Reprinted from

THEO CHEM

Journal of Molecular Structure (Theochem) 501–502 (2000) 509–518

A comparison of some global path-following methods, I.☆

Gy. Dömötör[a], L.L. Stachó[b], M.I. Bán[a,*]

[a]*Institute of Physical Chemistry, JATE University of Szeged H-6701, P.O. Box 105, Szeged, Hungary*
[b]*Bolyai Institute for Mathematics, JATE University of Szeged H-6720 Aradi Vértanúk tere 1, Szeged, Hungary*

ELSEVIER

ELSEVIER

Journal of Molecular Structure (Theochem) 501–502 (2000) 509–518

THEO CHEM

www.elsevier.nl/locate/theochem

A comparison of some global path-following methods, I.☆

Gy. Dömötör[a], L.L. Stachó[b], M.I. Bán[a,*]

[a]*Institute of Physical Chemistry, JATE University of Szeged H-6701, P.O. Box 105, Szeged, Hungary*
[b]*Bolyai Institute for Mathematics, JATE University of Szeged H-6720 Aradi Vértanúk tere 1, Szeged, Hungary*

Abstract

There are two different practical ways of global path-following (PF) on potential energy surfaces (PESs) of molecular reactions: (i) the Elber–Karplus (EK) method (and its improvements), and (ii) the family of DDRP methods. The early versions of the methods under (i) are based on minimizing a functional of the entire path and applying penalty functions as constraints, and are evaluated only for molecular mechanical (MM) PESs. The first true improvement in the EK sequels is the method of Chiu et al. who—instead of using penalty functions—introduce a redistribution of the grid points to substitute the constraints employed in former versions of the EK method, and perform PF on quantum mechanical (QM) PESs. In the present paper the mathematical foundations and the performances of the above methods have been compared. The superiority of the DDRP method in accuracy and stability over the other methods has been verified and tested by a difficult mathematical function simulating the conformational change in the catechol molecule.

Keywords: Elber–Karplus methods; DDRP method; Global path-following; Potential energy surfaces

1. Introduction

In the discussion of reaction mechanisms and especially of the course of a chemical reaction or conformational change in a molecular system, the reaction path (RP) concept plays a most important role. The RP is a sequence of steepest descent paths (SDP) that joins the minima pertaining to the stable states of reactants and products through saddle points (SPs) or transition states (TSs). Fukui's RP, the *intrinsic reaction coordinate* (IRC), is a SDP in mass-weighted Cartesian coordinates, assuming that the reaction takes place infinitely slowly. However, this definition can be extended to any coordinate system by defining the RP as a minimum energy RP (MERP) connecting the minima of reactants and products via the SP/TS. Naturally, from the chemical point of view, the most significant parts of the RPs and/or the potential energy surfaces (PESs) are the critical (or stationary) points (minima and SPs/TSs). Nevertheless, recently theoretical investigations of entire RPs (or reaction profiles) on PESs of molecular systems also attract considerable attention. PF procedures can be classified as local and global ones. Local PF procedures start by searching for a TS and then walking downhill on a SDP towards the minima belonging to reactants and products. Such procedures require many energy and gradient calculations due to the small steps to be taken to avoid leaving the MERP. However, in the majority of cases the location of the TS may not at all be a simple task. The other

☆ Dedicated to Professor R. Gáspár on the occasion of his 80th year.

* Corresponding author.

E-mail addresses: domotor@chem.u-szeged.hu (G. Dömötör), stacho@math.u-szeged.hu (L.L. Stachó), m.i.ban@chem.u-szeged.hu (M.I. Bán).

0166-1280/00/$ - see front matter
PII: S0166-1280(99)00465-0

disadvantage of local procedures is that only a well-predefined part of the PES can be searched for and serious problems are arising when bifurcations of RPs take place or very deep and winding canyons occur on the PES or if the RP passes through more than one TSs as it almost surely happens in the case of large molecular systems. Global methods can "scan" a wide region of the PES thoroughly and can find simultaneously a great number of stationary points, however, at the expense of countless calculations. In the following we are going to compare the theoretical foundations and functioning of the two main families [1] of global PF methods: the Elber–Karplus (EK) method [2] and its improvements [3–6] with special interest to Chiu's method [7], and the family of DDRP methods [8–19].

2. Discussion

In a former paper [17] we observed incorrectnesses in the mathematical formulation of the PF methods based on the EK strategy. In later articles [20,21] we carried out a rigorous mathematical analysis of the foundation of this strategy and showed through examples that one cannot expect useful results by such methods even for mathematical test functions and small chemical systems, to say nothing of large molecules with hundreds of atoms and degrees of freedom. In the present paper we give a short comparative summary of the mathematical bases of the global PF methods under consideration and a new aspect of their uses.

2.1. Mathematical description

The RP in a given system of N atoms is a smooth curve $C: [0,1] \to \mathbb{R}^{3N}$ with arc length proportional parameterization in the coordinate configuration space $\mathbb{R}^{3N}$ connecting two local minima of the energy function $U: \mathbb{R}^{3N} \to \mathbb{R}$ in a manner such that at each point $C(s)$ of the curve C, the function U has a local minimum on the hyperplane $H_s := \{p \in \mathbb{R}^{3N} : \langle p - C(t), (\mathrm{d}/\mathrm{d}s)C(s)\rangle = 0\}$ approaching orthogonally to the tangent vector of C through the point $C(s)$, whenever such a curve exists uniquely. For the sake of unambiguity throughout this section we are going to consider only simple systems where the graph surface of the energy function U admits only one reaction valley. EK developed a method [2] for finding a RP on the basis of the hypothesis that the RP minimizes the functional

$$A: g \mapsto \frac{\int_0^1 U(g(s))\|\partial g/\partial s\| \,\mathrm{d}s}{\int_0^1 \|\partial g/\partial s\| \,\mathrm{d}s} \tag{1}$$

of the *energy average* for all smooth curves $g: [0,1] \to \mathbb{R}^{3N}$ with arc-length proportional parameterization joining the two minima of the energy function U.

We have already proved theoretically [21] that

$$\inf A = \inf U \tag{2}$$

and this infimum (i.e. the largest lower bound of the given infinite set of numbers—the average functional values) will never be attained if the energy function has only two local minima (at the configurations of reactants and products) as is the case in most simple reactions. *Moreover, the RP is no stationary curve in the sense of variational calculus. The average functional A if U has only two local minima and the shape of a curve at low energy average values is in general very far from that of the RP.* Naturally, these results can easily be extended to local minima of any discretional numbers. Therefore it can be stated that *the mathematical foundation of the EK method for searching RP is incorrect.*

In this paper we investigate in detail the behaviour of the original EK method [2] and its best improvement, i.e. the algorithm of Chiu et al. [7], and also provide two further variants (called CHIU1 and CHIU2) of the latter. The numerical computer experiments lead to the following conclusion: most points of the resulting polygons are clustered near the points of minima a, b while the remaining ones can be regarded as an approximation of some smooth curve between a and b. Let us consider any procedure $\mathscr{P}$ with the following properties:

1. $\mathscr{P}$ produces for each $n = 2,3,\ldots$ a polygon $C_n := \{p_1^{(n)},\ldots,p_n^{(n)}\}$ with edges $(p_i^{(n)}, p_{i+1}^{(n)})$ of equal length δ_n such that $p_1^{(n)} = a, p_n^{(n)} = b$.
2. P_n minimizes some approximate form of the functional A for polygons of n equidistant points joining a with b.
3. There exists a curve C joining a with b such that

approximately length $(C)/\delta_n$ consecutive points of the polygon C_n lie on C while the remaining points of C_n are clustered at a distance $\approx \delta_n/2$ from the endpoints a and b.

Notice that this is the case in some numerical methods (e.g. in that of Chiu et al.) which minimize some (approximate) path integral average in accordance with our numerical experiences. We can estimate the value of δ_n as follows. Fix n and consider all polygons $\{p'_1,\ldots,p'_n\}$ of n points possessing the properties (1) and (3) with some unknown value δ instead of δ_n.

Then δ_n is approximately the minimum place in the variable δ of the estimate numerical path integral average $A_n(\delta)$ for a δ-equidistant n-point of which approximately length $(C)/\delta$ terms lie on C and the remaining points are located in a $(\delta/2)$-neighbourhood of a and b, towards the direction of the smallest curvature of the graph of U, the direction of the eigenvector of the Hessian of U. If we write

$$E := \int_C (U - \min U)\,\mathrm{d}s, \qquad L := \text{length}(C) \tag{3}$$

and if $U(x) - U(a)$, $U(x) - U(b) \le k\delta^2$ around the minima then

$$A_n(\delta) \approx \frac{E + \left(n - \dfrac{L}{\delta}\right)k(\delta/2)^2}{n\delta} + \min U. \tag{4}$$

Thus δ_n can be estimated as

$$\delta_n \approx [\delta :\ A_n(\delta) = \min A_n] = 2\sqrt{\frac{E}{kn}}. \tag{5}$$

Thus the value δ_n which measures the accuracy of the approximation is proportional to $1/\sqrt{n}$ with the smallest eigenvalue k as follows. The curve C is near the RP only when a and b are joined by a narrow valley. If the RP passes on ridge segments, the curve C will not approximate the RP. In light of the above observations we can interpret the results so far regarded as positive proofs for the use of the EK method and its numerical variants as follows. To our knowledge all the numerical examples were carried out with a relatively small number of points $n \approx 10$ such that length $(C)/n \approx \delta$. In case of narrow reaction valleys (as with the Müller–Brown function [22]) one can really verify that the points obtained lie near the RP. However, no computer experiments with significantly larger number of points (e.g. with $n \approx 100$) were formerly reported. From our experiences, in such cases the resulting sequence of points does not become clustered quickly in contrast with previous expectations suggested in the literature. In particular, even if the points outside some neighbourhoods of the minima fit the RP well, only the highest SP on the RP can be expected identifiable with some member of the resulting point sequence of an accuracy proportional to $1/\sqrt{n}$. Thus by increasing the number of points we can only slowly improve the density of the sequence below a critical and practically non-negligible value (denoted above by δ_n) even in the most harmless case of a uniquely deep narrow reaction valley. We remark that our DDRP method provides an equidistant sequence distributed on C and hence it gives an accuracy $\delta_n \approx$ length $(C)/n$ for the same numerical effort in the case of n points.

In worse cases with ramified reaction valleys and ridge sections on the RP (as with the "catechol" function) the approximating point sequences cannot "climb" up to these ridge sections because moving towards them with the point sequence the energy average associated increases. In such cases the resulting sequence of points can pass arbitrarily far from some SPs located on the RP.

2.2. A short comparison of the methods based on the EK strategy and the family of DDRP methods

2.2.1. The EK strategy

The general problem is to minimize the line integral

$$S(\mathbf{R}_\mathrm{i}, \mathbf{R}_\mathrm{f})_\mathbf{L} = \frac{1}{L}\int_{\mathbf{R}_\mathrm{i}}^{\mathbf{R}_\mathrm{f}} [E(\mathbf{R})\cdot \mathrm{d}\ell(\mathbf{R})]_\mathbf{L} \tag{6}$$

of the EK method [2] where $E(\mathbf{R})$ is a function (the total energy of the system in the Born–Oppenheimer approximation) dependent upon the N-dimensional nuclear position vector $\mathbf{R}$. $\mathrm{d}\ell(\mathbf{R})$ is a line element on the path $\mathbf{L}$ of length L between the initial and final configurations $\mathbf{R}_\mathrm{i}$ and $\mathbf{R}_\mathrm{f}$. To find the path $\mathbf{L}$ that minimizes $S(\mathbf{R}_\mathrm{i},\mathbf{R}_\mathrm{f})_L$ a discretized form of the line integral

$$S(\mathbf{R}_0, \mathbf{R}_{M+1})_\mathbf{L} = \frac{1}{L}\sum_{j=1}^{M+1} E(\mathbf{R}_j)\cdot\Delta\ell_j \tag{7}$$

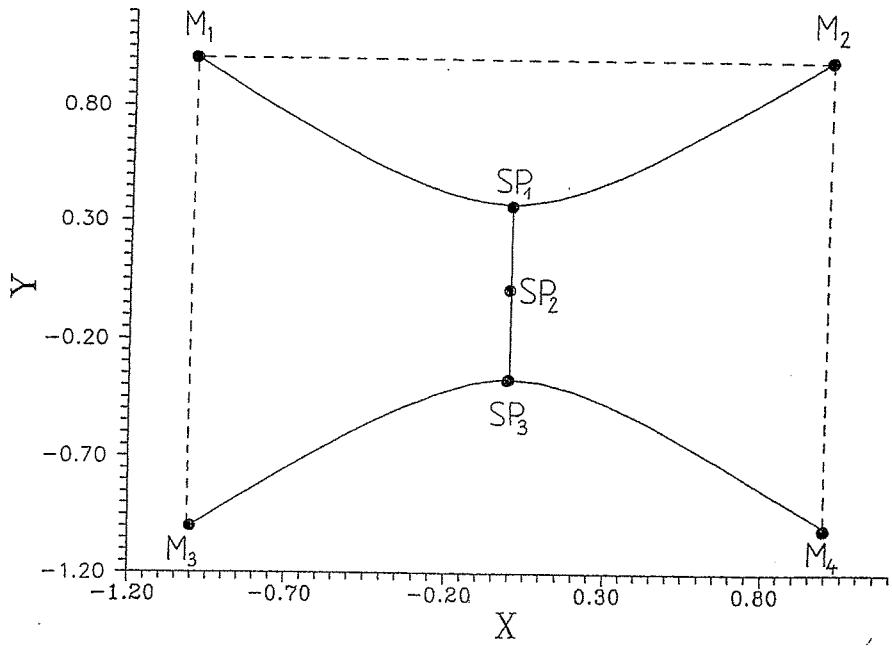

Fig. 1. The entire path of the SB ("catechol") function as calculated by the DDRP method, starting from the open tetragon joining the minima M_3, M_1, M_2 and M_4, as indicated by the broken line, showing the seven critical points and the two bifurcations (100 points).

is employed which includes M intermediate grid points. For eliminating the problems arising from a straightforward minimization of the discretized form of the line integral two penalty functions as constraints have been introduced. Unfortunately, the minimization of the original EK functional was found to be extremely sensitive to the choice of the parameters in the penalty functions, and later amendments [3–6] of the original EK method [2] did not set the method free from its inherent basic problems essentially. To overcome the problems remained Chiu et al. [7] evenly distributed the grid points along the path, nevertheless, also they minimized the discretized line integral.

2.2.2. *The DDRP method*

The algorithm for the simplest case based on the theory [8–10] is characterized by the following steps. Given a curve connecting two local minima of the energy function, the phase flow of the negative gradient takes it to converge to some sequences of the steepest descent (and ascent) paths between the two minima. In practice the successive phase curves can be approximated by the following algorithm. Choose $P_1, \ldots, P_n$ to be consecutive points representing a polygon C. First calculate the effect of the flow of $-\nabla U$ on the points for some (virtual) time. Thus we have to calculate the solutions of the ordinary differential equation $(\mathrm{d}/\mathrm{d}t)\, x(t) = -\nabla U(x(t))$. After some steps we homogenize the polygon by placing further points on or taking off in order to maintain the distance between the consecutive points approximately equal. The procedure consisting of two alternating steps (the production of a new approximation polygon followed by its homogenization) is continued by iteration and stops when the changes between two subsequent approximation polygons remain within a predetermined limit. The DDRP method does not start from the line integral or its discretized form and does not use penalty functions as constraints. The use of such criteria leads in the EK method and its sequels [2–6] to the controversial results we argue against. Chiu et al. [7] also starts from the line integral or its discretized form and use minimization. However, instead of employing penalty functions they introduce a redistribution of the grid points. Unfortunately, this redistribution is essentially the same as the homogenization procedure described earlier in Refs. [8–12]. On the other hand, instead of the minimization technique, the DDRP method is using the negative gradient to guide and control the shifts of the points thus giving uniquely the really safe direction for the

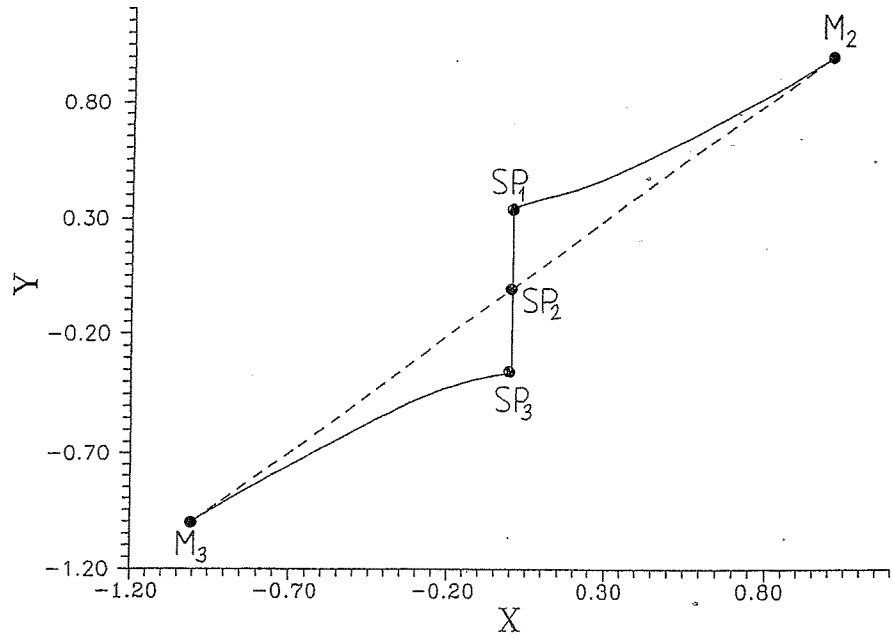

Fig. 2. Path segment 1 starting from the digon joining the minima M_3 and M_2 using the DDRP method (100 points).

PF. Besides it must be emphasized that if calculating the subsequent new approximation curves by the DDRP method computation times would be enormously decreased, especially in ab initio calculations, when we apply analytical gradients. The DDRP method produces the $(n + 1)$th approximation curve from the nth one in such a manner that the progression takes place in the direction of the negative gradient, and proportionally to it. The use of Hessians is unnecessary during the DDRP procedure. Moreover, complete, exact and unambiguous mathematical foundation exists only for the DDRP method in which the

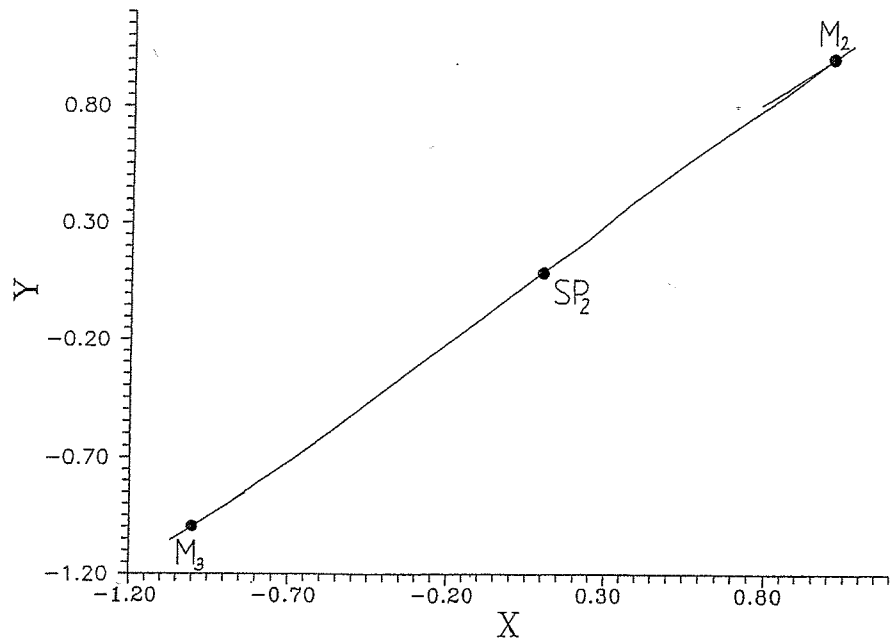

Fig. 3. Path segment 1 starting from the digon joining the minima M_3 and M_2 using the EK method with parameters $\lambda = \lambda' = 0.0001$ (25 points).

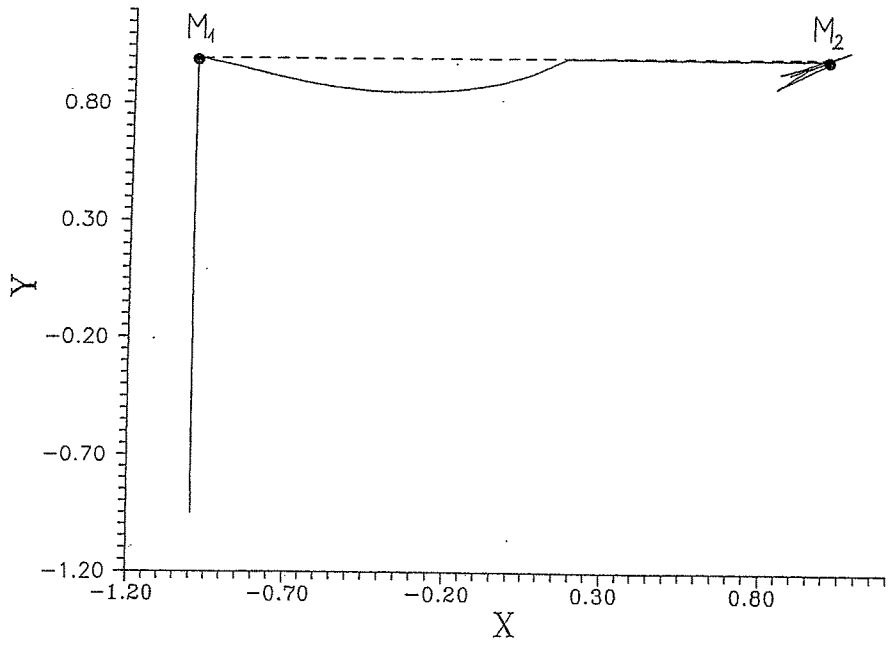

Fig. 4. Path segment 2 starting from the digon joining M_1 and M_2 using the EK method with parameters $\lambda = \lambda' = 0.1$ (25 points).

mathematical basis and the numerical realization are in consistent union and the result of the search is always a true SDP. Here we want to draw the attention to a new path concept occurring in a more recent version based on the EK strategy [23]. Although the path found by Olender and Elber *is not a true SDP* (as stated by the authors), nevertheless—after proper mathematical reformulation [24]—it may give a useful additional description of the RP.

2.3. Comparative results

As a test example the Stachó–Bán (SB) function

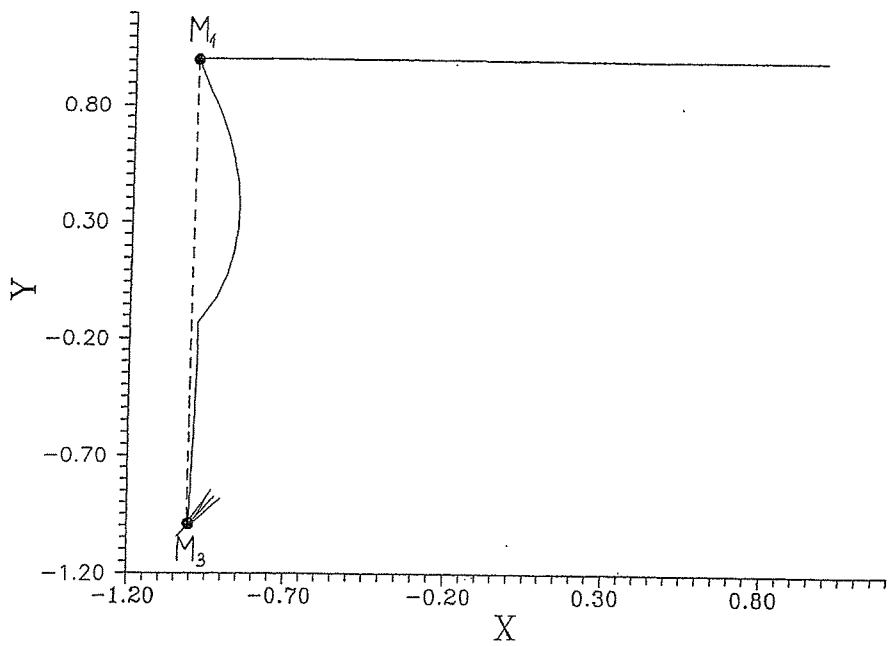

Fig. 5. Path segment 3 starting from the digon joining M_1 and M_3 using the EK method with parameters $\lambda = \lambda' = 0.1$ (25 points).

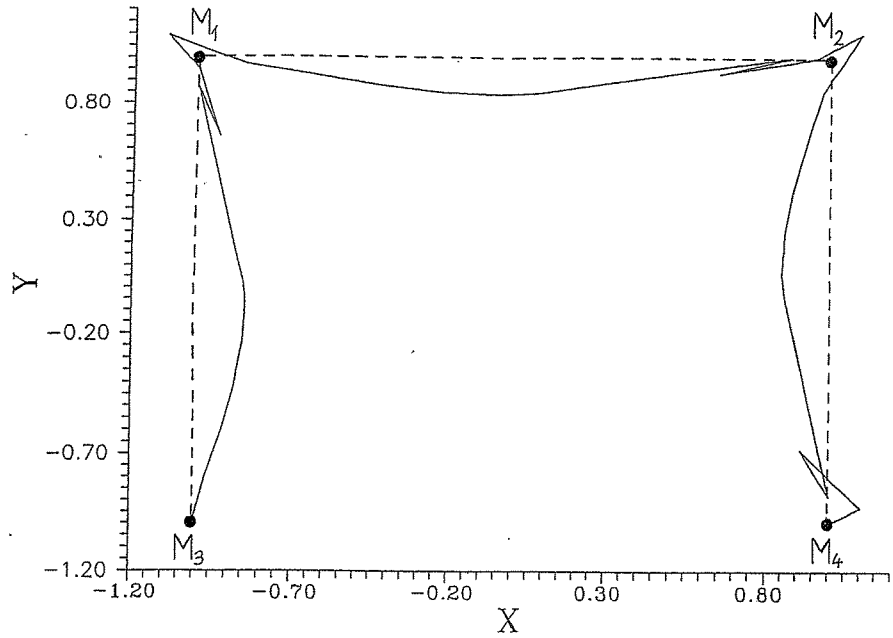

Fig. 6. Path segment 4 starting from the open tetragon joining the minima M_3, M_1, M_2 and M_4, as indicated by the broken line, using the EK method with parameters $\lambda = \lambda' = 1$ (31 points).

[8–10]

$$f(x) = \prod_{j,k=0,1} \|x - ((-1)^j, (-1)^k)\|^2 + (x_1^2 - 1)^3 - (x_2^2 - 1)^3 \quad (8)$$

has been employed. This is a two-variable mathematical function which has four minima (M_1, M_2, M_3 and M_4) at (−1,1); (1,1); (−1, − 1) and (1, − 1), three SPs (SP_1, SP_2 and SP_3) at (0, 0.37213), (0,0) and (0, − 0.37213) and two symmetrical bifurcations at SP_1 and SP_3, and it simulates quite well the

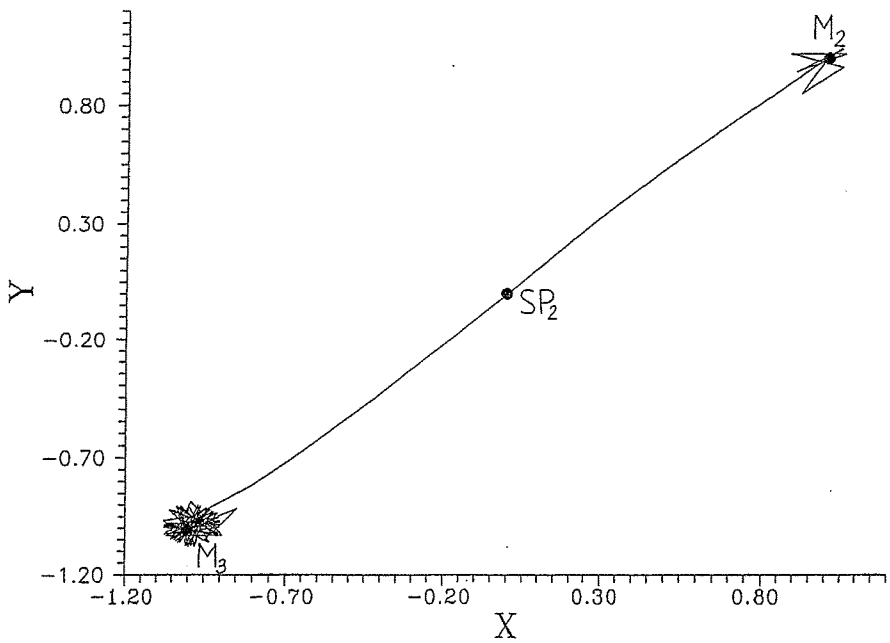

Fig. 7. Path segment 1 starting from the digon joining the minima M_3 and M_2 using the CHIU1 method. (100 points).

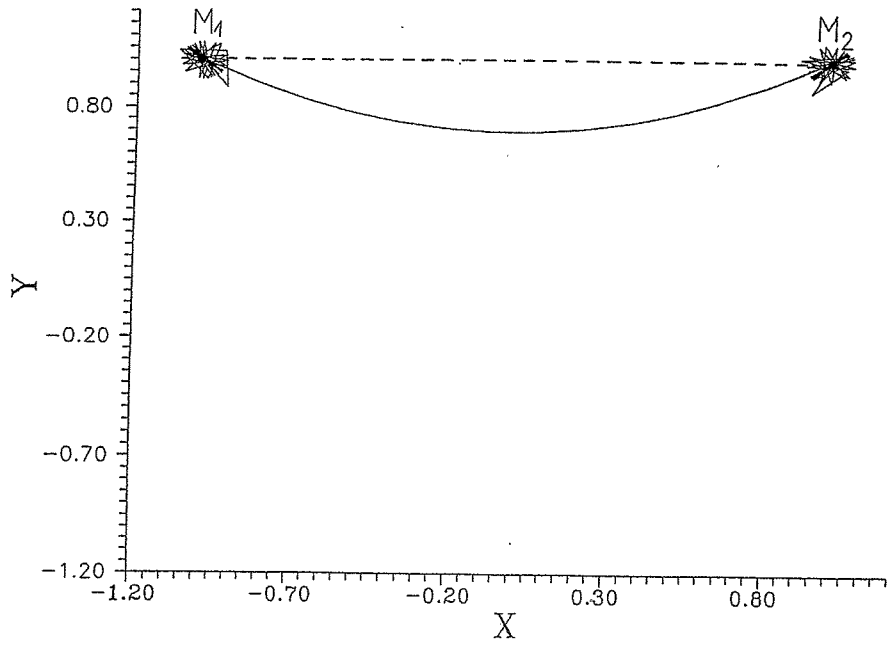

Fig. 8. Path segment 2 starting from the digon joining M_1 and M_2 using the CHIU1 method (100 points).

conformational change in the catechol molecule. We tried to find the entire path (Fig. 1) from different starting polygons using three different global PF methods: the EK method [2], Chiu's method [7] and the DDRP method [8–16]. For minimizing the discretized form of the line integral in the methods of EK and Chiu et al. the Powell minimization technique [25] was used. For the DDRP method only two trials (displayed in Figs. 1 and 2) were made regarding that the remaining cases had already been discussed in Refs. [8,10]. The EK method was employed in its original form and in the legends of figures the

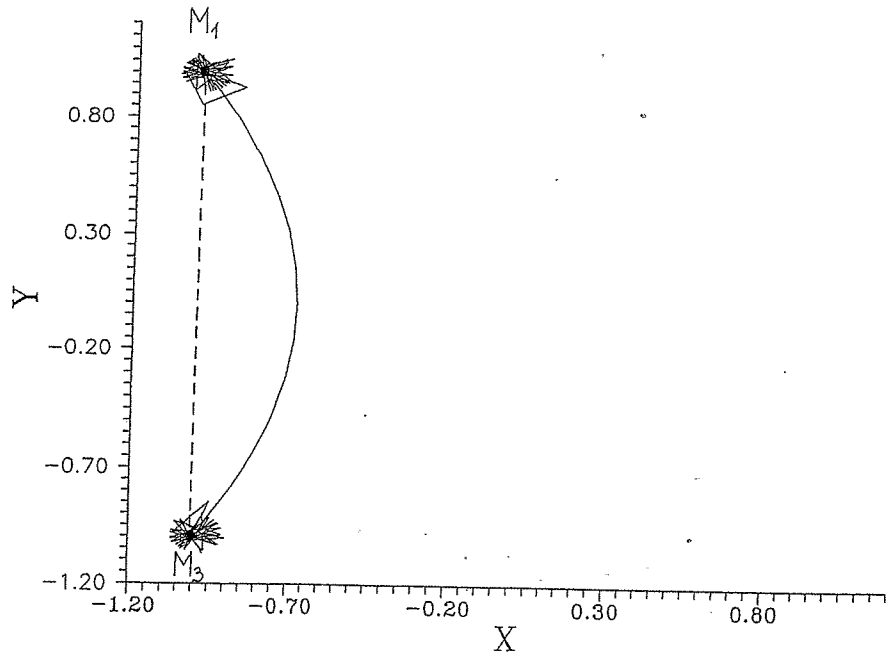

Fig. 9. Path segment 3 starting from the digon joining M_1 and M_3 using the CHIU1 method (100 points).

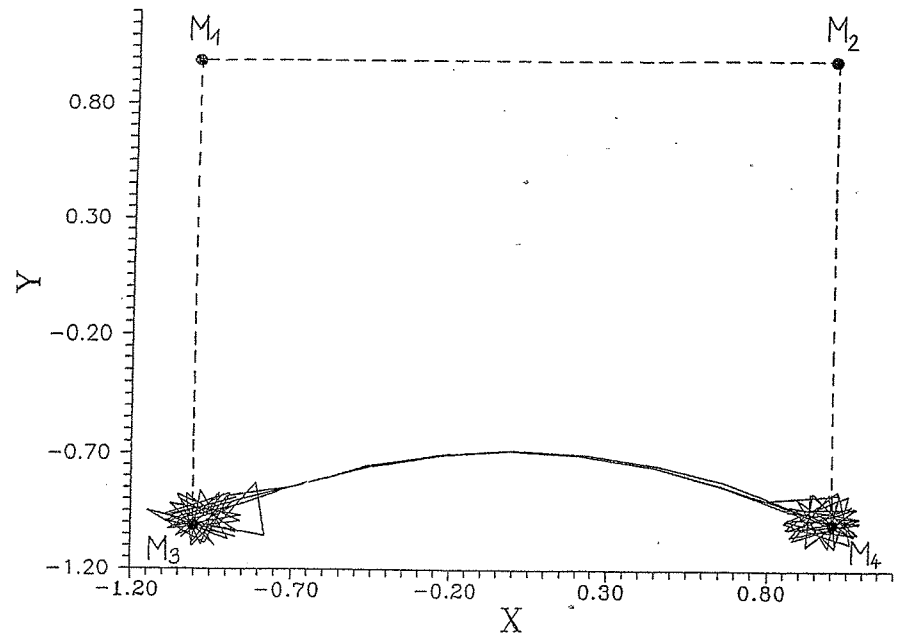

Fig. 10. Path segment 4 starting from the open tetragon joining the minima M_3, M_1, M_2 and M_4, as indicated by the broken line, using the CHIU1 method (100 points).

parameter values and the number of points used (in parentheses) have been indicated. Four trials (displayed in Figs. 3–6) starting from different minima connected with one another (indicated by broken lines) have been made. Since the original Chiu algorithm has not been described with proper thoroughness the principles of the method was used to prepare two algorithms on our own, denoted by

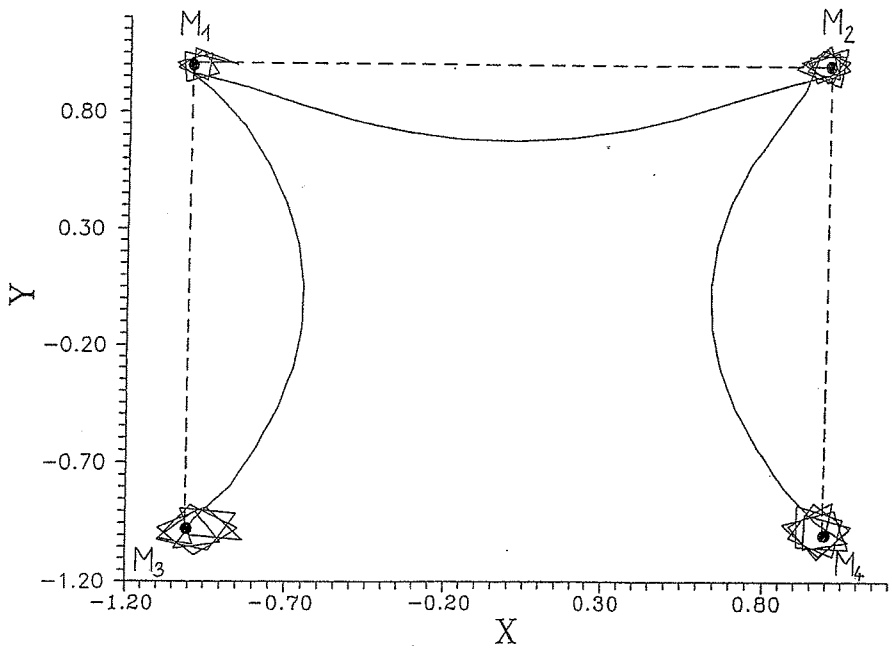

Fig. 11. Path segment 4 starting from the open tetragon joining the minima M_3, M_1, M_2 and M_4, as indicated by the broken line, using the CHIU2 method (100 points).

CHIU1 and CHIU2. In CHIU1 the discretized line integral in the form modified by Chiu et al.

$$S(R_0, R_{M+1})_L = \frac{1}{L} \sum_{j=1}^{M+1} \frac{E(R_j) + E(R_{j-1}\})}{2} \Delta\ell_j \qquad (9)$$

was used, and employing two constraints (fixed endpoints and equilateral approximation polygons) was minimized by the Powell method [25]. In CHIU2 the PF procedure used in the DDRP method was applied to Eq. (9). In this case we have not used minimization, however, we carried out instead the redistribution of the grid points along the path. Four trials for CHIU1 (Figs. 7–10) and one trial for CHIU2 (Fig. 11) are shown.

As one can see from the comparisons of the RPs in the figures only the DDRP method could produce the entire path with all its exact characteristics, starting from an open tetragon (Fig. 1). From the same starting position neither the EK method nor the Chiu variants could reproduce the entire path, and no points of the curves are even in the vicinity of the SPs (see Figs. 6, 10 and 11). The DDRP method starting from a digon connecting M_2 and M_3 reproduces the central part of the SB function with all the three SPs (Fig. 2) however the EK and Chiu's methods do not (Figs. 5 and 9). Nothing better comes out when starting from other minima (see Figs. 3, 4, 7 and 8). The points of the final curves are quite far from those of the true RP and from the SPs.

3. Conclusion

Our numerical computer experiments have shown that a single global PF method (from among the three investigated ones) has the ability to reproduce the entire true RP with all the critical points and bifurcations while the final curves produced by the other two methods (the EK method and the method of Chiu et al.) pass quite far from the true RP and no points of them get in the proximity of the SPs. This fact clearly indicates *the superiority of the DDRP method over other global PF methods.*

Acknowledgements

This work is supported by the Hungarian National Science Foundation, grant no. OTKA T 020743.

References

[1] H.B. Schlegel, in: D.R. Yarkony (Ed.), Modern Electronic Structure Theory. Part I, Advanced Series in Physical Chemistry, 2, World Scientific, Singapore, 1995, p. 482.
[2] R. Elber, M. Karplus, Chem. Phys. Lett. 139 (1987) 375.
[3] R. Czerminski, R. Elber, Proc. Natl Acad. Sci. 86 (1989) 6963.
[4] R. Czerminski, R. Elber, J. Chem. Phys. 92 (1990) 5580.
[5] R. Czerminski, R. Elber, Int. J. Quantum Chem.: Quantum Chem. Symp. 24 (1990) 167.
[6] S. Huston, J.W. Ponder, TINKER: Software Tools for Molecular Design, Version 3.5, October 1997 (Copyright: Jay William Ponder, 1990–1997).
[7] S.S.L. Chiu, J.J.W. McDouall, I.H. Hillier, J. Chem. Soc. Faraday Trans. 90 (1994) 1575.
[8] L.L. Stachó, M.I. Bán, Theor. Chim. Acta 83 (1992) 433.
[9] L.L. Stachó, M.I. Bán, J. Math. Chem. 11 (1992) 405.
[10] L.L. Stachó, M.I. Bán, Theor. Chim. Acta 84 (1993) 535.
[11] L.L. Stachó, M.I. Bán, Comput. Chem. 17 (1993) 21.
[12] Gy. Dömötör, M.I. Bán, L.L. Stachó, J. Comput. Chem. 14 (1993) 1491.
[13] M.I. Bán, Gy. Dömötör, L.L. Stachó, J. Mol. Struct. (Theochem) 311 (1994) 29.
[14] L.L. Stachó, Gy. Dömötör, M.I. Bán, J. Mol. Struct. (Theochem) 337 (1995) 99.
[15] L.L. Stachó, M.I. Bán, J. Math. Chem. 17 (1995) 377.
[16] Gy. Dömötör, M.I. Bán, L.L. Stachó, J. Comput. Chem. 17 (1996) 289.
[17] L.L. Stachó, Gy. Dömötör, M.I. Bán, T. Csendes, J. Mol. Struct. (Theochem) 398–399 (1997) 111.
[18] P.R. Surján, M. Kállay, Gy. Dömötör, L.L. Stachó, M.I. Bán, in: H. Kuzmany, J. Fink, M. Mehring, S. Roth (Eds.), Molecular Nanostructures, Proceedings of the IWEP'97, Kirchberg, Austria, 1997. World Scientific, Singapore, 1998, p. 147.
[19] Gy Dömötör, L.L. Stachó, M.I. Bán, J. Mol. Struct. (Theochem) 455 (1998) 219.
[20] L.L. Stachó, Gy. Dömötör, M.I. Bán, Chem. Phys. Lett. 317 (1999) 328.
[21] L.L. Stachó, Gy. Dömötör, M.I. Bán, J. Math. Chem. (2000) in press.
[22] K. Müller, L.D. Brown, Theor. Chim. Acta 53 (1979) 75.
[23] R. Olender, R. Elber, J. Mol. Struct. (Theochem) 398-399 (1997) 63.
[24] L.L. Stachó, M.I. Bán, Gy. Dömötör, Papers presented at WATOC'99, 5th World Congress of Theoretically Oriented Chemists, London, UK, 1–6 August 1999. Book of Abstracts, pp. 333–335.
[25] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes: The Art of Scientific Computing, Cambridge University Press, Cambridge, UK, 1986, p. 294.