13 Jacob, F. and Monod, J., *J. Mol. Biol.*, 1961, **3**, 318.
14 Altman, S., in *Advances in Enzymology* (ed. Meister, A.), John Wiley and Sons Inc., 1989, p. 1.
15 Cech, T. R., *Sci. Am.*, 1986, November, 76.
16. Berget, S. M., Moore, C. and Sharp, P. A., *Proc. Natl. Acad. Sci. USA*, 1977, **74**, 3171.
17. Chow, L. T., Gelinas, R. E., Broker, T. R. and Roberts, R. J., *Cell*, 1977, **12**, 1.
18. Brody, E. and Abelson, J., *Science*, 1985, **228**, 963.
19. Maniatis, T. and Reed, R., *Nature*, 1987, **325**, 673.
20. Gall, J. G., *Science*, 1991, **252**, 1499.
21. Cech, T. R., *Cell*, 1986, **44**, 207.
22. Walter, P. and Blobel, G., *Nature*, 1982, **299**, 691.
23. Walter, P. and Blobel, G., *Proc. Natl. Acad. Sci. USA*, 1980, **77**, 7112.
23a. Davis, T. L., Firulli, A. B. and Kinniburgh, A. J., *Proc. Natl. Acad. Sci. USA*, 1989, **86**, 9682.
24. Blackburn, E. H., *Nature*, 1991, **350**, 569.
25. Blackburn, E. H., *Trends Biochem. Sci.*, 1991, **16**, 379.
26. Furuichi, T., Dhundale, A., Inouye, M. and Inouye, S., *Cell*, 1987, **48**, 47.
27. Inouye, M. and Inouye, S., *Trends Biochem. Sci.*, 1991, **16**, 18.
28. Banne, R., *Biochim. Biophys. Acta*, 1989, **1007**, 131.
29. Simpson, L., *Science*, 1990, **250**, 512.
30. Blum, B., Sturn, N. R., Simpson, A. M. and Simpson, L., *Cell*, 1991, **65**, 543.
31. Hansen, H. A. S., Volkman, N., Pielke, J., Glotz, C., Weinstein, S., Makowski, S., Meyer, S., Wittmann, H. G. and Yonath, A., *Biochim. Biophys. Acta*, 1990, **1050**, 1.
32. Wittmann, H. G., in *Structure, Function and Genetics of Ribosomes* (eds. Hardesty, B. and Kramer, G.), Springer Verlag, New York, 1985, p. 1.
33. Burma, D. P., *J. Sci. Ind. Res.*, 1984, **43**, 388.
34. Gale, E. F., Cundliffe, E., Reynolds, P. E., Richmand, M. H. and Waring, J. J., *The Molecular Basis of Antibiotic Action*, J. P. Wiley and Sons, New York, 1981.
35. Davis, B., *Microbiol. Rev.*, 1987, **51**, 431.
36. Girshovich, A. S., Bochkareva, E. S. and Gudkov, A. T., *FEBS Lett.*, 1982, **150**, 99.
37. Zimmermann, R. A., Thomas, C. L. and Wower, J., in *The Ribosomes, Structure, Function and Evolution* (eds. Hill, W. E., Dahlberg, A. E., Garrett, R. A., Moore, P. B., Schlessinger, D., Warner, J. R.), American Society for Microbiology, Washington DC, 1990, p. 331.
38. Montandon, P. E., Wagner, R. and Stutz, E., *EMBO J.*, 1986, **5**, 3705.
39. Ryan, P. P. and Draper, D. E., *Proc. Natl. Acad. Sci. USA*, 1991, **88**, 6308.
40. Allen, P. M. and Noller, H. F., *J. Mol. Biol.*, 1989, **208**, 457.
41. Burma, D. P., Nag, B. and Tewari, D. S., *Proc. Natl. Acad. Sci. USA*, 1983, **80**, 4875.
42. Burma, D. P., Tewari, D. S. and Srivastava, A. K., *Arch. Biochem. Biophys.*, 1985, **239**, 427.
43. Noller, H. F., Vernita, H. and Zimniak, L., *Science*, 1992, **256**, 1416.
44. Piccirilli, J. A., McConnell, T. S., Zaug, A. J., Noller, H. F. and Cech, T. R., *Science*, 1992, **256**, 1420.
45. Burma, D. P., *J. Biosci.*, 1984, **6**, 419.
45a. Burma, D. P., Srivastava, A. K., Srivastava, S., Dash, D., Tewari, D. S. and Nag, B., *J. Biosci.*, 1985, **8**, 757.
46. Burma, D. P., Srivastava, S., Srivastava, A. K., Mahanti, S. and Dash, D., in *Ribosomes, Structure, Function and Genetics* (ed. Hardesty, B. and Kramer, G.), Springer Verlag, New York, 1986, p. 438.
47. Dey, D. N. and Burma, D. P., *Indian J. Biochem. Biophys.*, 1991, **28**, 369.
48. Agrawal, R. K., De, A. and Burma, D. P., *Indian J. Biochem. Biophys.*, 1992, **29**, 148.
49. Noller, H. F., *Nature*, 1991, **353**, 302.
50. Von Ahsert, U., Davies, J. and Schroeder, R., *Nature*, 1991, **353**, 386.

RESEARCH COMMUNICATIONS

Global search for optimal biomolecular structures using mutually orthogonal Latin squares—A novel algorithm

N. Gautham and Z. A. Rafi
Department of Crystallography and Biophysics, University of Madras, Guindy Campus, Madras 600 025, India

We propose a modified grid search technique to find global starting points for the optimization of biomolecular conformations, in which the multidimensional search space is reduced to a two-dimensional one using mutually orthogonal Latin squares (MOLS). If there are m dimensions in the conformational space, each of size n, the method involves the computation of the energy values at m^2 of the m^n possible points, these m^2 points being chosen using MOLS. Subsequent analysis of the m^2 energy values using simple statistics allows identification of the optimal conformation. The computation time compared to conventional grid search methods could be reduced by several orders of magnitude. We have succesfully applied the method to arrive at the gross optimal conformation for a mononucleotide, a dinucleoside monophosphate, and for a tetrapeptide.

THEORETICAL studies of biomolecular conformation often use optimization techniques to minimize the energy and arrive at a stable geometry. Some of the most frequently used techniques[1-3] are essentially local minimization methods and require an appropriate initial conformation. In the absence of previous structural knowledge about the system, these initial points in the multidimensional configuration space may be arrived at by a sampling of the entire space. If the sampled points are closely spaced, such a search may become prohibitive in terms of computation time. We describe here a method to drastically decrease the computation time required to identify possible starting conformations. The method uses mutually orthogonal Latin squares (MOLS).

A Latin square of order m is defined as an arrangement of m symbols in an $m \times m$ square such that each symbol occurs exactly once in every row and once in every column. Two Latin squares are orthogonal, if, when they are superimposed, each symbol of the first square occurs once and only once with each symbol of the second square. A set of MOLS is a set of Latin squares any two of which are orthogonal[4-6] (Figure 1). It has been shown that if $m = q$, a prime number, then one can construct $q-1$ MOLS of order q[7-10].

MOLS have been used in the design of agricultural or medical experiments as well as in cryptography[4,9,11,12]. We have now adapted the method for optimization.

In order to test the use of MOLS in optimization, we tried the method on eighteen different arbitrary functions of up to 11 variables designed specifically for this purpose, with known minima. The method successfully picked up the minima each time. Encouraged by this success, we then applied the method to optimize biomolecular structure.

The algorithm we used is as follows.

Step 1. Choose the function to the optimized, i.e. $F(x_i)$, $i = 1, n$, where x_i are the variables. In the applications described below, the function is the energy and the x_i are the torsion angles.

Step 2. Choose the range for each x_i. In the present case the range is 0° to 360°.

Step 3. Choose the order m of the MOLS grid such that (i) m is a prime number, (ii) $m \geq n+1$, and (iii) m = the number of points at which each variable is to be sampled.

	I	II	III	IV	V	VI	VII
A	x_1 y_1 z_1	x_7 y_6 z_5	x_6 y_4 z_2	x_5 y_2 z_6	x_4 y_7 z_3	x_3 y_5 z_7	x_2 y_3 z_4
B	x_2 y_2 z_2	x_1 y_7 z_6	x_7 y_5 z_3	x_6 y_3 z_7	x_5 y_1 z_4	x_4 y_6 z_1	x_3 y_4 z_5
C	x_3 y_3 z_3	x_2 y_1 z_7	x_1 y_6 z_4	x_7 y_4 z_1	x_6 y_2 z_5	x_5 y_7 z_2	x_4 y_5 z_6
D	x_4 y_4 z_4	x_3 y_2 z_1	x_2 y_7 z_5	x_1 y_5 z_2	x_7 y_3 z_6	x_6 y_1 z_3	x_5 y_6 z_7
E	x_5 y_5 z_5	x_4 y_3 z_2	x_3 y_1 z_6	x_2 y_6 z_3	x_1 y_4 z_7	x_7 y_2 z_4	x_6 y_7 z_1
F	x_6 y_6 z_6	x_5 y_4 z_3	x_4 y_2 z_7	x_3 y_7 z_4	x_2 y_5 z_1	x_1 y_3 z_5	x_7 y_1 z_2
G	x_7 y_7 z_7	x_6 y_5 z_4	x_5 y_3 z_1	x_4 y_1 z_5	x_3 y_6 z_2	x_2 y_4 z_6	x_1 y_2 z_3

Figure 1. Three mutually orthogonal Latin squares of order 7.

Step 4. Divide the range for each x_i into m points to be sampled. The search volume is thus m^n.

Step 5. Choose m^2 points in this m^n volume using MOLS. Figure 1 illustrates, as an example, the choice of 7^2 points from a total of 7^3.

Step 6. Calculate F at each of these m^2 points.

Step 7. Calculate the averages for each value of each variable ($n \times m$ averages) e.g. referring to Figure 1, for the value x_1 of the variable x, this would involve taking the average of F at the points represented by the subsquares A I, B II, C III, D IV, E V, F VI and G VII.

Step 8. For each variable, separately examine the averages for the optimum value.

Step 9. The set of optimum values of the variables gives the optimum configuration with respect to F.

This method was applied to the optimization of the three torsion angles of the mononucleotide 5′-TMP (Figure 2, the sugar pucker was held constant at C2′endo) by minimizing the energy, which was calculated using only the electrostatic and nonbonded terms of the semi-empirical expression, i.e.

$$E = \sum_{i>j} \frac{332\, q_i q_j}{\epsilon r_{ij}} - \sum_{i>j} \frac{A_{ij}}{r_{ij}^6} + \sum_{i>j} \frac{B_{ij}}{r_{ij}^{12}}.$$

The dielectric constant ϵ was taken to be 4, and A_{ij} and B_{ij} are the Lennard–Jones constants[13]. The partial charges q_i and q_j between the two interacting atoms separated by a distance r_{ij} were taken from Weiner *et al.*[14]. The three torsion angles were sampled at intervals of 10°, i.e. a total of 36 values. Since the nearest higher prime number is 37 the order of the Latin squares for each angle was taken to be 37 and each angle was

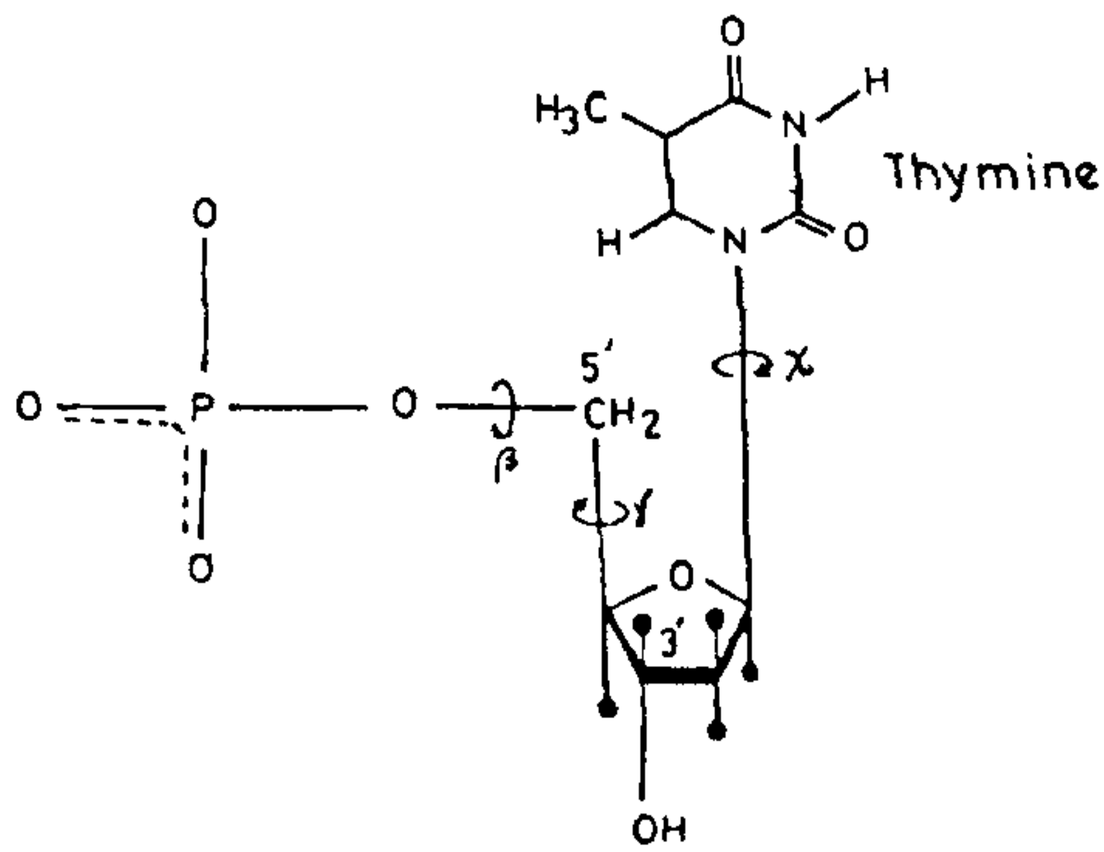

Figure 2. 5′-TMP. Only the three torsion angles indicated were varied (symbols follow IUPAC-IUB conventions[23]).

therefore sampled both at 0° and at 360°. Out of a possible 37^3 points in the configurational space, 37^2 were chosen by constructing a set of three MOLS of order 37. The energy was calculated at each of these points. The average energy at each value of the torsion angles was calculated by considering all the 37 different points where this value occurred, in the manner specified in the algorithm. The averages were then examined to arrive at the optimum value for each angle. This resulted in the set of angles ($\beta = 173.8°$, $\gamma = 32.9°$ and $\chi = 276.8°$) as the optimum one for the mononucleotide (Figure 3). This conformation compares very well with those observed in crystals of mononucleotides where β shows a preference towards *trans*, γ towards *gauche*$^+$ and χ for pyrimidine nucleotides tend to be *anti* when sugar is in the C2'endo conformation[15,16].

The second application was to the dinucleoside monophosphate ApA. Here only the nonbonded interactions were taken into consideration. This was because a) we wished to see whether the MOLS procedure as applied to larger dimensional searches is successful at a gross level and b) these interactions would have an easily recognizable optimum where the molecule was compactly folded. Seven torsion angles were varied (Figure 4). The sugar pucker was again kept constant at C2'endo. A set of seven MOLS of order 37 was used and the procedure described above was carried out. Since, however, an examination of the averages gave several possible optima for each of the 7 torsion angles, the 11 best values were picked up for each angle (with the provision that no two values were less than 20° from each other) and a set of 7 MOLS of order 11 was constructed and the procedure repeated. Once again there were several possibilities and a further set of 7 MOLS of order 7 was constructed. From the results of this cycle 3 possibilities were picked up for each angle and all the 3^7 combinations were checked. The angles and non-bonded energies corresponding to five best conformations are shown in Table 1. Figure 5 shows the conformation no. I in the table.

It may be seen that, for the dinucleotide, the MOLS procedure leads to many physically meaningful combinations of torsion angles. These combinations correspond to very different conformations, all of them however with van der Waals' energy values within 5 kcal mol^{-1} of the minimum. Among these are a few which lead to regular helical structures when repeated along a polymeric backbone[15-20]. Thus conformation no. IV of Table 1, with an energy value -28.17 kcal mol^{-1}, shows an *anti*, *t*, g^-, *t*, *t*, *t* and *anti* combination for the angles χ_1, ε, ζ, α, β, γ and χ_2 respectively which will lead to a helical structure with Watson–Crick pairing[15]. Conformation no. V is very close to the B-DNA structure[17]. The angles do not have exactly the preferred optimal values. Nevertheless, these conformations can be used as the starting points for a more conventional gradient minimization. Conformations nos. II and III are *cis* about the ζ bond. These normally unfavourable values[15] of the torsion angles are probably due to the non-inclusion of the terms for the torsion potential in the expression for the energy.

The third application was to the tetrapeptide *d*-val-*l*-ala-*l*-leu-*l*-ala. A recent solution of the crystal structure of this compound in this department[21] motivated the choice of this molecule. Nine torsion angles ($\chi_1, \psi_1, \phi_2, \psi_2, \phi_3, \chi_3^1, \chi_3^2, \psi_3, \phi_4$, in the standard IUPAC-IUB nomenclature[22]) were varied, with the rest of the procedure as for ApA except that both the electrostatic and non-bonded energy terms were used. Table 2 gives the five best conformations, while Figure 6 corresponds to conformation no. I in the table.

Again the MOLS procedure leads to many different physically meaningful combinations of the torsion angles, which are of comparable low energy values. The five best combinations differ only in the values of ψ_1, ϕ_2, ϕ_3 and χ_3^2. Both the crystal and the MOLS structures prefer the β strand region of the Ramachandran

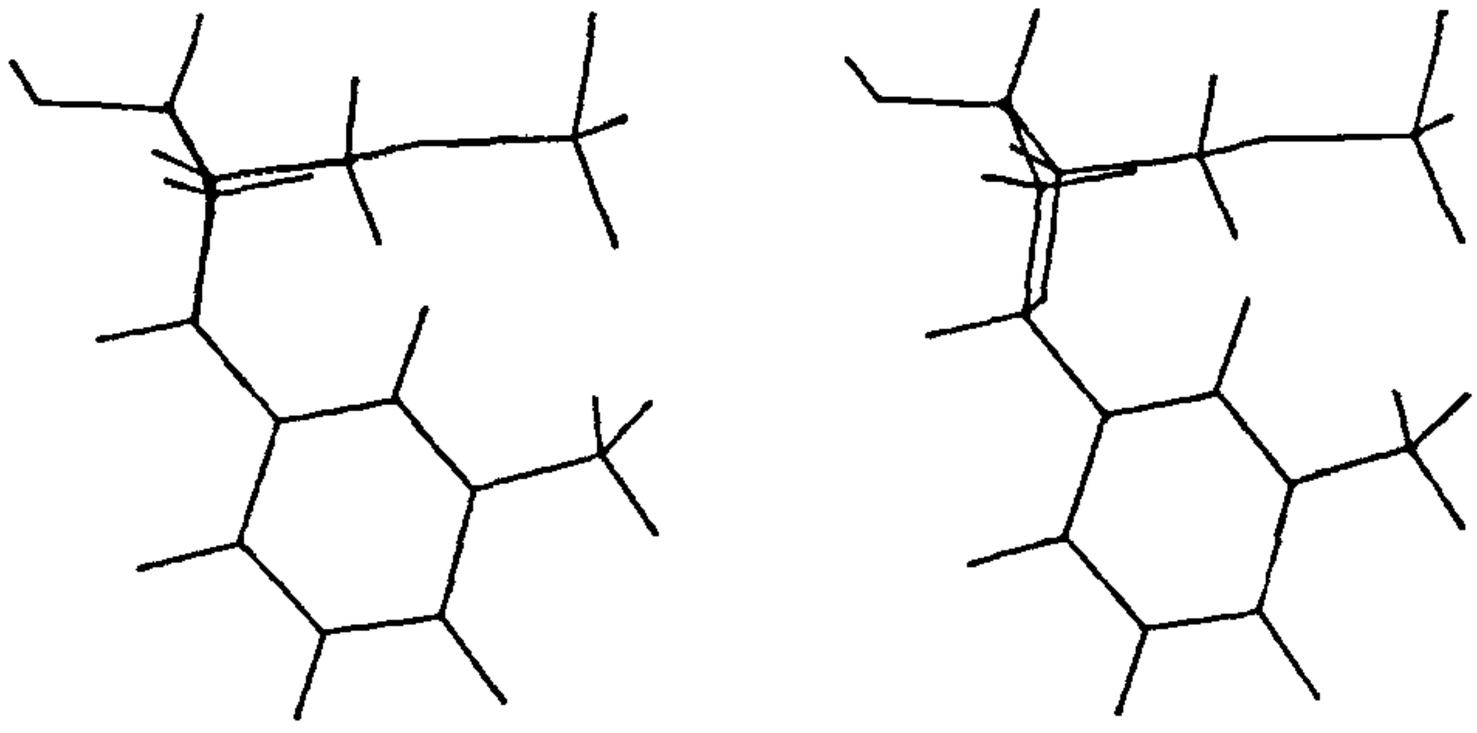

Figure 3. Stereo view of the optimum conformation of 5'-TMP as obtained by the MOLS procedure.

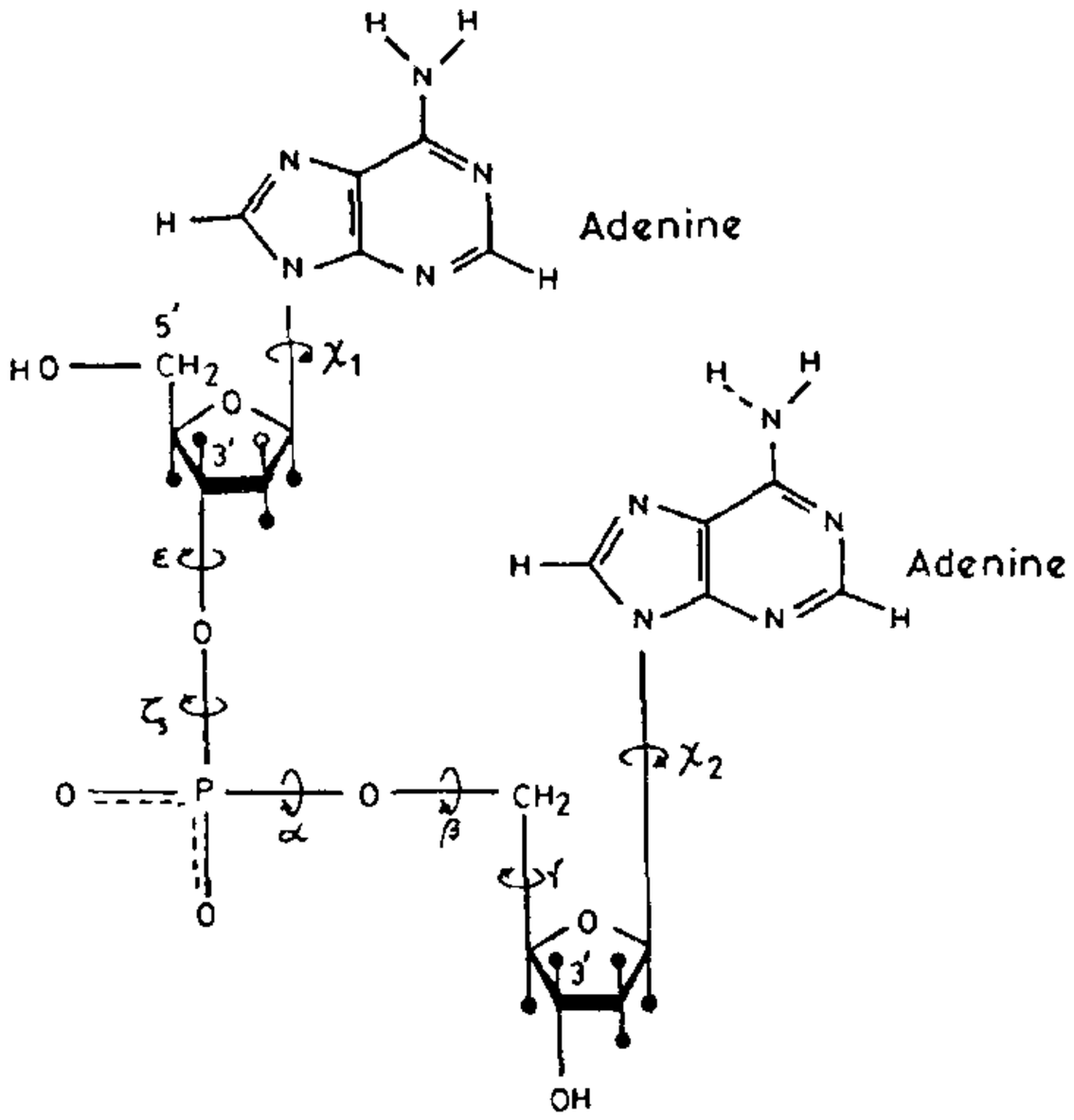

Figure 4. d(ApA). Only the seven torsion angles indicated were varied.

plot. All five MOLS combinations however lead to structures which are slightly more folded than the crystal structure. The value of the energy calculated by the function mentioned above is marginally greater for the crystal structure than for the structures indicated by MOLS showing that the present method indeed leads to the minimum of the given function.

The choice of the MOLS is not uniquely determined. Given a set of 5 symbols, for example, it is possible to construct about 2×10^{11} different sets of MOLS. We explored the effect of different choices on the final result in the case of ApA. Ten different trials were made. Each time a set of 7 MOLS of order 37 was generated using a random number generator. Subsequent analysis of the results showed no significant deviation in the optima picked up.

The MOLS procedure thus appears to be an ideal first step in exploring all possible biomolecular conformations, including those far from known minima. Conventional techniques do not allow such global searches at an economic computational cost. MOLS, in contrast, can drastically reduce the size of the computation problem to a manageable level when searching for fully globally optimal biomolecular conformations.

Table 1. The five 'best' conformations for the dinucleoside monophosphate ApA as obtained through the MOLS procedure.

Conformation no.	χ_1	ε	ζ	α	β	γ	χ_2	Energy (kcal mol^{-1})
I	62.9	188.6	37.0	252.0	167.6	328.9	42.2	−31.92
II	62.9	188.6	7.0	162.0	167.6	48.9	152.2	−29.86
III	212.9	188.6	7.0	162.0	167.6	48.9	112.2	−29.62
IV	132.9	188.6	257.0	162.0	167.6	118.9	2.2	−28.17
V	212.9	288.6	257.0	252.0	127.6	48.9	152.2	−27.56

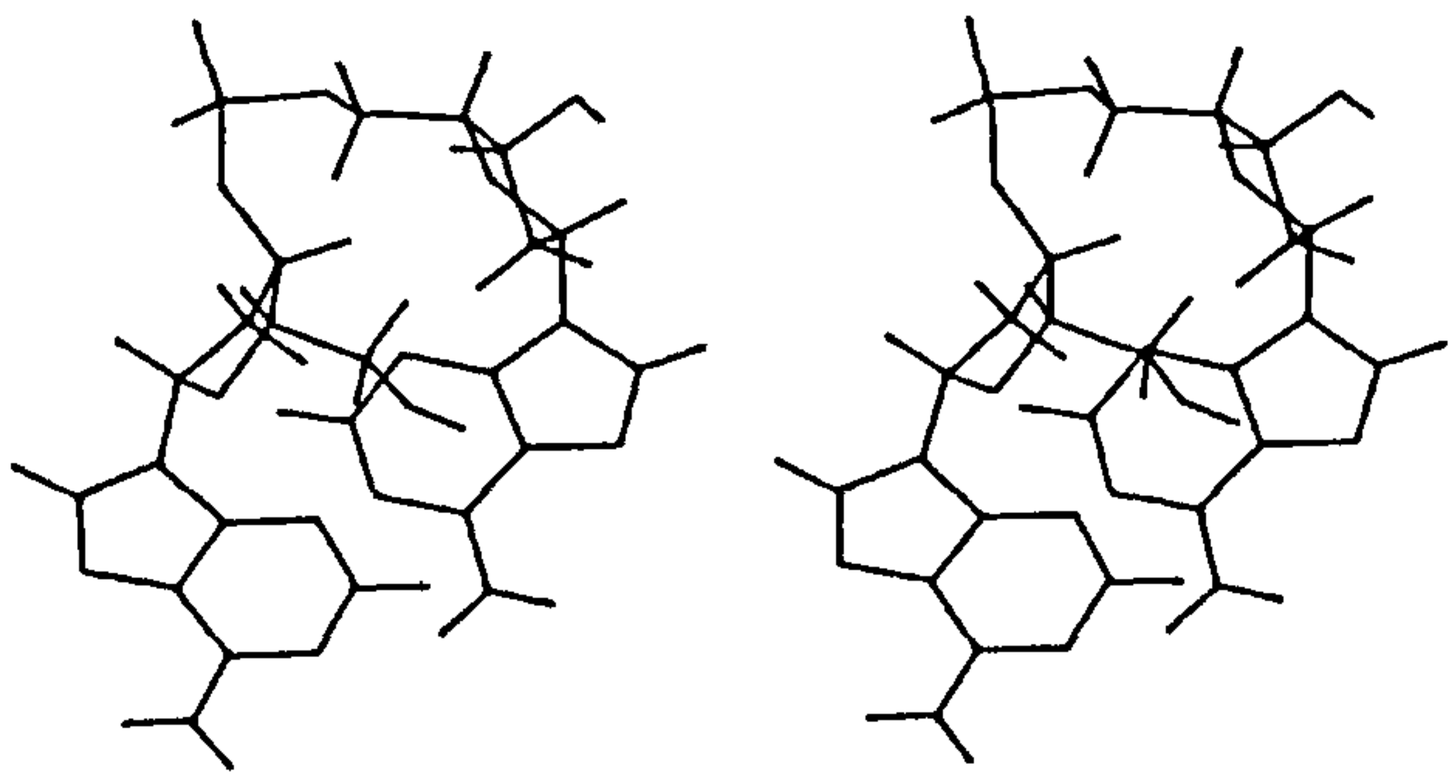

Figure 5. Stereo view of the conformation of d(ApA) with the minimum Ivan der Waals' energy, as obtained by the MOLS procedure (i.e. conformation no. I of Table 1).

Table 2. The five 'best' conformations for the tetrapeptide as obtained through the MOLS procedure.

Conformation no.	χ_1	ψ_1	ϕ_2	ψ_2	ϕ_3	χ_3^1	χ_3^2	ψ_3	ϕ_4	Energy (kcal mol^{-1})
I	55.5	−99.4	−108.4	97.7	−61.1	−155.3	52.1	100.0	−148.4	−0.92
II	55.5	−99.4	−108.4	97.7	−81.1	−155.3	52.1	100.0	−148.4	−0.48
III	55.5	−99.4	−148.4	97.7	−61.1	−155.3	52.1	100.0	−148.4	−0.39
IV	55.5	−49.4	−108.4	97.7	−61.1	−155.3	52.1	100.0	−148.4	−0.34
V	55.5	−99.4	−108.4	97.7	−61.1	−155.3	102.1	100.0	−148.4	−0.22
Crystal structure	−70.3	−148.9	−145.8	148.7	−116.2	−172.3	59.5	125.3	−111.4	−0.06

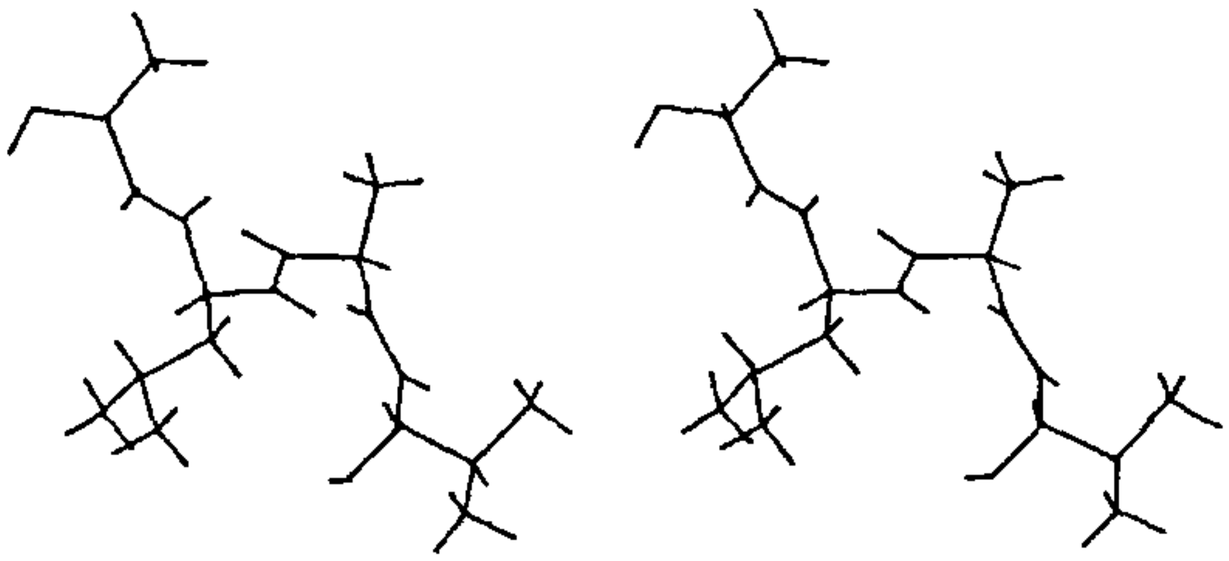

Figure 6. Stereo view of the conformation of the tetrapeptide with the minimum energy, as obtained by the MOLS procedure (i.e. conformation no. I of Table 2).

The present paper, we believe, represents the first report of the application of MOLS as an optimization technique. We have successfully demonstrated its utility in relation to biomolecular conformational searches. However it has not escaped our notice that the MOLS procedure as outlined here could find application in other areas of science and technology.

1. Weiner, S. J. and Kollman, P., *J. Comput. Chem.*, 1981, **2**, 287–303.
2. Schlick, T. and Overton, M., *J. Comput. Chem.*, 1987, **8**, 1025–1039.
3. Fletcher, R. and Powell, M. R. J., *Computer J.*, 1963, **6**, 163–168.
4. Fisher, R. A., *The Design of Experiments*, Oliver and Boyd, London, 1960, pp. 70–85.
5. Ito, K., in *Encyclopaedia Dictionary of Mathematics*, MIT press, Cambridge, Massachusetts, 1987, vol. II, pp. 891–892.
6. Mikheev, V. M., in *Encyclopaedia of Mathematics* (eds. Vinogradov, I. M., *et al*), Kluwer Academic Publishers, Dordrecht, 1990, vol. 5, pp. 362–364.
7. Ryser, H. J., *Combinatorial Mathematics*, Mathematical Association of America, 1963, pp. 79–84.
8. Hall, Jr. M., *Combinatorial Theory*, Blaisdel Publishing Co., Waltham, Massachusetts, 1967, pp. 192–194.
9. Liu, C. L., in *Introduction to Combinatorial Mathematics*, McGraw-Hill Book Co, New York, 1968, pp 359–370.
10. Lidl, R. and Niederreiter, H., *Finite Fields*, Addison-Wesley Publishing Co, Massachusetts, 1983, pp 513–515.
11. Finney, D. J., *Experimental Design and its Statistical Basis*, Cambridge University Press, London, 1955, pp. 56–67.
12. Denes, J. and Keedwell, A. D., *Latin Squares, New Developments in the Theory and Applications*, North Holland–Amsterdam, 1991.
13. Balasubramanian, R. and Seetharamulu, P., *Int. J. Biol. Macromol.*, 1983, **5**, 167–170.
14. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P., *J. Am. Chem. Soc.*, 1984, **106**, 765–784.
15. Sundaralingam, M., *Biopolymers*, 1969, **7**, 821–860.
16. Bansal, M. and Sasisekharan, V., in *Theoretical Chemistry of Biological Systems* (ed. Naray-Szabo, G.), Elsevier Science Publications, The Netherlands, 1986, pp. 127–218.
17. Arnott, S. and Hukins, W. L., *Biochem. Biophys. Res. Commun.*, 1972, **47**, 1504–1509.
18. Olson, W. K. and Flory, P. J., *Biopolymers*, 1972, **11**, 25–56.
19. Yathindra, N. and Sundaralingam, M., *Biopolymers*, 1973, **12**, 297–314.
20. Thiagarajan, P. and Ponnuswamy, P. K., *Biopolymers*, 1978, **17**, 533–553.
21. Pattabhi, V., (personal communication).
22. IUPAC-IUB Commission on Biochemical Nomenclature, *Biochem. Biophys. Acta*, 1971, **229**, 1–17.
23. IUPAC-IUB Joint Commission on Biochemical Nomenclature, *Eur J. Biochem.*, 1983, **131**, 9–15.

ACKNOWLEDGEMENTS. We thank Prof. N. Yathindra for discussions, suggestions and encouragement. We are grateful to Prof Vasantha Pattabhi for suggestions and for access to unpublished data.

28 February 1992; revised accepted 2 September 1992