

Molecular structure matching by simulated annealing. II. An exploration of the evolution of configuration landscape problems

M.T. Barakat and P.M. Dean

Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QJ, U.K.

Received 13 September 1989

Accepted 30 November 1989

Key words: Simulated annealing; Molecular matching; Molecular similarity; Dihydrofolate reductase; Configuration landscape

SUMMARY

This paper considers some of the landscape problems encountered in matching molecules by simulated annealing. Although the method is in theory ergodic, the global minimum in the objective function is not always encountered. Factors inherent in the molecular data that lead the trajectory of the minimization away from its optimal route are analysed. Segments comprised of the C_z atoms of dihydrofolate reductase are used as test data. The evolution of a reverse ordering landscape problem is examined in detail. Where such patterns in the data could lead to incorrect matches, the problem can in part be circumvented by assigning an initial random ordering to the molecules.

INTRODUCTION

In the previous paper [1], a method was presented for molecular matching using simulated annealing to minimize the difference distance matrix, derived from the atomic Cartesian coordinates, for two molecules. This is a general method and has the great advantage that large combinatorial problems can be handled with relative ease. For example, two molecules composed of 150 atoms can be matched by finding atom correspondences through the permutation set of 150; brute force procedures would require 10²⁶³ comparisons whereas simulated annealing samples a small number and effectively reduces the combinatorial problem to an ordering problem. The method of simulated annealing differs from iterative improvement optimization techniques by allowing uphill excursions to prevent trapping in local minima. In theory, the method is ergodic and should find the global minimum of any set of values in discrete space. However, in practice, there are circumstances where the method does not appear to be consistently ergodic.

The most efficient algorithm found in the previous paper was that employing the cooling schedule of Aarts et al. [2]. This method dynamically changes the temperature factor in response to the variation encountered in the objective function of each Markov chain as the algorithm proceeds. The performance of this algorithm in matching two sets of coordinates can be tested exactly by using identical, but differently ordered, coordinate sets. Extensive tests with random coordinates for problem sizes of 20, 70 and 150 points showed that the method perfectly matched all 10 test runs for 20 points; 8 out of 10 runs produced perfect matches for 70 points; and 7 out of 10 runs gave exact matches for 150 points. Careful scrutiny of the correspondence assignments for those runs which did not give a perfect match revealed that the failure was not due simply to getting a few correspondences wrong; mismatches were drastically different assignments. A similar problem was encountered with matches between molecular C_α atomic data for DHFR *Lactobacillus casei*. When half the molecule was matched against the whole molecule not all matches were perfect.

It can readily be shown that simulated annealing is ergodic; convergence proofs are given by van Laarhoven and Aarts [3]. Two explanations can be offered for non-ergodicity. Firstly, the algorithm may not converge because the final acceptance criterion has not been reached, although the trajectory may be on the way to a correct global minimum. This problem can be circumvented by minor adjustments to the program and is an annoyance rather than a difficulty. Secondly, incorrect solutions may be encountered which are the result of serious landscape problems; in this case the solutions are badly wrong and steps have to be taken to reduce the probability of these difficulties occurring. The issue of the configurational landscape in discrete optimization problems has rarely been addressed in the literature. Kirkpatrick and Toulouse [4] investigated the configuration landscape for long-range Ising spin glasses and the travelling salesman problem; they showed that at low temperatures the space is ultrametric. Landscape problems can be envisaged to be analogous to the following configuration. Consider a golf course with small hillocks and valleys but with very deep holes. It is easy to imagine that even with hill climbing an algorithm could be trapped in a deep hole, but not necessarily the deepest, and a false minimum produced since only a small fraction of the total configuration space can be sampled and the probability of getting out of a deep hole is small.

In molecular matching, configuration landscape problems may arise from elements of symmetry within the structures or, in the case of proteins, from regions of similar secondary structure showing correspondence. These difficulties are identified and analysed in this paper.

COMPUTING METHODS

The cooling schedule of Aarts et al. [2] employed in the previous paper [1] has been used in this study. The objective function is scaled, the length of the Markov chain is given by

$$l_M = 2 N_A(N_B - 1) \quad (1)$$

where N_A and N_B are respectively the number of atoms in molecules A and B and the initial temperature is set at 2.0. Computing was carried out on an IBM 3084Q. Where molecular matches have been obtained, the corresponding atom assignments were used in the McLachlan [5] algorithm to generate the best matched coordinates. Ribbon diagrams of protein structure were drawn using SYBYL on an Evans and Sutherland PS390.

Two types of input data are used here to analyse the configuration landscape problem: (a) molecular data taken from the crystal structure of DHFR (*L. casei*) [6] using the coordinates for the C_α atoms, and (b) almost linear data of 24 points ($N_A = 24$) where an end point is not colinear with the others but is located normal to the line at position 23.

RESULTS

Landscape problems with DHFR

In Table 8 of the previous paper [1], segments from DHFR (*L. casei*) C_α coordinate sets, where the sequence order for A was retained, were matched with a random order for the C_α atoms of B. The set N_A numbered 1–81 produced 5 perfect matches out of 10, but the 5 mismatches were very similar to each other, suggesting that the configuration landscape had two deep valleys, the global and the secondary minimum.

The mismatched subset of 1–81 (set A) atoms superimposed on the full set of 162 atoms (set B) is shown in Fig. 1a; a ribbon diagram illustrating the secondary structure is shown in Fig. 1b. Although the ribbon diagrams appear to show a very badly matched structure, this could be deceptive since many of the C_α positions appear to be closely paired. However, the real test lies in the rms separation; an rms of 4.0 Å denotes a reasonable match and could account for a deep secondary minimum in the landscape. Segment 1–81 contains two helices, Leu²³–Thr³⁴ and Gly⁴²–Phe⁴⁹. Portions of these helices are mismatched. The first helix in set A is mismatched by residues 44–49 in set B and the second helix in set A is matched by 27–33 in set B.

The set numbered 82–162 exhibited 4 matches at the global minimum; mismatches fell into two groups with very different alignments. The mismatches are illustrated as atom positions in Figs. 2 and 3 together with the corresponding ribbon diagrams to show secondary structure. No correlations in secondary structure alignments were immediately obvious with the 82–162 segment.

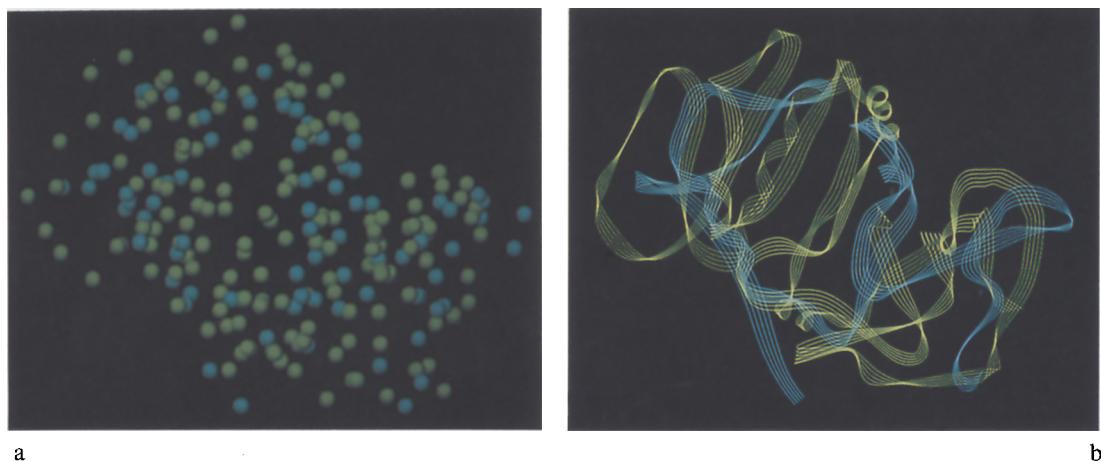


Fig. 1. Mismatch between segment 1–81 (blue) and 1–162 (yellow) C_α -atoms of DHFR *L. casei*. Rms separation = 4.0 Å. a: Atom positions; b: Ribbon diagram of polypeptide backbone.

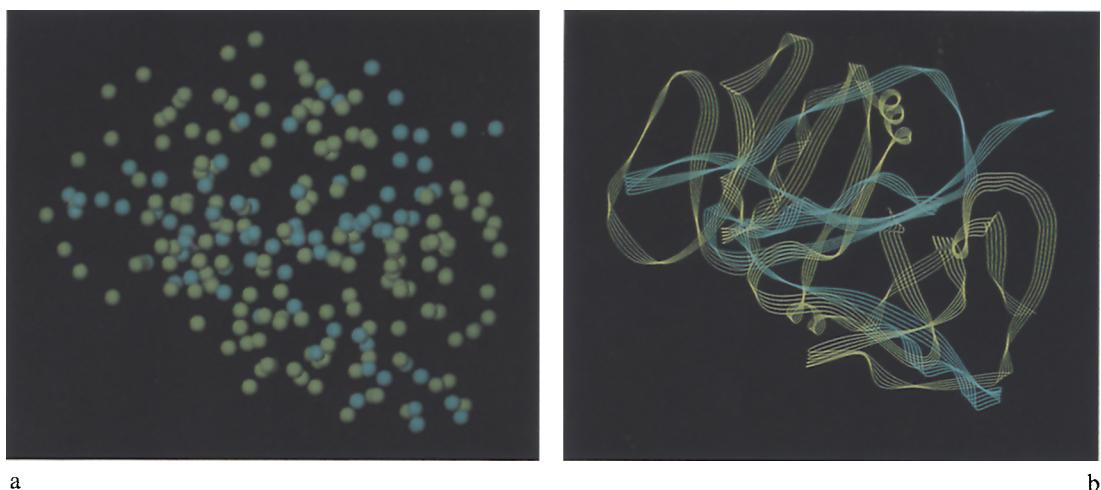


Fig. 2. Mismatch between segment 82–162 (blue) and 1–162 (yellow) C_α-atoms of DHFR *L. casei*. Rms separation = 11.7 Å.
a: Atom positions; b: Ribbon diagram of polypeptide backbone.

The rms values for the mismatches are 11.6 Å and 12.1 Å suggesting that these local minima represent very poor matches.

Landscape problems with 20 C_α atom segments of DHFR

The discovery of a landscape problem, with a subset of 81 atoms matched against the full set of 162 atoms, prompted further examination with smaller consecutive 20-atom segments using identical coordinate data but with the order for B randomised. Table 1 illustrates the results for

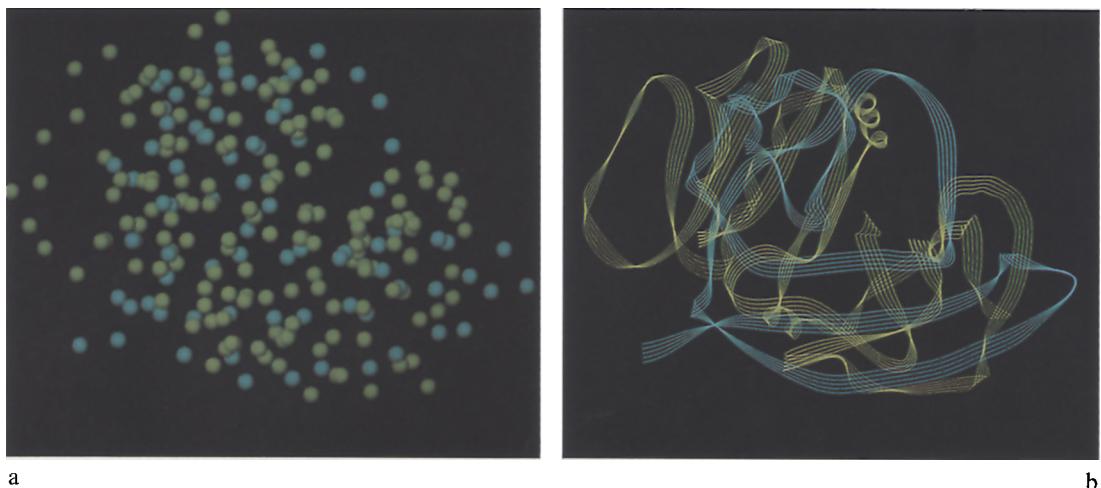


Fig. 3. Mismatch between segment 82–162 (blue) and 1–162 (yellow) C_α-atoms of DHFR *L. casei*. Rms separation = 12.1 Å.
a: Atom positions; b: Ribbon diagram of polypeptide backbone.

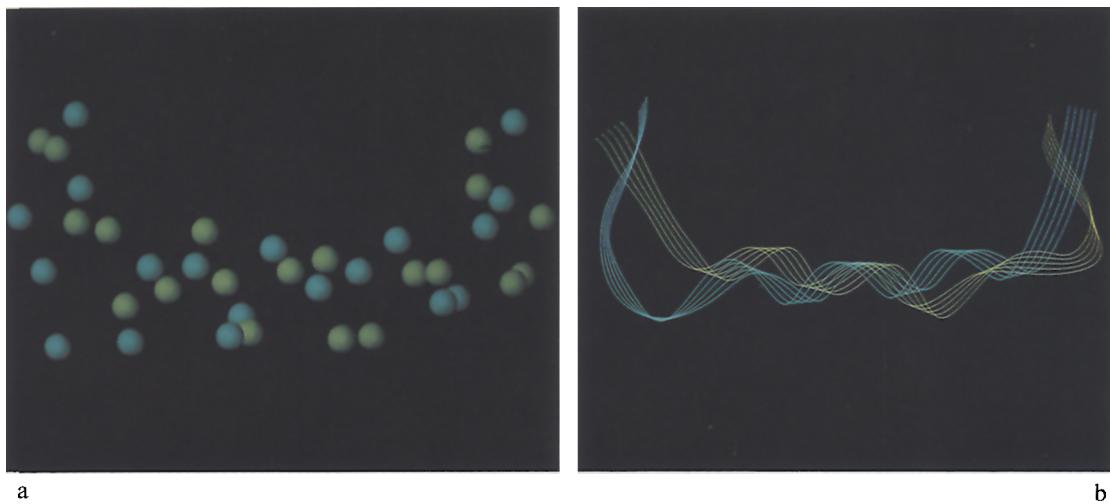


Fig. 4. Mismatch between the 20 atom segment $C_\alpha 21\text{--}40$ against itself. Rms separation = 2.6 Å. a: Atom positions; b: Ribbon diagram of polypeptide backbone.

each segment averaged over 10 runs. Segments in the range 21–80 and 121–160 produced occasional mismatches; all other segments produced consistently correct assignments in all 10 runs.

Segment 21–40 contains the long helix from Leu²³–Thr³⁴. Figures 4a and b show the C_α positions together with the ribbon diagram of secondary structure. It can be seen that the direction of the helix is reversed in the mismatch. All the mismatches of the segment show a similar reversal. Pieces of the segment at the end of the helix are positioned similarly to ‘fool’ the algorithm. The rms separation is 2.6 Å.

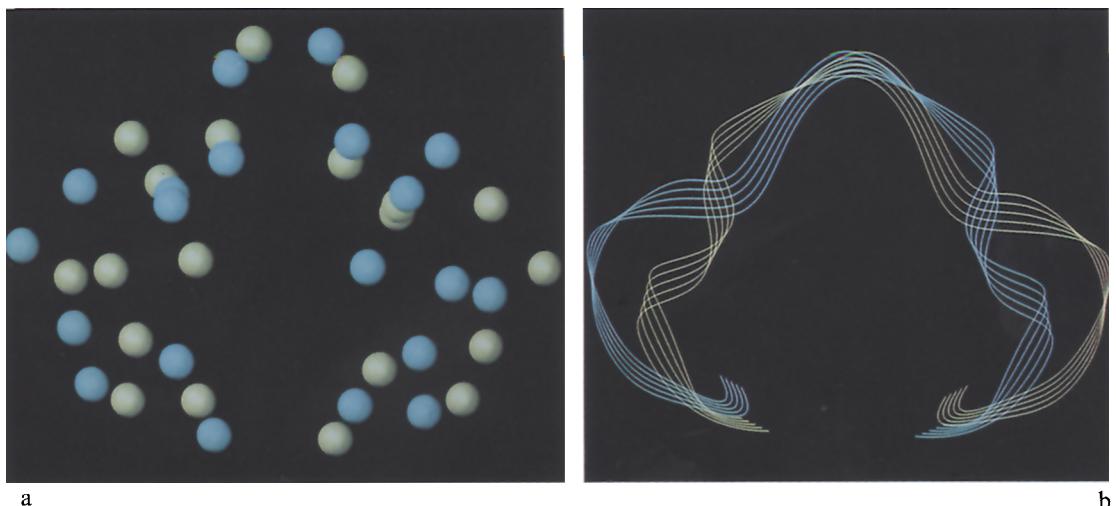


Fig. 5. Mismatch between the 20 atom segment $C_\alpha 41\text{--}60$ against itself. Rms separation = 2.8 Å. a: Atom positions; b: Ribbon diagram of polypeptide backbone.

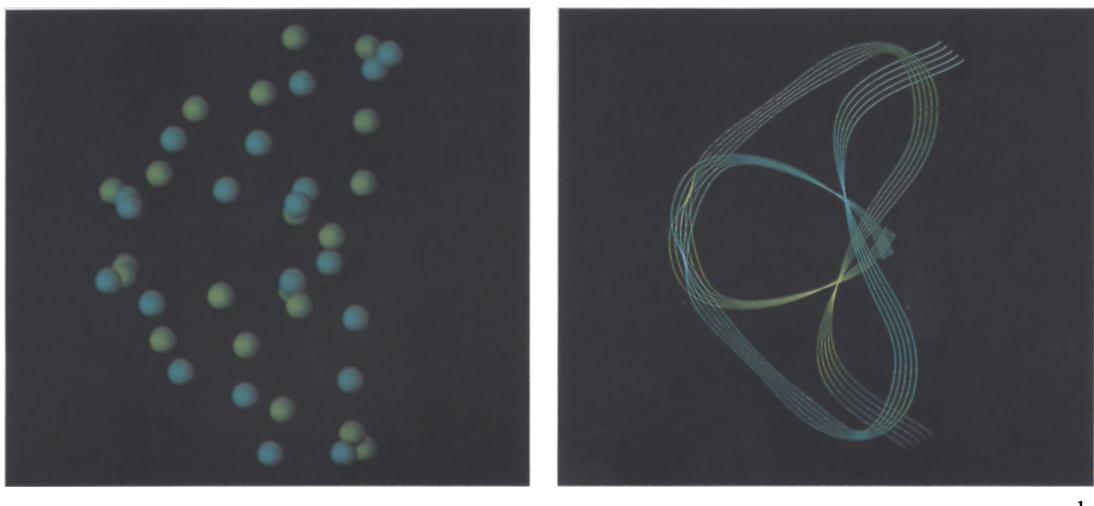
TABLE 1
ANNEALING DATA USING THE AARTS et al. [2] DYNAMIC TEMPERATURE FACTOR^a

C _α -atoms in set A	C _α -atoms in set B	Initial DDM statistic	Final DDM statistic	Correct assignments (%)
1–21	1–21	0.242	0.000	100
21–40	21–40	0.220	0.025	40.0
41–60	41–60	0.241	0.026	74.0
61–80	61–80	0.200	0.022	72.0
81–100	81–100	0.223	0.000	100
101–120	101–120	0.303	0.000	100
121–140	121–140	0.343	0.046	50.0
141–160	141–160	0.295	0.012	90.0

^aSegments of 20 C_α-atoms from *L. casei* DHFR were matched. The order of segment B was randomised initially (that for A remained ordered). A scaling constant of C=8 was used. The values represent the averages over 10 runs.

Segment 41–60 contains a helix from Gly⁴²–Phe⁴⁹, a turn from Lys⁵¹–Leu⁵⁴ and a sheet from Glu⁵⁶–Thr⁶³. There appears to be an axis of pseudo-symmetry passing through the turn (Figs. 5a and b). This results in the mismatch being a reversal of the ordering of the two structures with an rms separation of 2.8 Å.

Segment 61–80 gives a poor mismatch with few parts of the structure appearing to be superimposed (Figs. 6a and b). This segment contains a portion of a sheet up to Thr⁶³, a turn from Gln⁶⁵–Tyr⁶⁸ followed by a second turn from Ala⁷⁰–Ala⁷³; a small sheet follows from Val⁷⁴–Val⁷⁶ and the segment ends with a helical portion from Asp⁷⁸. The mismatch is a reverse ordering. The rms separation between the two structures is 4.2 Å.



a

b

Fig. 6. Mismatch between the 20 atom segment C_α61–80 against itself. Rms separation = 4.2 Å. a: Atom positions; b: Ribbon diagram of polypeptide backbone.

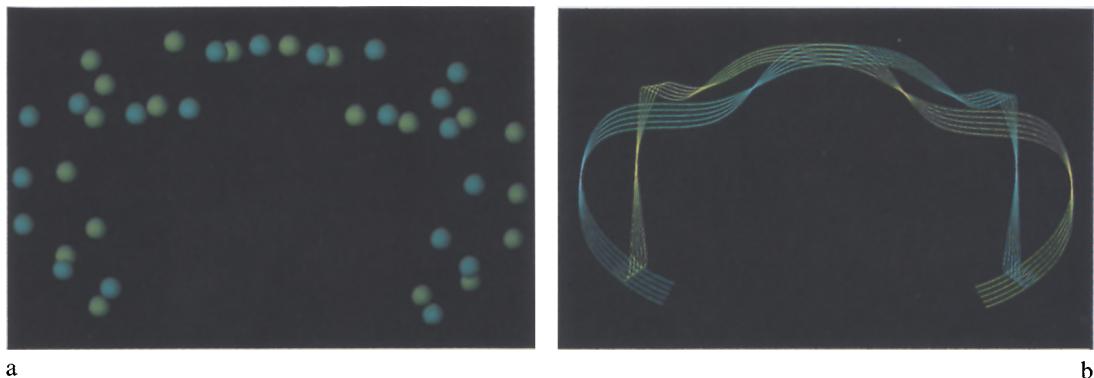


Fig. 7. Mismatch between the 20 atom segment $C_\alpha 121\text{--}140$ against itself. Rms separation = 3.1 Å. a: Atom positions; b: Ribbon diagram of polypeptide backbone.

Segment 121–140 contains a turn with a loop and the mismatch is a sequence reversal with a long section of atoms superimposed (Figs. 7a and b). The rms separation is 3.1 Å.

Segment 141–160 contains a portion of a sheet up to Val¹⁴⁴, three linked turns from Asp¹⁴⁶–Thr¹⁵² followed by a section of a sheet. Figures 8a and b show that the two segments are reversed by a hairpin turn so that the long strands are positioned parallel; the rms separation is 4.6 Å.

The principal feature of all the mismatches obtained with these five 20-atom segments taken from DHFR *L. casei* is that each mismatch is caused by a different ordering which results in an effective rotation of the segments through 180° round an axis of pseudo-symmetry. This characteristic incorrect ordering gives rise to a low value for the objective function in a significant number of runs. In order to understand how these mismatches arise, we have designed a simple test data set which consists of 24 points, 23 points are equally spaced and colinear, the 24th point is positioned orthogonal to the line at position 23 and the distance between 23–24 is the same as that between other adjacent points. This data set forms structure A and is termed the ‘bent line’; structure B is identical to A so that after annealing the global minimum will have A and B ordered identically; the other nearest local minimum will have A and B in the reverse order. The following experiments were performed to study the evolution of mismatches.

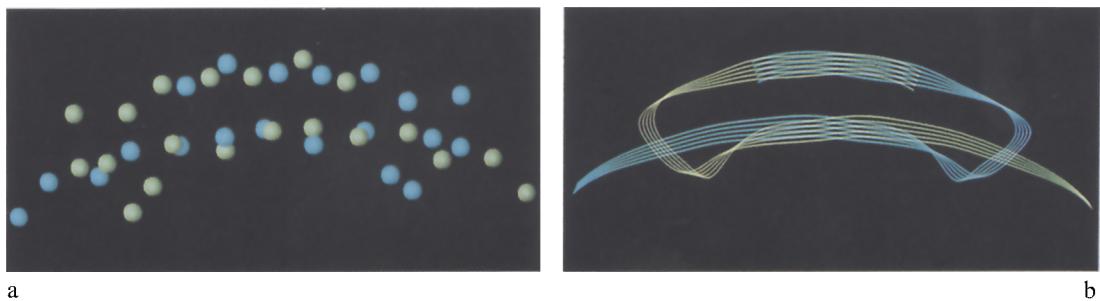


Fig. 8. Mismatch between the 20 atom segment $C_\alpha 141\text{--}160$ against itself. Rms separation = 4.6 Å. a: Atom positions; b: Ribbon diagram of polypeptide backbone.

Evolution of the landscape problem

The configuration landscape for 80 successive configurations ($N_A = N_B = 24$) starting from a random order for B, is shown in Fig. 9. The value for the objective function oscillates between 265–300 whilst all configurations are accepted. It is evident that any annealing algorithm has to cope with oscillations in the objective function. However, if the ‘hills’ in the landscape are exceptionally large the configuration may become trapped. Figures 10, 11 and 12 show key determinants for the annealing profile for two different starting conditions using the bent line test data.

In Fig. 10a the order for B is initially randomised; the average value of the objective function at the end of each Markov chain exhibits constant oscillations as the annealing progresses. The final result is a mismatch with the order for B reversed compared with A. In contrast, the profile shown in Fig. 10b has the initial order for B reversed. The average value for the objective function shows little change over the temperature range 2.0–0.75 but thereafter declines sharply with large oscillations. The final order has B reversed compared with A.

The standard deviation of the objective function for both trajectories is plotted in Figs. 11a and b. The two profiles are distinct from each other. The peaks in Fig. 11a approximately follow the profile shown in Fig. 10a. A different profile for the annealing of an initial reversed ordering is shown in Fig. 11b. Maximum deviation occurs in the transition phase at around a temperature of 0.5.

The acceptance ratio for the two profiles is illustrated in Figs. 12a and b. Once again the profiles mirror those of Fig. 10. However, there is a major difference, by a factor of two, in the acceptance ratio at the beginning of the cooling schedules. The profile with the high acceptance ratio remains

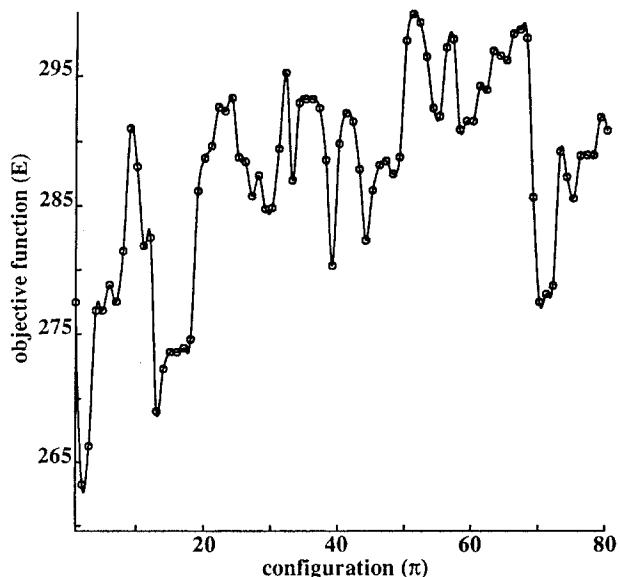


Fig. 9. The configuration landscape profile of the objective function for 80 configurations of the bent line problem without annealing.

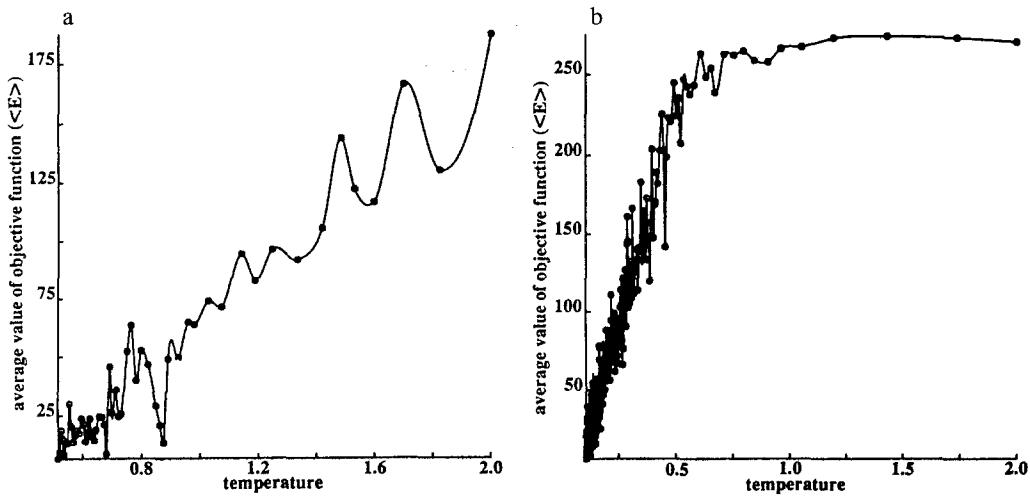


Fig. 10. The decrease in the average value of the objective function as the temperature is lowered ($C = 8$; $N_A = N_B = 24$; the coordinate set for A is identical to B). The points form a bent line with 23 linear points. a: The order for set B was initially randomised; b: The order for set B was initially the reverse of A.

effectively unchanged in the range $T = 2.0$ to $T = 0.75$, where the initial ordering is reversed (Fig. 12b).

The acceptance probability, and hence the acceptance ratio, depends on the Boltzmann probability distribution.

$$P(s'|s) = e^{-\Delta E C / 3k_B T} \quad (2)$$

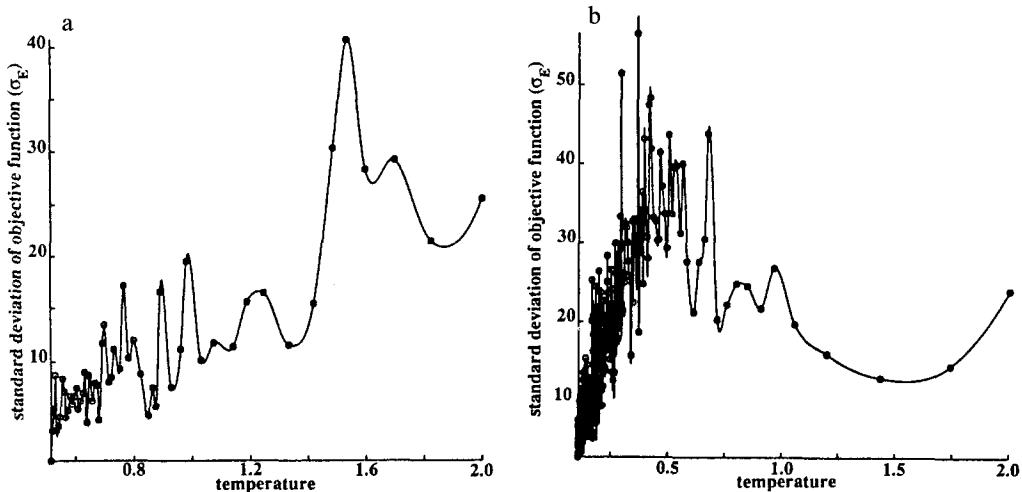


Fig. 11. The variation in the standard deviation of the objective function with temperature ($C = 8$; $N_A = N_B = 24$; the coordinate set for A is identical to B). The points form a bent line with 23 linear points. a: The order for set B was initially randomised; b: The order for set B was initially the reverse of A.

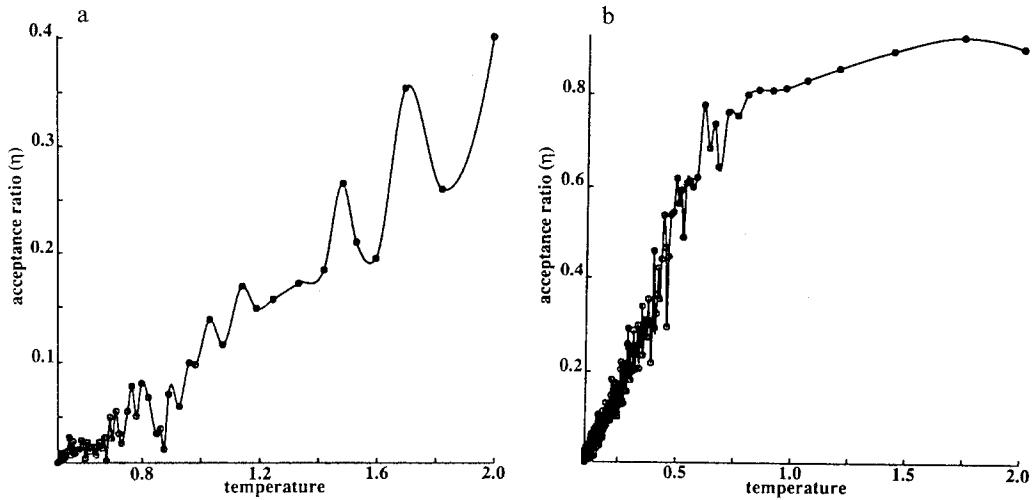


Fig. 12. The change in the acceptance ratio with temperature ($C = 8$; $N_A = N_B = 24$; the coordinate set for A is identical to B). The points form a bent line with 23 linear points. a: The order for set B was initially randomised; b: The order for set B was initially the reverse of A.

If T and C are held constant, only ΔE and $\sigma_{\Delta E}$ will affect the acceptance probability. The variation in $\sigma_{\Delta E}$, for the two conditions of random or reverse orders for set B, has been analysed in detail for different numbers of points in the bent line ($N_A = N_B$; $N_A = 20, 30, 40, \dots, 160$), where the last point is non-colinear. Fig. 13 shows the regression lines for $\sigma_{\Delta E}$ against N_A for the two different conditions; line *a* has a random order for B, line *b* has a reverse order for B; each point is derived from a population of 100000 values of ΔE . The data points along each line are well correlated. The gradients of both lines are significantly different. The slope of line *b* is approximately 10 times that of line *a*. Substitution in Eq. (2) for positive values of ΔE and $\sigma_{\Delta E}$ enables us to predict the total acceptance probability for the two different orderings. Consider the following two cases:

Case a: let $\Delta E = 3$, $T = 2$, $C = 8$, $\sigma_{\Delta E} = 1$, with set A ordered and B random, then the initial acceptance at a level of $+ 3\sigma_{\Delta E}$ is found to be 0.018 and the total acceptance probability = 0.509.

Case b: let $\Delta E = 3$, $T = 2$, $C = 8$, $\sigma_{\Delta E} = 10$, with set A ordered and B reversed, then the initial acceptance at a level of $+ 3\sigma_{\Delta E}$ is found to be 0.670 and the total acceptance probability = 0.835.

In the annealing subroutines, the standard deviation of ΔE is initially calculated using $N_A(N_B - 1)$ values before annealing. These values are obtained by performing random pair swaps on set B, each time returning the order of set B to its initial state. For case a, where set A is ordered and B starts in a random order, the first value for E is likely to be high. After one swap the value for E will also be high, so ΔE will be small; subsequent ΔE values will be of similar size, so $\sigma_{\Delta E}$ will be small and ΔE values may also be negative. The average ΔE for case a is approximately zero. On the other hand, with case b (B is initially the reverse order) the first value of E will be close to zero. After one swap, the value for E will be higher, and ΔE will be large and positive. Subsequent ΔE values will vary according to the following pattern: if two consecutive points in B are swapped, ΔE will be small; if two widely separated points are swapped, ΔE will be large. The

average ΔE for case b is always positive. Therefore, after $N_A(N_B - 1)$ swaps, $\sigma_{\Delta E}$ would be larger in case b than in case a.

We now consider the annealing trajectory for the two identical sets of the bent line ($N_A = 24$) where B starts as a reversal of the ordering of A. Even though the initial acceptance of three standard deviations of positive ΔE values in the first Markov chain is high (0.67), all negative ΔE values created cause a swap back towards the original reverse ordering. Therefore, there will always be a considerable retention of configurations which show a reversal trend in the ordering. From the outset this would lock the trajectory towards a reversal configuration for the bent line problem. On the other hand, if B starts in a random order, then the initial acceptance probability for positive ΔE values is extremely small at 0.018. In this case annealing would embark on a particular final trajectory at an early stage to generate either a correct match or a reversal mismatch with approximately equal probability. Multiple tests on 10 runs using these two starting conditions are illustrated in Table 2. Random initial orderings for B gave 60% correct assignments with a marked reduction in the initial DDM statistic. No correct assignments were generated from an initial reversed ordering for B despite its higher acceptance; all the final assignments were reversals of A and the DDM statistic was unchanged.

A similar line of reasoning should hold for any configuration landscape problem where a reversal of a correctly matched ordering gives a solution near to the global minimum in the objective function. For example, it was noticed that in the *L. casei* DHFR matching, segment 21–40 C_x was either exactly matched with itself or had a predominantly reversed order. Multiple annealing tests with random or reversed initial orders for B are illustrated in Table 2. As predicted, no correct assignments from 30 runs were produced from a reversed initial ordering. In contrast, about 50% of the solutions were correct matches if a random initial ordering was used.

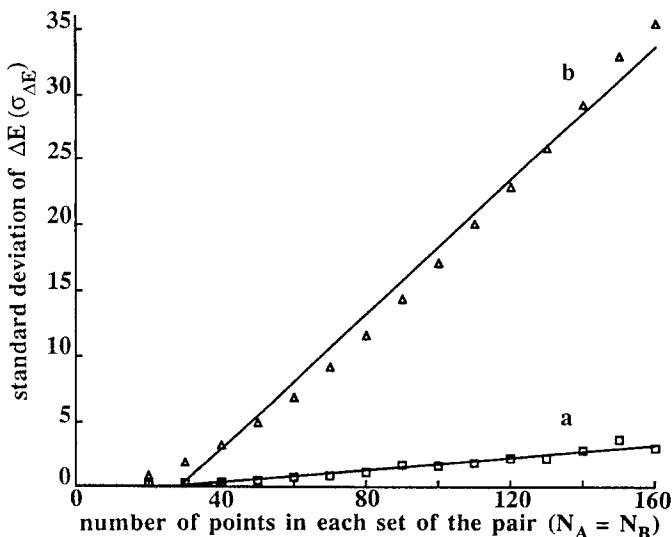


Fig. 13. Regression lines for the relation between the standard deviation of ΔE and the number of points in a straight line (N_A). $N_A = N_B$ in each situation and 100000 ΔE values (with no annealing) were computed at each value of N_A . Line a , is the regression line with set B initially randomised. The regression equation for the line is $\sigma_{\Delta E} = 0.0242 (\pm 0.0016) N_A - 0.6459 (\pm 0.1584)$; ($r = 0.9773$). Line b , is the regression line with set B initially in the reversed order to A. The regression equation for the line is $\sigma_{\Delta E} = 0.2586 (\pm 0.0090) N_A - 7.389 (\pm 0.9009)$; ($r = 0.9922$).

TABLE 2
ANNEALING DATA USING THE AARTS et al. [2] DYNAMIC TEMPERATURE FACTOR^a

Points in set A	Points in set B	Initial ordering for B	Initial DDM statistic	Final DDM statistic	Correct assignments (%)
'bent' line	'bent' line	random	0.243	0.005	60.0
'bent' line	'bent' line	reversed	0.011	0.011	0.0
21–40 C _α	21–40 C _α	random	0.222	0.038	53.3
21–40 C _α	21–40 C _α	reversed	0.053	0.048	0.0

^aThe results of matching a 'bent' line structure (24 points) with itself are shown; the values represent the averages over 10 runs. The 21–40 C_α-atom segment from *L. casei* DHFR was also investigated over 30 runs, comparing it with itself. The order of segment B was either randomized or reversed initially (that for A remained ordered). A scaling constant of C=8 was used.

DISCUSSION

Simulated annealing is a very effective method for minimizing a function in discrete space. Our intention here has not been to detract from the value of simulated annealing, but to highlight some of the pitfalls that could lead the unwary into prematurely abandoning the method as a useful tool in molecular matching problems. The preceding paper demonstrated the great power of simulated annealing in solving very large combinatorial optimization problems. It is the occasional failures that need to be examined so that we can firstly understand why a non-optimum trajectory has arisen, and secondly take steps to reduce the occurrence of these intermittent failures. Although in principle the method is ergodic if the system is cooled slowly from an infinitely high temperature, in practice, where there are restrictions on computing time, short cuts have to be introduced which can lead to non-ergodicity. Non-ergodicity arises from a difficult configuration landscape where the minimization becomes trapped in a deep well close to that of the global minimum. This paper has drawn attention to one type of landscape problem where the molecular structures being matched have an approximate rotational axis of symmetry. Consistent failures to find the global minimum can arise from sequential numbering schemes with linked, or bonded, atoms. However, if the numbering system is randomised, then some landscape problems can be circumvented and the annealing algorithm, even in a landscape intentionally designed to be difficult, can find the global minimum in 50% of the tests.

The discovery of a significant landscape problem in molecular matching occurred whilst attempting to match the first half of DHFR with the whole of the enzyme. Mismatches were produced at a rate of 50% and were essentially the same wrong atom assignments. It appeared that portions of helices were mismatched, although it is not certain whether the evolving mismatch was seeded by incorrect helix assignments. The mismatch produced by the first half of the enzyme was much better, rms separation $\sim 4 \text{ \AA}$, than that generated by the second half, rms $\sim 12 \text{ \AA}$.

Extensive explorations with smaller 20 C_α-atom segments of the enzyme showed that the structures possessing approximate rotational symmetry generate a configuration landscape problem and lead to mismatches with a reverse ordering. Thus the following secondary structural arrange-

ments can give rise to problems: a central helix, β -strand-turn- β -strand, helix-turn- β -strand and hairpin turns; other combinations of secondary structure could also be envisaged to cause configuration landscape problems through the formation of a pseudo-dyad axis. The ribbon diagrams of secondary structure, although containing all the main chain atoms and not just the C_α -atoms, are an aid in revealing how the mismatches have occurred in the small 20-atom segments. However, in the larger 81-atom segments the ribbon diagrams, because they emphasize linkages between connected atoms, direct attention from the atom positional correspondences and appear to us to be less useful in throwing light on large configuration landscape problems. The difficulty in generating the correct assignments for two helical matches can readily be appreciated since the distance matrices of the helical sections will be close to identical whether they are matched correctly or in a reverse sequence.

Experiments with identical sets of the bent line were performed to monitor how the algorithm dealt with data sets with a strong axis of symmetry. It is interesting to note how the standard deviation of ΔE (computed at the start of the annealing process) depends on the initial position of set B. When set B starts in the reversed order to set A, the acceptance ratio is higher than that for a randomly arranged set B. The algorithm is therefore attempting to counteract the difficult landscape when presented with an initial orientation favouring a reversal. However, even though the acceptance of three standard deviations of positive ΔE values in the first Markov chain is high (0.67), all the runs give reversed matches at the end. These findings suggest that, unless set B is randomised at the outset, the scaling constant (C) will have to be reduced to allow an initial acceptance probability close to 1. One further development for the program to overcome the possibilities of reversed matches would be to store the final match and reanneal the reverse order of this match from a temperature of 1.5. This procedure would produce a second match. The minimum of the two matches should be in the correct order.

No attempt has been made here to study the detailed mathematical properties of the configuration landscape of molecular matching problems. In future work it will be important to search for signatures of ultrametricity. A metric space has a distance, d , which obeys the triangular inequality

$$d(A,C) \leq d(A,B) + d(B,C) \quad (3)$$

for the three points A, B, and C.

In an ultrametric space there is a stronger inequality

$$d(A,C) \leq \max \{d(A,B), d(B,C)\} \quad (4)$$

The triangles are either equilateral or isosceles with small base. The configuration landscape for the travelling salesman problem has been proved to be ultrametric [4]. Two local minima are separated by a barrier between them which is an ultrametric distance. The phenomenon of ultrametricity in molecular matching may help us to understand the problem of mismatching. What has to be established in the analysis of the configuration landscape is how this distance between two configurations is related to the number of moves separating them. It might be possible to use cluster analysis to chart the evolution of particular configurations by constructing the dendrogram at the end of each Markov chain. Comparison between successive dendograms should reveal

whether certain intermediate configurations force the trajectory into a particular final configuration. For example, this procedure should enable us to distinguish whether or not the reversed helical sections in the mismatch of segment 1–81 shown in Fig. 1 are the cause of the mismatch. Once the ultrametricity of a particular type of annealing problem is understood, Rammal et al. [7] speculate that hard landscape problems may be overcome by enlarging the configuration space during the annealing process in order to destabilize previous local minima without creating new ones.

ACKNOWLEDGEMENTS

We thank Dr. R.A. Lewis for providing the molecular graphics. P.M.D. wishes to thank the Wellcome Trust for personal financial support through the Principal Research Fellowship Scheme. Part of the work was supported by the Cambridge SERC Centre for Molecular Recognition.

REFERENCES

- 1 Barakat, M.T. and Dean, P.M., *J. Comput.-Aided Mol. Design*, 4 (1990) 295.
- 2 Aarts, E.H.L., Korst, J.H.M. and van Laarhoven, P.J.M., *J. Stat. Physics*, 50 (1988) 187.
- 3 van Laarhoven, P.J.M. and Aarts, E.H.L., *Simulated Annealing: Theory and Applications*, Reidel, Dordrecht, 1988.
- 4 Kirkpatrick, S. and Toulouse, G., *J. Physique*, 46 (1985) 1277.
- 5 McLachlan, A.D., *Acta Crystallogr. A*, 38 (1982) 871.
- 6 Bolin, J.T., Filman, D.J., Matthews, D.A., Hamlin, R.C. and Kraut, J., *J. Biol. Chem.*, 257 (1982) 13650.
- 7 Rammal, R., Toulouse, G. and Virasoro, M.A., *Rev. Mod. Physics*, 58 (1986) 765.