for all non-hydrogen atoms. Hydrogen atoms were located from difference maps and refined iso-

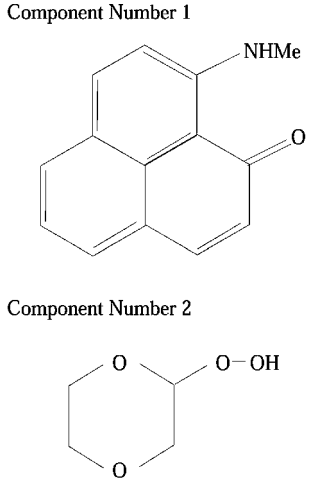

Figure 1
Chemical diagram of 9-(methylamino)-1*H*-phenalen-1-one 1,4-dioxan-2-y1 hydroperoxide solvate ($C_{14}H_{11}NO \cdot C_4H_8O_4$).

tropically. Since, in this article, we are mainly concerned with methodological questions, we decided to use synthetic data instead of the original experimental pattern. For this purpose, we reconstructed the diffraction pattern for solving the structure of $C_{14}H_{11}NO \cdot C_4H_8O_4$ from the positions of the known structure (Yatsenko *et al.*, 1998) using *PowderCell99* software (Kraus & Nolze, 1999). In this way, we obtained a 2θ (angle) *versus* I (intensity) diagram. The software also provides methods to index (h, k, l) each reflection extracted from the 'perfect' experimental result. Choosing Mo $K\alpha$ radiation ($\lambda = 0.71073$ Å) and $2\theta_{\max} = 50.67^\circ$ for our data collection, we obtained 2823 reflections. The so-obtained artificial pattern was used as the input for the ELP algorithm, leading to a configuration that can be compared with the known structure in order to verify the validity of our algorithm.

The paper is organized as follows. The techniques, SA and ELP, are briefly reviewed in §2. Our results are presented and discussed in §3. Finally, we present our conclusions in §4.

2. Methods

The use of any global optimization technique requires, as a first step, the choice of a suitable cost function. For the purpose of crystal structure determination, such a cost function can be defined by

$$E = \left\{ \sum_{j=1}^{N_k} [|F_{\rm cal}(\mathbf{k}_j)| - |F_{\rm obs}(\mathbf{k}_j)|]^2 \Big/ \sum_{j=1}^{N_k} |F_{\rm obs}(\mathbf{k}_j)|^2 \right\} \times 100, \quad (1)$$

where N_k is the total number of reflections, $F_{\rm obs}(\mathbf{k}_j)$ is the observed structure factor, and $F_{\rm cal}(\mathbf{k}_j)$ is the calculated structure factor. Here, the structure factors are given by

$$F(\mathbf{k}_j) = \sum_{i=1}^{N} f_i \exp[i2\pi(\mathbf{k}_j \cdot \mathbf{x}_i)], \quad (2)$$

with N the total number of atoms in a unit cell, and $\mathbf{x}_i$ the fractional coordinates of the ith atom. The scattering factor of an atom (the so-called atomic form factor) f_i is given by

$$f_i(\sin\theta/\lambda) = \sum_{q=1}^{4} a_{qi} \exp(-b_{qi} \sin^2\theta/\lambda^2) + c_i, \quad (3)$$

where a_{qi}, b_{qi} and c_i are the coefficients of the ith atom and can be obtained from the *International Tables for X-ray Crystallography* (Ibers & Hamilton, 1974).

Note that, because of the periodic boundary conditions and the symmetry of the space group, only the atoms in an asymmetric unit need to be considered. With the above cost function, the crystal structure corresponds to the deepest minimum in the energy landscape ($E = 0$) or, in the language of statistical physics, to the expectation value of the energy at $T = 0$. Hence, the optimization process is equivalent to a canonical ensemble simulation at $T = 0$, which in a simulation is approximated by a sufficiently low temperature. In this picture, the failure of conventional minimizers results from the roughness of the energy landscape with its huge number of local minima separated by high barriers. This is because the probability to cross an energy barrier of height ΔE decreases exponentially with $\exp(-\Delta E/k_B T)$ for simulations in a canonical ensemble and the simulation will become trapped in one of the multitude of local minima.

In SA, one tries to avoid trapping by modelling the crystal growth process in nature. During a simulation, the temperature is lowered very slowly with Monte Carlo dynamics from a sufficiently high initial value to a 'freezing' temperature T_f where the system undergoes no significant changes with respect to the Monte Carlo iteration. If the rate of temperature decrease is slow enough for the system to stay in thermodynamic equilibrium, then it is ensured that the system can avoid being trapped in local minima and that the global minimum will be found. It could be shown that convergence to the global minimum can be secured for a logarithmic annealing schedule (Geman & Geman, 1984), but constraints in available computer time enforce the choice of faster annealing schedules where success is no longer guaranteed. In earlier work (Hsu *et al.*, 2001), we found as an optimal annealing schedule a cooling where

$$T_i = \alpha T_{i-1} \quad \text{and} \quad T_1 > T_2 > T_3 > \ldots > T_{N_T}. \quad (4)$$

The initial temperature T_1 is set such that the acceptance ratio is about 0.5, and the total number of temperatures N_T in the annealing process has to be chosen to be large in order that the cooling rate

$$\log\alpha = (\log T_{N_T} - \log T_1)/(N_T - 1) \quad (5)$$

is slow enough for the system to stay in (quasi-)equilibrium at any stage of the annealing approach. For every temperature T_i, N_S Monte Carlo sweeps are performed. A Monte Carlo sweep is a sequence of M Metropolis steps, with M being the number of atoms in an asymmetric unit. In every Metropolis step, one tries to change the position of an atom according to

$$x'^j = x^j + \Delta x^j. \quad (6)$$

Here, x^j, $j = 1$, 2 and 3, are the fractional coordinates of the given atom, $\Delta x^j = r_s^j \cdot \eta_j$ is the displacement in the direction j, r_s^j is a scale factor to ensure equal acceptance ratios in three directions, and η_j is a random number between 0.5 and -0.5. Such a proposed move of an atom is then accepted with a probability $\min[1, \exp(-\Delta E/T)]$.

In ELP, new configurations are proposed in the same way as in SA by moving the selected atoms according to equation (6). However, ELP avoids trapping in local minima in a way that is different from SA: the problem of exponential slow convergence in low-temperature simulations is circumvented through the use of a modified energy function designed in such a way that it steers the search away from regions that have already been explored. For this purpose, we use a simple modification of the Metropolis algorithm where configurations are accepted with probability $\min[1, \exp(-\Delta E'/T)]$. Here, T is a low temperature and $\Delta E'$ is the replacement of the commonly used energy difference $\Delta E = E_{\rm new} - E_{\rm old}$ according to

$$\Delta E' = \Delta E + \varepsilon \frac{H(E_{\rm new}, t) - H(E_{\rm old}, t)}{H(E_{\rm new}, t) + H(E_{\rm old}, t)}. \quad (7)$$

In the above expression, ε is a constant and of order $\mathcal{O}(n_F)$ with n_F being the number of degrees of freedom of the molecule in the asymmetric unit. $H(E, t)$ is the accumulative histogram of configurations with cost function E at 'time' t (in Monte Carlo sweeps) of the simulation. The histogram is updated at each Monte Carlo step; hence the 'time' dependence of $H(E, t)$: if a move is accepted, $H(E_{\rm new}) = H(E_{\rm new}) + 1$; otherwise, $H(E_{\rm old}) = H(E_{\rm old}) + 1$. As a result, the search process keeps track of the number of prior explorations of a particular energy region and biases against revisiting the same energy.

Unlike in a typical low-temperature region, the probability weight of a local minimum conformation decreases in an ELP simulation with the time the system stays in that minimum, and consequently the probability to escape from the minimum increases. Hence, ELP utilizes the interplay of two factors. Given equal frequencies $H(E, t)$, the simulation will favor low energies. This will ensure that no irrelevant conformations with large cost function E are sampled. However, soon the system will run into a local minimum. After some time, the energy landscape is locally deformed by our approach in such a way that the local minimum is no longer favored and the system will explore values of the cost function E. It will then either fall in a new and different local minimum or walk through this high-energy region until the corresponding histogram entries all have similar frequencies. At that point, the form of the original energy landscape is restored and the system again has a bias toward low energies.

In both SA and ELP, we also monitor the convergence of the diffraction pattern towards the observed pattern through the use of another quantity, the so-called residual value or R factor, defined by

$$R = \sum_{j=1}^{N_k} \big||F_{\rm obs}(\mathbf{k}_j)| - |F_{\rm cal}(\mathbf{k}_j)|\big| \Big/ \sum_{j=1}^{N_k} |F_{\rm obs}(\mathbf{k}_j)|. \qquad (8)$$

By construction, the correct structure has a residual value $R = 0$. However, we found that a less-stringent criteria can be used to determine whether the correct structure was found. We found that all configurations with $R < 0.3$ resemble closely the (known) crystal structure of our compound (Hsu *et al.*, 2001). For this reason, we choose $R < 0.3$ as a criterion to identify the correct crystal structure.

3. Results and discussion

In order to compare the previous SA results (Hsu *et al.*, 2001) with the ELP results, we take for each run of both methods the same Monte Carlo sweeps and initial configurations. In all our SA runs, the initial temperature was chosen to be 0.6 and a sequence of $N_T = 80$ temperatures was generated with cooling rate α of 0.95. At each temperature, we considered two cases: Monte Carlo sweeps $N_M = 2048$ and $N_M = 4096$. The lowest energy E and R factor were stored in our 20 SA runs, which started from 20 different randomly chosen initial configurations.

In the ELP simulations, the total number of Monte Carlo sweeps was set to $N_H = N_M \times N_T$, which ensures equal statistics for the two methods. As the lowest temperature, we chose $T_L = 0.001$, and the constant ε was set to 24, which is the number of non-H atoms in the asymmetric unit. Again, the lowest energy E and R factor were stored in our 20 ELP runs, which started from the same 20 initial conformations as used for the 20 SA runs.

In order to obtain more refined structures, we introduced in both the SA and ELP runs an additional global update which takes into account the fact that the atomic form factors of atoms C, N and O are close to one another in value and difficult to distinguish in numerical simulations. For this reason, we create randomly N_p pairs of atoms and switch the positions of the atoms in each pair once for each temperature in a SA run and every N_M Monte Carlo sweeps in the ELP run. If $\Delta E < 0$, the switch is accepted; otherwise it is rejected.

In order to visualize the different behavior of the two optimization methods, we present in Fig. 2 the time series of energy E for both SA and ELP. Fig. 3 is the time series of the R factor for the same two runs. Both runs relied on $N_H = N_M \times$

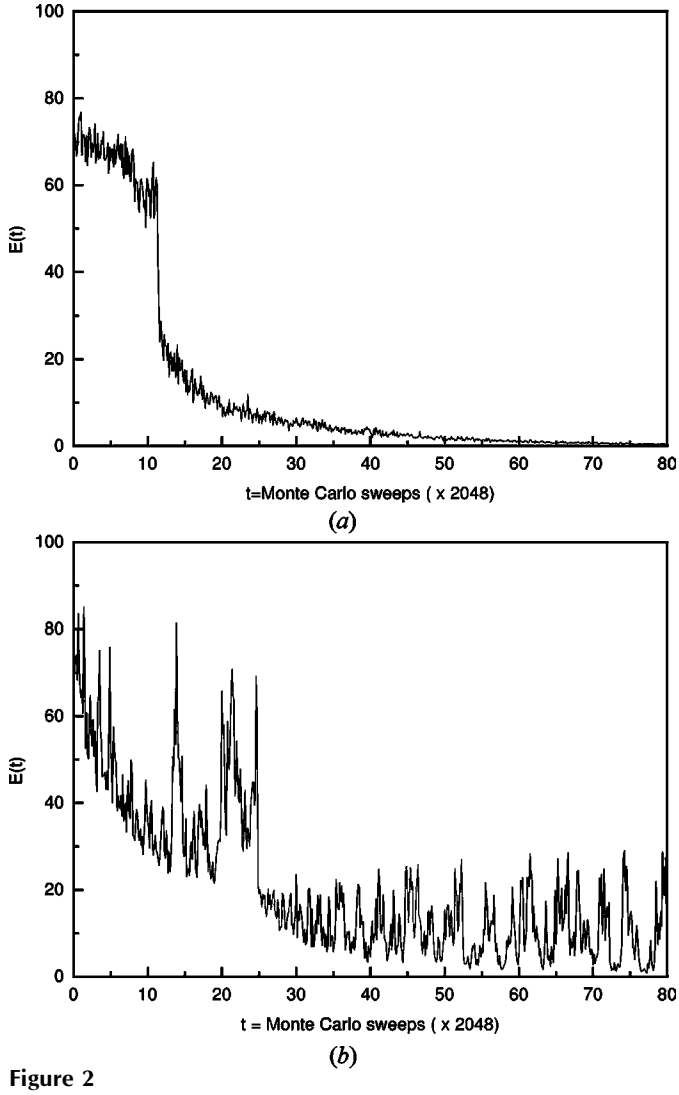

Figure 2
The time series of energy E for (a) SA and (b) ELP runs with $N_H = 2048 \times 80$ Monte Carlo sweeps.

$N_T = 2048 \times 80$ Monte Carlo sweeps, and in both runs the correct structure was found. While in SA the fluctuations decrease with Monte Carlo time (*i.e.* with decreasing temperature), the ELP run shows a different behavior. At the beginning, the simulation behaves like a low-temperature canonical simulation, and the energy decreases rapidly until it becomes trapped in a local minimum. Through the following Monte Carlo sweeps, entries in the corresponding histogram bin are accumulated and the energy landscape is locally deformed until the simulation escapes this local minimum to find a lower local minimum. This process is repeated until the simulation finds the target structure ($R < 0.3$), after 30×2048 Monte Carlo sweeps. The so-obtained lowest energy structure corresponds to the crystal structure of $C_{14}H_{11}NO \cdot C_4H_8O_4$ (see Fig. 4). Within the remaining time, the ELP run performs essentially a random walk in and out of this structure.

For a quantitative analysis, we have listed the results of the 20 SA and ELP runs in Tables 1 and 2. In 20 SA runs with a total number of $N_H = 2048 \times 80$ Monte Carlo sweeps, the correct crystal structure ($R < 0.3$) was found three times, *i.e.* with a probability of 20%. On the other hand, in 20 ELP runs with the same number of Monte Carlo sweeps, that structure was found with 65% probability (13 out of 20 runs). Hence, under this condition the performance of ELP is by a factor of four better than that of SA. Our previous work showed that it would require $N_H = 20480 \times 80$ Monte Carlo sweeps (*i.e.* an increase of the statistics by a factor of ten) to find the crystal structure with a probability of 70% (Hsu *et al.*, 2001). On the other hand, doubling the total number of Monte Carlo sweeps to $N_H = 4096 \times 80$ ensures in ELP runs that the correct structure is found with 90% probability (while the probability in SA runs increases to only 25%).

In order to find the crystal structure with similar probability in SA runs, we would have to raise the statistics to 102400×80 Monte Carlo sweeps (Hsu *et al.*, 2001). Hence, we conclude that ELP is of the order of ten times more efficient than SA in finding the crystal structure when we compare the total computer time needed to reach that structure with a set probability. Note that, in ELP runs, not only is the probability of finding the crystal structure enhanced but the new method also leads to, on average, lower energies and smaller R factors. Only when restricted to the cases where the correct structure was found does SA find, on average, lower energies than ELP. This is because SA works in its final phase (at low temperatures) as a local optimizer, improving solutions found earlier

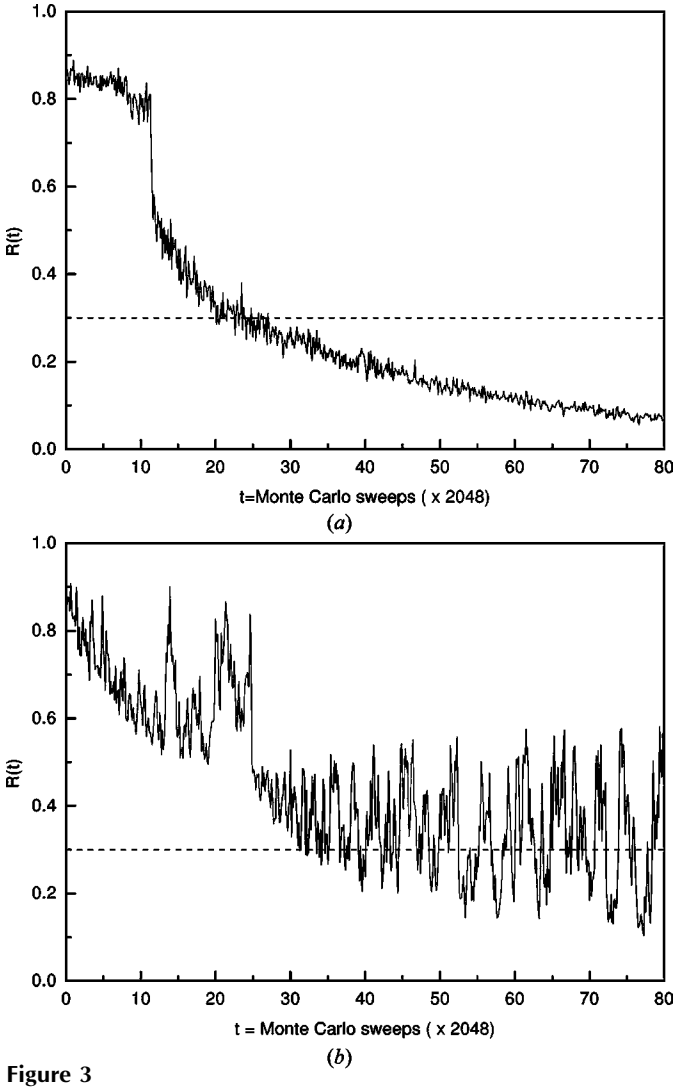

Figure 3
The time series of R factor for (*a*) SA and (*b*) ELP runs with $N_H = 2048 \times 80$ Monte Carlo sweeps. The criterion $R = 0.3$ is characterized by the dashed line.

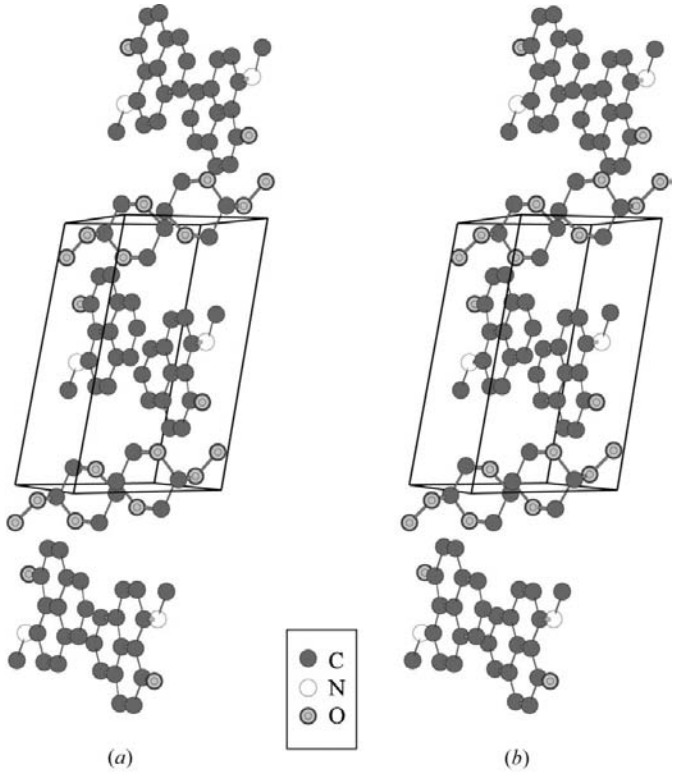

Figure 4
(*a*) The crystal structure of 9-(methylamino)-1*H*-phenalen-1-one 1,4-dioxan-2-y1 hydroperoxide solvate ($C_{14}H_{11}NO \cdot C_4H_8O_4$). The box marks the unit cell. No hydrogen atoms are shown. (*b*) The lowest energy configuration for this molecule as found in an ELP run of $N_H = 2048 \times 80$ Monte Carlo sweeps.

Table 1
The lowest values of energy E and R factor obtained by SA for the cases $N_H^1 = 2048 \times 80$ and $N_H^2 = 4096 \times 80$ with switch, and the average values of E and R for the 20 runs, ave(20), and for the runs which produce the correct structures, ave(correct).

No.	$E(N_H^1)$	$R(N_H^1)$	$E(N_H^2)$	$R(N_H^2)$
1	22.74	0.50	16.02	0.41
2	19.74	0.46	16.25	0.42
3	16.51	0.43	0.37	0.07
4	20.02	0.47	16.79	0.43
5	17.16	0.42	16.48	0.43
6	16.42	0.42	21.58	0.47
7	17.37	0.43	16.56	0.42
8	17.16	0.43	0.37	0.07
9	0.26	0.06	16.29	0.43
10	16.48	0.41	0.34	0.06
11	0.36	0.07	16.13	0.41
12	0.38	0.07	17.93	0.44
13	16.43	0.43	20.58	0.48
14	0.33	0.06	0.34	0.06
15	18.86	0.45	18.52	0.45
16	19.40	0.45	0.33	0.06
17	20.64	0.46	17.27	0.43
18	16.92	0.43	17.79	0.44
19	18.68	0.44	22.24	0.49
20	16.87	0.43	16.94	0.43
ave(20)	29.27	0.73	13.46	0.35
	±5.66	±0.15	±1.73	±0.04
ave(correct)	0.33	0.07	0.35	0.06
	±0.02	±0.00	±0.01	±0.00
%	20		25	

Table 2
The lowest values of energy E and R factor obtained by ELP for the cases $N_H^1 = 2048 \times 80$ and $N_H^2 = 4096 \times 80$ with switch, and the average values of E and R for the 20 runs, ave(20), and for the runs which produce the correct structures, ave(correct).

No.	$E(N_H^1)$	$R(N_H^1)$	$E(N_H^2)$	$R(N_H^2)$
1	1.14	0.12	3.10	0.19
2	0.77	0.10	2.86	0.18
3	0.98	0.11	1.50	0.14
4	3.05	0.19	9.31	0.32
5	1.15	0.12	0.68	0.09
6	1.29	0.12	1.70	0.14
7	0.82	0.10	0.32	0.06
8	17.30	0.43	0.42	0.07
9	5.77	0.26	0.30	0.06
10	17.74	0.43	0.35	0.07
11	3.05	0.19	0.77	0.09
12	17.15	0.43	6.35	0.27
13	0.95	0.11	0.24	0.05
14	1.21	0.12	1.79	0.15
15	1.29	0.12	4.40	0.22
16	1.04	0.11	15.74	0.40
17	16.46	0.43	5.80	0.26
18	17.47	0.43	1.24	0.12
19	16.61	0.43	0.49	0.07
20	17.54	0.43	0.55	0.08
ave(20)	7.14	0.24	2.90	0.15
	±1.67	±0.03	±0.85	±0.02
ave(correct)	1.73	0.14	1.83	0.13
	±0.38	±0.01	±0.44	±0.02
%	65		90	

in the run. This refinement effect is missing in ELP runs; however, our ELP solutions are already sufficiently close to the correct structure (see also Fig. 4) that further refinement (for instance by minimizing the structures with a standard local optimizer) seems not to be necessary. Hence, the results of Tables 1 and 2 indicate that ELP is much better suited than SA to determine crystal structures from diffraction data. We are currently investigating whether this also holds true for larger molecules, such as proteins.

4. Conclusions

To summarize, we have employed two methods, simulated annealing and energy landscape paving, to determine the crystal structure of an organic compound from simulated X-ray diffraction data comprising integrated intensities. Our results indicate that ELP is more efficient (by an order of ten) in finding the crystal structure than the simulated-annealing approach.

This work was partially supported by the National Science Foundation under grant No. CHE-9981874. Part of this article was written while UH was a visitor at the Computing Center of Academia Sinica, Taipei, Taiwan. He thanks all members of the center for their hospitality during his stay in Taipei.

References

Andreev, Y. G. & Bruce, P. G. (1998). *J. Chem. Soc. Dalton Trans.* **24**, pp. 4071–4080.
Andreev, Y. G., Lightfoot, P. & Bruce, P. G. (1997). *J. Appl. Cryst.* **30**, 294–305.
Andreev, Y. G., MacGlashan, G. S. & Bruce, P. G. (1997). *Phys. Rev. B*, **55**, 12011–12017.
Attfield, J. P., Sleight, A. W. & Cheetham, A. K. (1986). *Nature (London)*, **322**, 620–622.
Beurskens, P. T., Admiraal, G., Beurskens, G., Bosman, W. P., Gelder, R. de, Israél, R. & Smits, J. M. M. (1994). *The DIRDIF Program System*, Technical Report, Crystallography Laboratory, University of Nijmegen, The Netherlands.
Chen, Y. & Su, W. P. (2000). *Acta Cryst.* A**56**, 127–131.
Cirujeda, J., Ochando, L. E., Amigó, J. M., Rovira, C., Rius, J. & Veciana, J. (1995). *Angew. Chem. Int. Ed. Engl.* **34**, 55–57.
Clearfield, A., McCusker, L. B. & Rudolf, P. R. (1984). *Inorg. Chem.* **23**, 4679–4682.
Dinnebier, R. E., Stephens, P. W., Carter, J. K., Lommen, A. N., Heiney, P. A., McGhie, A. R., Brard, L. & Smith, A. B. III (1995). *J. Appl. Cryst.* **28**, 327–334.
Geman, S. & Geman, D. (1984). *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**, 721–741.
Hansmann, U. H. E. & Wille, L. (2002). *Phys. Rev. Lett.* **88**, 068105-1–068105-4.
Harris, K. D. M. & Tremayne, M. (1996). *Chem. Mater.* **8**, 2554–2570.
Harris, K. D. M., Tremayne, M., Lightfoot, P. & Bruce, P. G. (1994). *J. Am. Chem. Soc.* **116**, 3543–3547.
Hauptman, H. A. (1986). *Science*, **233**, 178–183.
Hsu, H. P., Lin, S. C. & Hansmann, U. H. E. (2001). *Phys. Rev. E*, **64**, 56707.
Ibers, J. & Hamilton, W. C. (1974). Editors. *International Tables for X-ray Crystallography*, Vol. IV. Birmingham: Kynoch Press. (Present distributor: Kluwer Academic Publishers, Dordrecht.)
Kariuki, B. M., Zin, D. M. S., Tremayne, M. & Harris, K. D. M. (1996). *Chem. Mater.* **8**, 565–569.
Kraus, W. & Nolze, G. (1999). *PowderCell for Windows, V2.1*, Federal Institute for Materials Research and Testing, Berlin, Germany.
Ladd, M. F. C. & Palmer, R. A. (1980). Editors. *Theory and Practice of Direct Methods in Crystallography*. New York: Plenum Press.
Lightfoot, P., Cheetham, A. K. & Sleight, A. W. (1987). *Inorg. Chem.* **25**, 3544–3547.

Lightfoot, P., Glidewell, C. & Bruce, P. G. (1992). *J. Mater. Chem.* **2**, 361–362.
Louër, D. & Louër, M. (1987). *J. Solid State Chem.* **68**, 292–299.
Louër, M., Brochu, R., Louër, D., Arsalane, S. & Ziyad, M. (1995). *Acta Cryst.* B**51**, 908–913.
Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.
Reck, G., Kretschmer, R.-G., Kutschabsky, L. & Pritzkow, W. (1988). *Acta Cryst.* A**44**, 417–421.
Sheldrick, G. M. (1993). *SHELXL93*. University of Göttingen, Germany.
Su, W. P. (1995*a*). *Acta Cryst.* A**51**, 845–849.
Su, W. P. (1995*b*). *Physica (Utrecht) A*, **221**, 193–201.
Tremayne, M., Kariuki, B. M. & Harris, K. D. M. (1996). *J. Appl. Cryst.* **29**, 211–214.
Tremayne, M., Kariuki, B. M. & Harris, K. D. M. (1997). *Angew. Chem. Int. Ed. Engl.* **36**, 770–772.
Weeks, C. M., Hauptman, H. A., Smith, G. D, Blessing, R. H., Teeter, M. M. & Miller, R. (1995). *Acta Cryst.* D**51**, 33–38.
Woolfson, M. M. (1987). *Acta Cryst.* A**43**, 593–612.
Yatsenko, A. V., Tafeenko, V. A., Paseshnichenko, K. A., Stalnaya, T. V. & Solodar, S. L. (1998). *Z. Kristallogr.* **213**, 466–472.
Zimmer, M. F. & Su, W. P. (1998). *Phys. Rev. E*, **58**, 5131–5142.