

ON THE INVESTIGATION OF THE LOW-ENERGY CONFORMATIONAL SPACE OF MOLECULES

C. TOSI, R. FUSCO, L. CACCIANOTTI

Istituto Guido Donegani, Via G. Fauser 4, I-28100 Novara, Italy

and

V. PARISI

Dipartimento di Fisica, II Università di Roma, Via O. Raimondo, I-00173 Rome, Italy

Received 12 September 1988

(Received by the publisher 5 July 1989)

Abstract

A new perspective on traditional energy minimization problems is provided by a connection between statistical thermodynamics and combinatorial optimization (finding the minimum of a function depending on many variables). The joint use of a new method for uncovering the global minimum of intramolecular potential energy functions, based on following the asymptotic behavior of a system of stochastic differential equations, and an iterative-improvement technique, whereby a search for relative minima is made by carrying out local quasi-Newton minimizations starting from many distinct points of the energy hypersurface, proved most effective for investigating the low-energy conformational space of molecules.

1. Introduction

The investigation of the “low-energy conformational space” of molecules, and in particular the search for the global minimum of intramolecular potential energy functions, represents one of the most challenging problems of molecular mechanics. No methods capable of determining unambiguously the point of lowest energy of a conformational energy surface in a deterministic way are known. The usual minimization methods, regardless of their inherent peculiarities, suffer from the limitation of being local methods. In other words, they end up in the minimum “nearest” to the starting conformation so that, unless a previous detailed knowledge of the energy surface is available, it is almost an exceptional accident that they determine the conformation with the lowest possible energy in a given force field.

There are several procedures for solving the global energy minimum conformation problem in some particular instances. With the so-called "grid search" technique [1], the energy is evaluated in correspondence with the vertices of hypercubes generated by dividing up the full angular range of each torsion angle into narrow intervals (generally 10° to 30°). Once an array of energy values has been produced, it is possible to assign them to a number of "clusters" of low-energy conformations, and to carry out local minimizations within each cluster; that of the final conformations which has the lowest energy is taken as the global minimum. A more effective version of this method, generally referred to as "augmented grid search", is based on the division of the main chain non-hydrogen atoms of any acyclic structure into overlapping groups, the location of energy minima of all sub-groups on rotating all torsion angles through 360° , and the evaluation of the lowest energy for all possible combinations of the local torsional energy minima [1].

The above methods are time-consuming and not very efficient and, in addition, are restricted to fairly small molecules. For larger molecules, the problem can be made tractable by turning to an "iterative improvement" technique: one starts with the molecule in a known conformation and applies a standard rearrangement operation to all parts of the system in turn until a rearranged conformation that improves the objective function (i.e. that has a lower energy) is discovered. The rearranged conformation then becomes the new conformation of the molecule, and the process is continued until no further improvements (gains in energy) can be found. This search frequently gets stuck in a local minimum, so the process should be carried out many times, starting from different, randomly generated conformations. The best known example of the iterative improvement technique to the macroscopic rearrangement processes modeled by statistical mechanics is the Metropolis Monte Carlo procedure [2] which is, however, normally used to determine the optimal configuration of a system where intermolecular forces are at work (e.g. a rigid solute molecule surrounded by many rigid solvent molecules) rather than the best conformation of a single molecule where interactions of atoms with each other are described by certain intramolecular potential energy functions.

Based on the iterative improvement approach is a method proposed by Rao et al. [3] for finding the lowest-energy conformation of molecules, which does not require the starting point to be close to the actual solution, does not calculate gradients, and has been claimed to be successful even in cases involving ill-conditioned equations [4]. This method uses an algorithm of unconstrained global optimization developed by Bremermann [5] for arriving at the minimum value of a fourth-order function. Despite a number of successful applications, the latter method has some severe limitations, in particular the fact that, because gradients are not calculated, points of the energy surface crossed in the descent to the lowest-energy point do not correspond to local minima, and the amount of information on the shape of the surface gained at the end of a computation is very small.

It is worth mentioning two other algorithms for the global minimization of functions that, as far as we are cognizant, have not been applied until now to the conformational case. The “tunneling” algorithm by Levy and Montalvo [6] consists of two phases: a minimization phase whereby the current function value is lowered until a local minimum is found, and a tunneling phase which has the purpose of finding a point, other than the last minimizer found, such that when employed as the starting point for the next minimization phase, the new stationary point will have a function value no greater than the previous minimum found. The algorithm developed by Donnelly and Rogers [7] uses the trajectories of a discrete dynamical system to sample the domain of an objective function in the search for global minima. The effectiveness of this algorithm was demonstrated through its application to the optimization of the relative geometry of two rigid propane molecules interacting via van der Waals forces.

On account of this situation, it is then tempting to revisit the problem of the investigation of the most stable conformations of molecules, with the goal of developing techniques capable of maximizing the ratio between the information obtained on the low-energy conformational space and the time spent to obtain it.

2. New algorithms of search for energy minima

A fundamental question in seeking the global minimum is whether the fact that it has the lowest possible energy for that particular molecule in that particular force field has implications that can be exploited to find a solution to the problem of its discovery.

Since the conformational states of a molecule are weighted by their Boltzmann probability factors, it would seem that cooling a molecule to the absolute zero of temperature would suffice to uncover its minimum-energy state. Actually, however, low temperature is a necessary, but not sufficient condition for finding the ground state. As keenly remarked by Kirkpatrick et al. [8], experiments that determine the low-temperature state of a material – e.g. the growth of a single crystal from a melt – are done by first melting the substance and then lowering the temperature slowly after spending a long time at temperatures near the freezing point. Should this not be done, the substance would get out of equilibrium and form a glass with only metastable, locally optimal structures and no crystalline order. In analogy with annealing in solids, Kirkpatrick et al. [8] developed a method of optimization of the properties of very large and complex systems.

A good description of how a simulated-annealing method can be used to guide a search towards the absolute minimum was made by Wille [9], with illustration on the problem of the minimum-energy configuration of equal charges confined to a sphere.

The use of similar ideas from the adiabatic perturbation theory led Aluffi-Pentini, Parisi and Zirilli [10,11] to propose a new algorithm, called SIGMA [12,13],

for global optimization of real-valued functions defined in the N -dimensional real Euclidean space. This algorithm was subsequently extended to the case where the function whose global minimum we are seeking is the conformational energy of a molecule [14]. In the current version of the program, only the torsional geometry is allowed to change, i.e. both valence geometry (bond lengths and angles) and possible cyclic sub-structures are kept rigid.

The method looks for a point of absolute minimum by following the solution trajectories of the stochastic differential equation

$$dx = -\nabla f(x)dt + \epsilon dw(t) \quad (1)$$

obtained by adding the random perturbing force $\epsilon dw(t)$, where $w(t)$ is a standard N -dimensional Wiener process, to the “steepest descent” ordinary differential equation $dx = -\nabla f(x)dt$, which represents the motion of the atoms comprising the molecule under the action of the potential field. Equation (1), known as the Smoluchowski–Kramers equation, is a singular limit, valid when the inertial terms are negligible (i.e. when many collisions occur in each time unit), of the second-order equation of Langevin, which describes the motion of a molecule in a medium in thermal equilibrium with it. Equation (1) has been widely used to study physical phenomena, such as chemical reactions and diffusion of atoms in crystals [15].

The original version of SIGMA was used to test thirty-seven problems [12], and always gave good results, both on ill-conditioned problems and on problems with many minima (up to ten billions). However, for complex conformational problems where the potential is theoretically of infinite, but practically of limited range (e.g. the Lennard–Jones potential), it happened that the minimum found by SIGMA, although very low, was not the absolute one. A more recent version of SIGMA, called SIGRAC and based on the method of conjugate gradients [16], has shown a comparable performance on the thirty-seven problems mentioned above, while it works much better than SIGMA when applied to conformational cases.

In all cases we have dealt with so far, the global optimization algorithms described above have proved able to find a minimum which, within the intrinsic limitations of any probabilistic method, has been assumed to be the global one. The likelihood of this assumption would certainly be higher if the result could be borne out in some way by an independent technique.

For this purpose, we have developed an algorithm which, although not directly aimed at the detection of the global minimum, has proved to great use in investigating the low-energy regions of conformational surfaces. The algorithm, called LECSA (Low-Energy Conformational Space Analysis) [17,18], may be indicated as a “random-search-plus-local-optimization” technique. It is similar to the iterative-improvement techniques described in the first section, but with the substantial difference that each low-energy point is produced independently of those previously

generated. Starting from a generic point of the energy surface, obtained with a random-number generation technique, a local minimization is carried out with a quasi-Newton method. Several criteria are used to speedup the calculation such that, in its current version [18], LECSA detects a number of minima (always, so far, including the global one) approximately an order of magnitude larger than that detected by SIGMA-SIGRAC, with computer times an order of magnitude smaller. These criteria are based on the “distance” of starting conformations from the minima already found (evaluated along both the energy axis and the torsion angle axes), as well as the information collected about the preferred angular ranges of the internal rotations.

3. A test case: the sugar-phosphate-sugar fragment

As an example of the performance of the joined application of the SIGMA-SIGRAC and LECSA methods, we have chosen the dideoxyribose-phosphate (“sugar-phosphate-sugar”, or SPS) fragment shown in fig. 1. This molecule was studied in

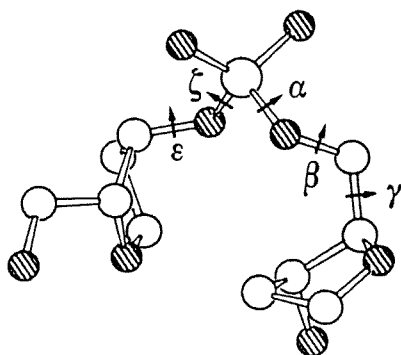


Fig. 1. Computer-drawn representation of SPS in a low-energy conformation (graphical program SCHAKAL by E. Keller, Institute of Crystallography, University of Freiburg i Br., FRG).

detail over the past few years with the goal of developing a potential energy function for polynucleotides with parameters best fitted to *ab initio* energies, computed for various sets of the five internal rotation angles ϵ , ζ , α , β and γ [19]. Note that SPS was taken as a test case for the application of the original version of the global optimization algorithm [14,20]: the conformation it took as the global minimum was subsequently shown to be the second lowest minimum, both by applying LECSA and SIGRAC and by running SIGMA in different conditions [21].

The fifty deepest minima of SPS detected by LECSA are listed in table 1 (cf. also table I of [21]); tables 2 and 3 describe a global minimization “history” for SIGMA and SIGRAC, respectively. A few comments about the contents of tables 1–3 appear in order at this point.

Table 1
Deepest minima of SPS as found by LECSA

No.	E	ϵ	ζ	α	β	γ
1	-97.81	-80.1	72.7	127.6	-97.7	61.5
2	-94.63	176.5	176.4	121.2	-95.6	55.5
3	-94.11	178.4	-178.9	72.6	-107.4	58.5
4	-92.79	-75.6	175.8	-115.6	91.3	-31.8
5	-92.49	-78.6	-51.3	-169.0	-103.2	57.1
6	-92.27	-177.8	60.0	125.9	-97.3	58.9
7	-92.26	-77.2	179.3	63.9	-110.8	56.8
8	-92.05	-77.0	-47.1	-94.7	89.9	-71.0
9	-91.31	-75.8	-173.8	120.0	-95.0	54.6
10	-91.30	-84.0	64.6	-116.8	-88.7	55.0
11	-91.26	-171.5	-64.9	-177.1	-105.0	59.2
12	-91.04	-79.2	-54.8	-53.6	-100.0	55.8
13	-91.00	-76.6	-179.5	-5.0	-90.2	55.2
14	-90.87	-77.6	74.6	166.2	93.7	41.9
15	-90.46	-77.1	75.8	-163.3	101.9	-56.7
16	-90.33	178.0	174.8	-77.4	57.7	-89.5
17	-90.26	178.1	-76.8	-109.6	93.8	-55.2
18	-90.10	-174.8	-72.7	-67.6	-101.4	56.6
19	-90.06	-176.5	-171.5	-52.3	111.6	-59.2
20	-89.88	-78.1	-50.4	-118.3	-88.4	54.9
21	-89.85	-77.9	-49.8	145.6	88.5	50.0
22	-89.81	177.7	54.8	-116.3	-88.7	55.4
23	-89.80	-78.2	-170.2	-83.9	57.8	-94.7
24	-89.59	177.6	179.6	68.0	-68.7	-55.7
25	-89.54	173.9	177.2	-109.2	91.2	51.7
26	-89.51	-78.0	-175.2	-119.5	93.1	47.7
27	-89.31	-166.9	-69.3	178.7	-105.1	65.8
28	-89.25	-175.5	-69.5	-117.7	-88.5	55.1
29	-89.17	-79.1	-52.0	-117.4	91.8	-32.7
30	-89.12	-76.5	176.2	-44.6	-97.8	54.9
31	-89.03	177.8	55.2	151.5	88.6	48.0
32	-88.95	-175.0	-69.9	146.9	87.1	49.5
33	-88.91	-75.8	-175.5	69.0	-67.7	-55.3
34	-88.70	-76.6	-175.6	28.9	74.6	-80.4
35	-88.39	-78.0	-51.6	167.5	-74.8	-55.3
36	-88.37	-177.8	69.6	78.9	-71.2	-68.5
37	-88.17	-83.0	65.0	179.3	-106.6	56.6
38	-88.05	171.9	170.4	6.7	92.9	-58.0
39	-88.01	-71.8	87.4	41.6	58.8	-92.0
40	-87.63	178.8	179.8	-49.0	-99.1	55.2
41	-87.44	178.8	176.3	28.8	93.9	-75.7
42	-87.42	-178.6	-174.0	46.2	108.9	-53.8
43	-87.29	-74.4	80.1	56.2	108.7	-56.2
44	-87.14	-178.1	177.3	32.6	91.4	-80.4
45	-87.12	179.6	179.9	-2.5	-91.0	56.8
46	-87.08	-179.1	-174.8	27.0	96.4	44.4
47	-87.01	-75.6	-177.0	-117.7	-88.4	55.0
48	-86.85	178.5	-179.8	-118.0	-88.4	54.9
49	-86.80	-179.5	58.2	51.9	109.2	-55.8
50	-86.70	-75.6	-177.1	147.9	87.3	49.3

Table 2
Minimization history of SPS with SIGMA

NFEV	E	Type	No.	ϵ	ζ	α	β	γ
1	-68.29			180.0	180.0	180.0	180.0	180.0
12182	-83.04	1		178.8	179.9	-60.0	-64.9	-57.3
15528	-82.11	-1		178.7	-179.1	178.3	-68.8	-57.6
23818	-88.91	1	33	-75.8	-175.5	69.0	-67.7	-55.2
41652	-79.86	0		-177.0	-72.7	169.2	175.7	51.9
61500	-92.26	1	7	-77.2	179.4	64.0	-110.9	56.8
103122	-88.95	-1	32	-174.9	-69.8	146.9	87.1	49.6
118291	-90.06	-1	19	-176.4	-171.4	-52.3	111.6	-59.2
131133	-83.46	-1		-79.6	-55.7	-54.5	162.0	-51.6
172682	-92.01	0	→8	-77.5	-47.2	-94.1	87.5	-72.2
184723	-85.62	-1		-174.3	-68.9	176.6	-69.6	-56.6
197800	-85.11	-1		-80.5	-58.0	-71.5	70.8	-168.7
250658	-88.77	0	→22	-175.7	60.4	-114.2	-89.2	57.7
254965	-68.78	-1		178.7	-179.6	177.1	179.8	-175.5
276285	-89.54	-1	25	173.9	177.2	-109.2	91.3	51.7
282849	-83.04	-1		178.8	179.9	-60.0	-64.9	-57.3
295585	-84.80	-1		-75.5	-175.2	49.0	109.3	-55.5
303808	-87.29	-1	43	-74.4	80.2	56.2	108.8	-56.2
314093	-89.59	-1	24	177.6	179.7	68.1	-68.7	-55.7
356348	-91.55	-1	8	-78.6	-52.5	-88.8	75.2	-79.1
410041	-87.28	-1	43	-74.9	80.1	56.9	108.7	-56.1
414213	-94.11	1	3	178.5	-178.9	72.6	-107.4	58.5
438691	-97.81	1	1	-80.1	72.7	127.6	-97.7	61.6
537012	-85.92	-1		179.8	57.0	-173.2	103.2	-55.9
540620	-97.81	1	1	-80.1	72.6	127.6	-97.7	61.5
552885	-87.07	-1	46	-179.1	-175.0	27.0	96.5	44.3
555507	-97.81	1	1	-80.1	72.6	127.6	97.7	61.5
576853	-88.39	-1	35	-78.0	-51.6	167.5	-74.8	-55.3
577986	-97.81	1	1	-80.1	72.7	127.6	-97.7	61.5
581472	-97.81	1	1	-80.1	72.7	127.6	-97.7	61.5

(i) The minima listed in table 1 are arranged in order of increasing energy, not in the chronological succession they had been detected by LECSA.

(ii) In tables 2 and 3, NFEV denotes the number of function evaluations. The entry "Type" indicates the fate of a single trial (cf. ref. [13]): 0 means that the trial did not converge within the maximum allowed number of observation periods; -1 means that the trial converged to a minimum whose energy is higher than at least one of the energies found so far by the program (perhaps in the course, not at the end, of a former trial); 1 means that the trial converged to a minimum with energy lower than all of the energies encountered so far by the program. "No." is the running

Table 3
Minimization history of SPS with SIGRAC

NFEV	E	Type	No.	ϵ	ζ	α	β	γ
1	-68.29			180.0	180.0	180.0	180.0	180.0
148691	-92.79	0	→ 4	-75.7	175.8	-115.7	91.4	-31.8
151562	-68.78	-1		178.7	-179.6	177.1	179.8	-173.5
154767	-68.78	-3		178.7	-179.6	177.1	179.8	-173.5
157872	-68.78	-1		178.7	-179.6	177.1	179.8	-173.5
161296	-68.78	-1		178.7	-179.6	177.1	179.8	-173.5
166723	-68.78	-1		178.7	-179.6	177.1	179.8	-173.5
205440	-90.46	-1	15	-77.0	75.8	-163.3	101.9	-56.7
266093	-88.70	-1	34	-76.8	-175.5	28.9	74.5	-80.4
291068	-94.11	1	3	178.5	-178.9	72.6	-107.4	58.4
314533	-86.07	-3		-78.0	-50.4	152.0	97.3	-61.3
399238	-85.90	0		179.6	57.3	-173.3	102.8	-55.5
457406	-71.21	0	→ 4	-121.1	166.6	-64.7	109.9	-89.3
496668	-84.14	-1		-170.5	-71.1	-74.2	-69.8	-56.4
532410	-97.81	1	1	-80.1	72.7	127.6	-97.7	61.5
565747	-84.14	-1		-170.6	-71.1	-74.1	-69.9	-56.4
623475	-82.86	-1		179.1	56.2	179.2	-68.9	-58.3
637976	-86.70	-1	50	-75.6	-177.1	147.9	87.3	49.3
774523	-85.02	0	→ 22	-176.5	50.6	-98.6	-100.1	58.2
900718	-94.63	-1	2	176.5	176.4	121.1	-95.6	55.5
919771	-89.85	-1	21	-77.9	-49.9	145.6	86.6	50.0
986017	-92.26	-1	7	-77.2	179.4	64.0	-110.9	56.8
992902	-97.81	1	1	-80.1	72.7	127.6	-97.7	61.5

number of minima in table 1. When Type = 0, an arrow indicates the minimum (if contained in table 1) in which SPS gets down after an independent local minimization.

(iii) Minimum No. 1 was computed five times by SIGMA and two times by SIGRAC before minimizations stopped. This was the consequence of setting a parameter called NSUC at these values before starting the runs. Now, while NSUC = 2 for SIGRAC has led to minimum No. 1 in *all* runs (with different starting conformations) we have carried out up to now, NSUC = 5 for SIGMA has sometimes failed in reaching minimum No. 1; in addition to the aforementioned case where SIGMA got stuck in minimum No. 2 see, for example, ref. [22].

4. Conclusions

The availability of several molecular mechanical programs that face the problem of conformational analysis under a number of different viewpoints puts us in a good position to obtain multifaceted pieces of information. The particular technique,

or combination of techniques, to be applied in any given situation depends on many factors: previous knowledge from other sources on the structural features of the molecule under examination, availability of computer time, kind of answers hopefully given to the questions raised about this molecule, etc.

In principle, SIGRAC is a most reliable technique for the detection of global minimum. Because of this very reliability, NSUC may be kept to very low values: NSUC = 1 has failed in one case up to now (i.e. the first minimum regarded as potentially global by the program was not the true absolute minimum), but NSUC = 2 has never failed. Unfortunately, the number of function evaluations is high, typically of the order of 10^6 , and every evaluation requires approximately 0.7 sec CPU time on our UNIVAC 1100/72 system for a molecule of the size of SPS.

SIGMA is approximately five times faster than SIGRAC, but NSUC should be given higher values (we have had cases where a molecule got trapped in a relative minimum even with NSUC = 5), so that the degree of confidence is not as high as with SIGRAC.

A preliminary investigation of the whole low-energy surface of a molecule with LECSA is always advisable: this not only gives us a useful cross-check, but also suggests a convenient choice of starting conformations for SIGMA and/or SIGRAC (while the final conformation found with both these programs clearly does not depend on the initial one, the time needed to perform the whole calculation does).

Furthermore, LECSA enables us to compute the statistical weight of conformers and the interconversion pathways for those minima that merge into lower-energy minima from which they are separated through saddle points with energy smaller than a given threshold: for example, as discussed in ref. [21], minimum No. 3 of SPS (cf. table 1) merges into No. 2 after overcoming a barrier of approximately 5 kJ mol^{-1} at 90° along the α -axis.

References

- [1] D.N.J. White, *Molecular Mechanics, Computer-Aided Molecular Design 1984* (IBC Technical Services Ltd., London, 1985).
- [2] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller, *J. Chem. Phys.* 21(1953)1087.
- [3] G. S. Rao, R.S. Thyai and R.K. Mishra, *J. Theor. Biol.* 90(1981)377.
- [4] N.S. Goel, G.S. Rao, M. Ycas, H.J. Bremermann and L. King, *J. Theor. Biol.* 35(1972)399.
- [5] H.J. Bremermann, *Math. Biosci.* 9(1970)1.
- [6] A.V. Levy and A. Montalvo, *SIAM J. Sci. Stat. Comput.* 6(1985)15.
- [7] R.A. Donnelly and J.W. Rogers, Jr., private communication.
- [8] S. Kirkpatrick, C.D. Gelatt, Jr. and M.P. Vecchi, *Science* 220(1983)671.
- [9] L.T. Wille, *Nature* 324(1986)46.
- [10] F. Aluffi-Pentini, V. Parisi and F. Zirilli, *J. Optim. Theory and Applic.* 47(1985)1.
- [11] F. Aluffi-Pentini, V. Parisi and F. Zirilli, *ACM T.O.M.S.*, submitted.
- [12] F. Aluffi-Pentini, V. Parisi and F. Zirilli, *Comp. J.*, in press.

- [13] F. Aluffi-Pentini, V. Parisi and F. Zirilli, ACM T.O.M.S., submitted.
- [14] C. Tosi, R. Pavani, R. Fusco, F. Aluffi-Pentini, V. Parisi and F. Zirilli, Rend. Accad. Naz. Lincei, Classe Sci. Fis. Mat. Nat., Ser. VIII, 78(1985)149.
- [15] Z. Schuss, in: *Theory and Applications of Stochastic Differential Equations* (Wiley, New York, 1980) Ch. 8.
- [16] F. Aluffi-Pentini, V. Parisi and F. Zirilli, to be published.
- [17] R. Fusco, L. Caccianotti and C. Tosi, Nuovo Cim. D8(1986)211.
- [18] R. Fusco, L. Caccianotti and C. Tosi, *Proc. First Donegani Workshop on Strategies for Computer Chemistry* (Kluwer Acad. Publ., Dordrecht, 1989) p. 227.
- [19] C. Tosi, E. Clementi and O. Matsuoka, Biopolymers 17(1978)51.
- [20] C. Tosi, *Theoretical Chemistry in Drug Design: A New Method of Global Optimization*, Conference Papers, *Computer-Aided Molecular Design 1985* (IBC Technical Services Ltd., London, 1986).
- [21] C. Tosi, R. Fusco and L. Caccianotti, Nuovo Cim. D8(1986)219.
- [22] C. Tosi, Internal Report No. 24/88, Istituto Donegani (1988).