ELSEVIER

24 September 1999

Chemical Physics Letters 311 (1999) 328–334

CHEMICAL PHYSICS LETTERS

www.elsevier.nl/locate/cplett

On the Elber–Karplus reaction path-following method and related procedures

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

[a] *Bolyai Institute, JATE University of Szeged, Aradi Vértanúk tere 1, H-6720 Szeged, Hungary*

[b] *Institute of Physical Chemistry, JATE University of Szeged, P.O. Box 105, H-6720 Szeged, Hungary*

Received 10 March 1998; in final form 9 February 1999

Abstract

It is shown through a mathematical proof and by using simple test examples that the fundamental principles of the method of Elber and Karplus (EK) for determining reaction paths are incorrect. Therefore the method, including its improved versions, and the results obtained with the algorithms based on the stategy of EK, even when they are in concordance with experimental data, should be accepted with reservations.

1. Introduction

In the last two decades, a number of algorithms have been developed for following reaction paths (RP) or intrinsic reaction coordinates (IRC) involving stationary points corresponding to local minima and transition states (TS) on potential energy surfaces (PES) of chemical systems or reactions. Such algorithms can be divided into two main groups: (i) the local (or 'direct') methods [1–11] follow the RP either from the minima uphill towards the saddle point (SP) presuming that there is a single SP along the RP between the two minima representing the stable states of reactants and products, or downhill from the SP on a steepest descent path (SDP) towards the minima, and (ii) the global (or 'indirect') methods [12–21] which carry out the search in the whole configuration space orthogonal to the RP. One of the latter methods – used especially for large molecules with a very large number of degrees of freedom – is the Elber-Karplus (EK) method [13], still in wide use. The method is based upon a principle stating that the RPs are energy average minimizing curves. We believe this mathematical basis to be unsound [22]. Czerminski and Elber [14] amended the original EK method and the modified version has been interfaced with a molecular mechanics program [23]. A recent method using the strategy of the EK method proposed by Chiu et al. [24] for the prediction of whole RPs of large molecular systems eliminates some of the deficiencies of the original EK method. Although the energy average functional has not been minimized, the improved versions of the EK method always refer to the minimization of the line integral (Eq. (1) of EK). This is incorrect and, moreover, there has been some confusion about which variant was used in a particular application [25–27]. Nevertheless, when employing certain techniques, the discretized form (Eq. (2) of EK) can be used in practical calculations. In any case, these techniques must be described clearly as

* Corresponding author. Fax: +36-62-420943; e-mail: m.i.ban@chem.u-szeged.hu

0009-2614/99/$ - see front matter
PII: S0009-2614(99)00832-5

part of the mathematical formulation because there always is a danger; if the mathematical foundation of a method is not sound and/or the numerical realization or algorithm is not consistent with the mathematical formulation, any later mending may not make the method sound. One of our present aims is, therefore, to verify this by a mathematical proof (see Section 2.1) and to show that the procedure without special interventions tailored to the actual problem does not work even for small systems (see Section 3.1). On the other hand, extricating the EK method from its fundamental defects leads to a variant of the DDRP method, a fact which has been overlooked by a number of authors [24–28]. To show the points they have in common and in which they differ, a comparison of the two path-following methods is given in Section 3. As to our mathematical investigation [29], the method of Olender and Elber [30] does not supply conventional RPs; however, it operates with a new concept of the RP which is probably not a true SDP in the mathematical sense, yet the solutions of such a variational problem may approximate or even replace the old concept of the RP. Another purpose of this Letter is to distance our global DDRP method [17–21] from the methods based on the EK strategy [28]. The DDRP method also proved to be successful in the interpretation of the properties of large systems (e.g. fullerenes [31]) by using a quantum-chemical method in calculating the energy functional.

To summarize, it can be concluded that, although starting from wrong theoretical arguments, the EK method and its sequels may produce realistic results. Nevertheless, the correctness of such results has not been verified; our experience shows that unambiguous results related to the RP cannot be obtained by the EK method even for quite simple systems.

2. Discussion

2.1. Theoretical considerations

In a forthcoming paper [29], we give a rigorous mathematical investigation of the energy average functional

$$A(C) = \int_C U \, \mathrm{d}s / L(C)$$

where $\int_c U \mathrm{d}s = \lim_{n\to\infty} \sum_{k=1}^{n} UC[k\beta/n + (n-k)\times \alpha/n] \| C[k\beta/n + (n-k)\alpha/n] - C[(k-1)\beta/n + (n-k+1)\alpha/n] \|$ is the energy integral and $L(C) = \lim_{n\to\infty} \sum_{k=1}^{n} \| C[k\beta/n + (n-k)\alpha/n] - C[(k-1)\beta/n + (n-k+1)\alpha/n] \|$ is the length of a curve $C:[\alpha,\beta] \to \mathbf{R}^N$ joining two local minima $a = C(\alpha)$ and $b = C(\beta)$ of the energy function $U:\mathbf{R}^N \to \mathbf{R}$. Here only the main points of the mathematical investigation will be described. According to the fundamental hypothesis of the EK method [13], each minimized curve of the average A is an RP between a and b. One can disprove this argument even with elementary mathematical tools. Let us start from any curve C joining the points a and b and let c be a global minimum point of U (possibly $c = a$ or $c = b$). Let C', C'' be any curves leading from b to c and back from c to b, respectively, and for every $n = 1,2,\ldots$ let $C_n = C \cup C' \cup K_n \cup C''$ where K_n is a closed curve which starts and ends at the minimum point c and passes in the sphere S_n of radius $1/n$ with a length at least $n \cdot L(C \cup C' \cup C'')$. Clearly the curves C_n join a with b and larger and larger pieces of them pass in smaller and smaller neighbourhoods of the minimum c. Therefore their energy averages converge with $\lim_{n\to\infty} A(C_n) = U(c) = \min U$. Conclusion: *If the energy function is not constant along the RP (that is always the case for chemical systems) and C^* is an arbitrary curve between a and b then*

$$\inf A = \min U < A(C^*) .$$

In accordance with the above construction, the numerical attempts for minimizing the average A with discretized curves lead to polygons with points clustering mostly around low energy values [13] hence the efforts of the numerical realizations to build constraints into the minimization of A to force convergency of the 'reasonable' piece of the approximating polygons by deleting the clustering part of them. At this point we emphasize that, besides the minimization of A, for a fair description of the resulting curves the mathematical model of constraint should also be explicitly mentioned. However, there is a theoretical indication showing that a curve other than the RP can be obtained by such methods even in realistic cases. In Ref. [29] we prove the following fact. *A local minimum of the functional A*

– in the sense of variational calculus – cannot be a RP (with non-constant energy values).

Next we review two numerical implementations of the (theoretically false) principle of minimal energy average for constructing approximate RP when using the EK strategy. The first is the original EK method [13] that introduces penalty functions along with a Powell minimization [32] of A with the apparent aim to inhibit clustering of points generated by the iteration. Numerical experiences with a 2-variable function show (see Section 3.1) that the choice of different penalties may lead to completely different results. The second one is that of Chiu et al. [24] which is the best improvement so far. They introduced a correction in the numerical EK method by suggesting the use of equidistant redistribution of the points on the approximating polygons. It is possible to give an infinitesimal mathematical interpretation of the limit curves of Chiu's method with the conclusion that it produces curves which can be characterized by an Euler–Lagrange-type [33] equation giving no RP description. We remark also that given a set of generalized spherical coordinates [33], there is a unique polygon up to an isometry joining a and b with equal consecutive side lengths such that the direction vectors of the consecutive sides should have the given angles of the generalized spherical coordinates. In this manner, we can produce a more stable variant of Chiu's method which relies upon unconstrained minimization with respect to spherical coordinates. As for the numerical results with Chiu's method, we have found that they provide practically good but theoretically inaccurate RP approximations if the RP is a minimum energy RP (MERP) passing at the bottom of a deep narrow valley as in the case of the Müller–Brown [21,34] function. On the other hand, if the RP is not a MERP, as in the case of the Stachó–Bán model function of the conformational changes in a catechol molecule [17], we shall have remarkably different behaviour. In a neighbourhood of a piece of RP where the graph of a 2-variable energy function changes from valley to ridge [17,19] equidistant EK-polygons bypass the piece of the RP on the ridge at significantly lower energy levels. From a theoretical point of view, this observation may be the starting point for some criticism of the usual RP concept; nevertheless, it can offer no relief for the EK method.

3. Comparison of the two path-following strategies: EK and DDRP

To make distinctions between the strategies of the two global path-following methods EK and DDRP, without going into details of the DDRP method [17–21] it is worth mentioning its main charateristic features and those of the methods based on the strategy of the EK method [13,14,23,24,30] The DDRP method does not start from the line integral, it does not minimize this 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) to the controversial results we argue against. Chiu et al. [24] are also starting from the line integral (or from 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 *homogenization* procedure described in Refs. [17–21]. On the other hand, 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 search, and the parameters used in the search serve only for controlling convergency. Moreover, exact, complete and unambiguous mathematical foundations exist only for a single global path-following method, the DDRP method, in which the mathematical basis and the numerical realization are in consistent union, and, as proven by the mathematical arguments, *the result of the search is always a true SDP.*

3.1. Tests with mathematical functions

(a) In order to follow theoretically and also practically the considerations described in Section 2.1, we have chosen the test function

$$F(r_1,r_2) = \left[(r_1-1)^2+(r_2-2)^2\right] \cdot \left[(r_1-2)^2+(r_2-1)^2\right] - 10$$

where

$$r_1 = \sqrt{(x_1-x_2)^2+(y_1-y_2)^2+(z_1-z_2)^2}\,,$$

$$r_2 = \sqrt{(x_2-x_3)^2+(y_2-y_3)^2+(z_2-z_3)^2}\,.$$

The variables r_i can be regarded as the distances $r_1 = \|P_1 - P_2\|$ and $r_2 = \|P_2 - P_3\|$ between the points $P_k = (x_k, y_k, z_k)$ $(k = 1,2,3)$ in the 3-dimensional space $\mathbf{R}^3$. It can easily be realized that this test example is a simulated collinear H_3 (or $H_2 + H$) system except that its minima are not at infinity; instead of using the actual potential energies, an analytic function has been employed to facilitate the calculations. Therefore the test function F, like the potential energy functions of a system of 3 atoms, is invariant under all isometric transformations (shifts, reflections, rotations) of $\mathbf{R}^3$. Thus the algorithm involving the penalty functions (4) and (7) in Ref. [13] can be applied to F. The function F has its minimum for $(r_1, r_2) = (1,2),(2,1)$ and its remaining stationary point is a SP with $(r_1, r_2) = (1.5,1.5)$. Furthermore, the RP of F in the configuration space (r_1, r_2) is the straight line segment joining (1,2) with (2,1).

In our first numerical experiment, we have started the EK algorithm with the straight-line segment $L_0 = \{(P_1, P_2(t), P_3): 0 \leqslant t \leqslant 1\}$ where $P_1 = (-1.5,0,0)$, $P_3 = (1.5,0,0)$ and $P_2(t) = (t - 0.5,0,0)$ which is a RP itself. In the course of the calculations, we represented the curves $L_0, L_1, L_2, \ldots$ by 21 points. At the beginning we set $R_i = (x_{1,i}, y_{1,i}, z_{1,i}, x_{2,i}, y_{2,i}, z_{2,i}, x_{3,i}, y_{3,i}, z_{3,i})$ $(i = 0, \ldots, 20)$ that is $R_i = (P_1, P_2(\frac{i}{20}), P_3)$. With the notation

$$\Delta\ell_i = \|R_i - R_{i-1}\|,$$

$$\Delta\ell = \tfrac{1}{20}(\Delta\ell_1 + \cdots + \Delta\ell_{20}), r$$

$$\Delta t_i = R_i^0 \times M(R_i - R_i^0) + M(R_i - R_i^0)$$

used in Ref. [13] where R_i^0 is the position vector of the ith point at the beginning of the calculation and M stands for the diagonal matrix of atomic masses (the unit matrix in our experiment) and the product $R_i^0 \times (R_i - R_i^0)$ is an abbreviation for $(P_{1,i}^0 \times (P_{1,i} - P_{1,i}^0), P_{2,i}^0 \times (P_{2,i} - P_{2,i}^0), P_{3,i}^0 \times (P_{3,i} - P_{3,i}^0))$ in terms of the usual vectorial product. The two penalty functions introduced in Ref. [13] are

$$Q_1(\lambda, L) = \sum_i \lambda(\Delta\ell_i - \Delta\ell)^2,$$

$$Q_2(\lambda', L) = \sum_i \lambda'(\Delta t_i)^2$$

where λ and λ' are two free positive parameters. We have chosen the values of λ and λ' to be equal (as it was done by EK [13]). According to the suggestions in Ref. [13], the functional

$$T(R_0, \ldots, R_{20}, \lambda, \lambda')_L = S(R_0, \ldots, R_{20})_L + Q_1(\lambda, L) + Q_2(\lambda', L)$$

should be minimized by the Powell algorithm [32] where

$$S(R_0, \ldots, R_{20})_L = \frac{1}{\|L\|} \sum_i F(R_i) \Delta\ell_i$$

for a curve L is obtained during the iterations (started from L_0) by Powell minimizations of T and represented by the points $R_0, \ldots, R_{20}$ where $\|L\| = \Sigma_i \Delta\ell_i$. $S(R_0, \ldots, R_{20}, \lambda, \lambda')$, is a discretized approximation of the averaged path integral

$$S^*(L) = \|L\|^{-1} \int_L F(R)\, \mathrm{d}\ell(R).$$

Some characteristic results are in Tables 1 and 2. It is clear that for large values of λ the penalty functions prohibit the decrease of the target functional $S(R_0, \ldots, R_{20})_L$. Thus, in the range $1000 \leqslant \lambda \leqslant 10000$, the decrease of S is nearly blocked. This blocking effect is weaker for smaller values of $\lambda (= \lambda')$, and with the choice $\lambda = 0.0001$ we reach almost the theoretical infimum value -10. However, in accordance with the theoretical predictions of Section 2.1, the points $R_0, \ldots, R_{20}$ of the final curve L (with $\lambda = \lambda' = 0.0001$) cluster in a small neighbourhood of the two minima of F. Therefore, the curve L

Table 1
Some characteristic values of the target functional and lengths of the path-following polygon in the range $1 \leqslant \lambda = \lambda' \leqslant 10000$

$\lambda (= \lambda')$	10 000	1000	100	10	1
$S(R_0, \ldots)_L$	−9.866 668 8	−9.866 680 7	−9.867 974 4	−9.866 799 4	−9.882 370 9
$\|L\| (= \Sigma_i \Delta\ell_i)$	1.414 210 6	1.414 273 8	1.414 815 4	1.420 177 3	1.527 970 3

Table 2
Some characteristic values of the target functional and lengths of the path-following polygon in the range $0.0001 \leqslant \lambda = \lambda' \leqslant 0.1$

$\lambda(=\lambda')$	0.1	0.01	0.001	0.0001
$S(R_0,\ldots)_L$	−9.954 222 0	−9.980 810 8	−9.999 986 5	−9.999 998
$\|L\|(=\Sigma_i \Delta\ell_i)$	2.573 447 3	5.331 357 7	6.967 402 2	11.902 849

thus obtained can by no means be considered as a final approximation of the RP. Choosing any initial curve with parameters in the range $100 \leqslant \lambda = \lambda' \leqslant 1000$ the algorithm does not essentially change the points R_i. Of course, the reason for this effect is not the fact that we have found the RP. It is obvious from the third line of Tables 1 and 2 that at small values of the penalty parameters $\lambda = \lambda'$ the length of the curve obtained by the minimization procedure becomes larger and larger.

(b) One of the best examples most frequently used in testing ab initio methods and path-following procedures is the Müller–Brown (MB) problem [21,34]. Through scrutiny of this test example, not only the wrong working of the EK method has been illustrated for a case quite far from the complexity of a large chemical or biochemical system with a large number of atoms and degrees of freedom but, going beyond this aim, it has been proven that by presuming $\lambda = \lambda'$ the RP (and its characteristics) of the MB problem can not be reproduced over a wide range (between 0.0001 and 3000) of parameter values. Fig. 1 shows the best curves approximating the RP of the MB problem calculated by the EK method starting the calculations from a straight line, with $n = 25$ equidistant points on it, between the two farthest minima. Between $0.0001 \leqslant \lambda \leqslant 1.0$ the RP approximation curves are similar to curves 4–5 of Fig. 1 and, in the 500–3000 range, they basically coincide and lie between curves 1 and 2. Even the best

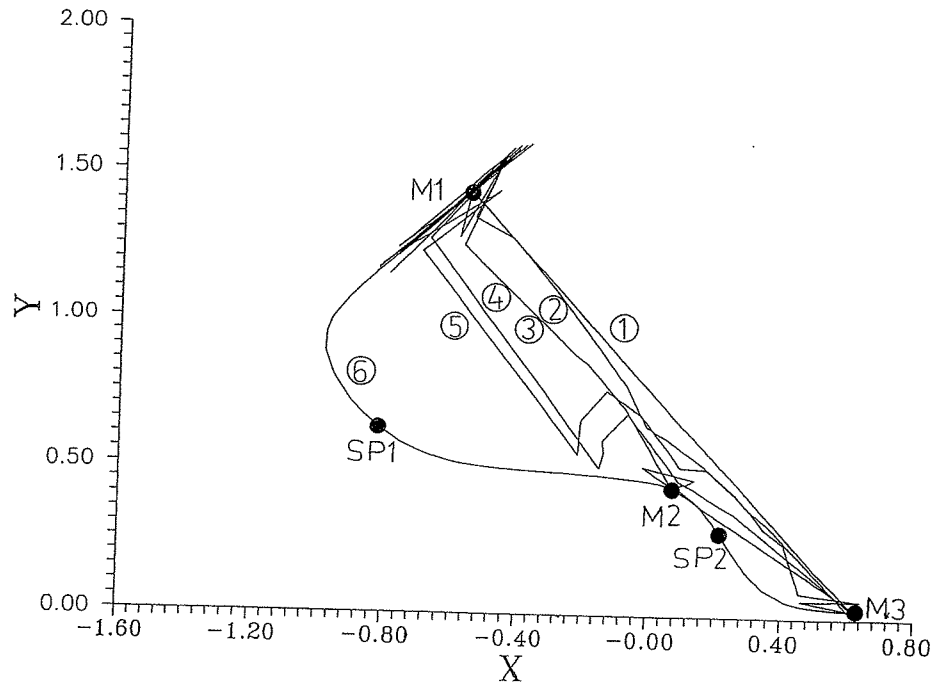

Fig. 1. Path-following of the MB problem by the EK method. The numbered curves 1–5 were obtained by using the parameter values ($\lambda = \lambda'$) as follows: (1) 500, (2) 30, (3) 0.2, (4) 0.1 and (5) 5. Curve (6) is the final MB RP showing the positions of the stationary points as obtained by the DDRP method.

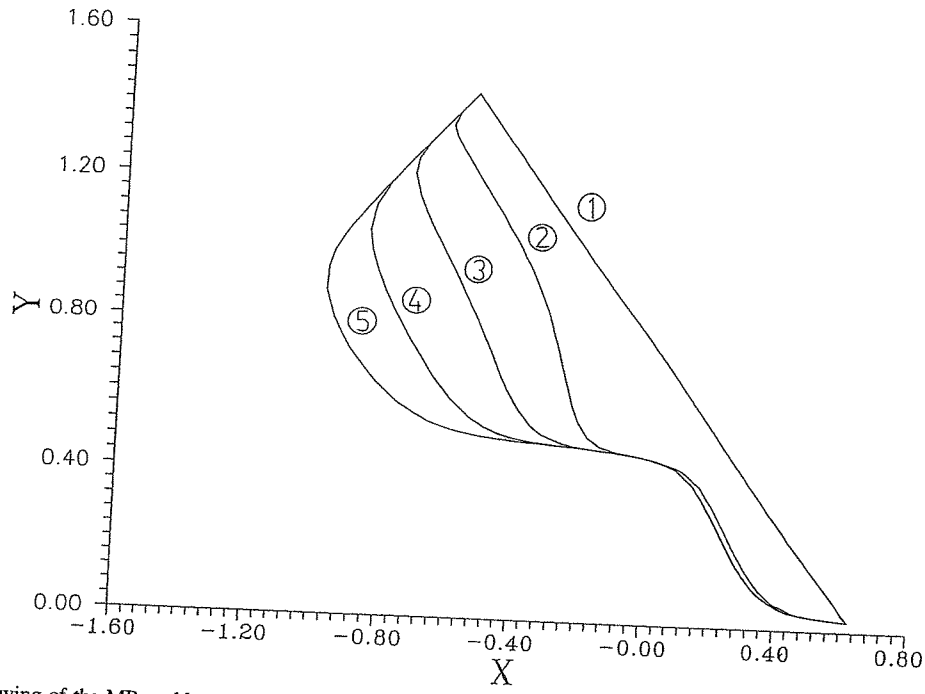

Fig. 2. Path-following of the MB problem by the DDRP method. Typical RP approximations have been selected from the series of curves between the initial straight line joining the two farthest minima and the polygon converged to the real RP [21].

approximation RPs in the 1.0–500 range (curves 2–5) differ from each other and cannot follow smoothly the bends of the real RP as calculated by the DDRP method [21] (curve 6). Moreover, most of the EK curves miss the stationary points and even pass far from the third minimum (M2 of Fig. 1) and the two SPs (SP1 and SP2 of Fig. 1). It is especially true for curve 2 of Fig. 1 which was adjusted to get a best value (101.230; experimental value, 106.035) for the barrier height (the function value of SP1 using M1 as the reference point) by choosing the parameters $\lambda = \lambda' = 0.1$. The determination of the RP by the DDRP method is illustrated in Fig. 2. Starting again from a straight line with $n = 25$ equidistant points, the consecutive approximation curves/polygons fit smoothly closer and closer to the final real RP, giving all the three minima and the two SPs [21]. Here no penalty functions (or adjusting parameters) are needed except parameters controlling the convergency of the procedure, the speed of the search and the smoothness of the curves. The end of the search is indicated not by minimized function values but by the final curve which must not vary significantly by shifting the points in the direction of the negative gradient when the procedure is continued (satisfying e.g. the Hausdorff criterion [20]).

4. Conclusions

(i) It has been proven that the line integral used in the methods employing the EK strategy without constraints has no minimum.

(ii) Using polygons with points/vertices of discrete numbers and without using penalty functions the results obtained are not useful.

(iii) When using penalty functions, the resulting curve can be interpreted as a sort of RP approximation; however, the larger the values of the parameters, the more the minimization of the penalty functions and not of the energy of the system will take place.

(iv) The best way to get the right RP is not by minimizing the energy averages calculated in the vertices of the polygons, as done by the EK strategy, but by searching for the consecutive approximating

polygons along the negative gradient jointly with the homogenization procedure as proposed by the DDRP strategy. Then the end of the search will be indicated not by the minimum but by the condition when the convergency of the RP-approximating curve/polygon reaches a predetermined limit or two subsequent curves are practically unchanged. The DDRP method does not require special starting positions (e.g. endpoints in minima) and/or an initial straight line for the 0th approximation of the RP. The DDRP method is very stable and no system has been found for which it failed.

Acknowledgements

This work was supported by the Hungarian Scientific Research Fund OTKA, Grant No. T 020743.

References

[1] J.W. McIver, A. Komornicki, J. Am. Chem. Soc. 96 (1974) 5798.
[2] K. Ishida, K. Morokuma, A. Komornicki, J. Chem. Phys. 66 (1977) 2153.
[3] M. Dupuis, H.F. King, J. Chem. Phys. 68 (1978) 3998.
[4] C.J. Cerjan, W.H. Miller, J. Chem. Phys. 75 (1981) 2800.
[5] S. Bell, J.S. Crighton, J. Chem. Phys. 80 (1984) 2464.
[6] M.J.S. Dewar, E.F. Healy, J.J.P. Stewart, J. Chem. Soc. Faraday Trans. 2 (1984) 227.
[7] I. Bálint, M.I. Bán, Int. J. Quantum Chem. 25 (1984) 667.
[8] J. Baker, J. Comput. Chem. 7 (1986) 385.
[9] P.G. Jasien, R. Shepard, Int. J. Quantum Chem. Quantum Biol. Symp. 22 (1988) 183.
[10] P. Valtazanos, S.T. Elbert, S. Xantheas, K. Ruedenberg, Theor. Chim. Acta 78 (1991) 287.
[11] C. Gonzalez, H.B. Schlegel, J. Chem. Phys. 95 (1991) 5853.
[12] L.R. Pratt, J. Chem. Phys. 85 (1986) 5045.
[13] R. Elber, M. Karplus, Chem. Phys. Lett. 139 (1987) 375.
[14] R. Czerminski, R. Elber, J. Chem. Phys. 92 (1990) 5580.
[15] D.A. Liotard, Int. J. Quantum. Chem. 43 (1992) 723.
[16] S. Fisher, M. Karplus, Chem. Phys. Lett. 194 (1992) 252.
[17] L.L. Stachó, M.I. Bán, Theor. Chim. Acta 83 (1992) 433.
[18] L.L. Stachó, M.I. Bán, J. Math. Chem. 11 (1992) 405.
[19] L.L. Stachó, M.I. Bán, Theor. Chim. Acta 84 (1993) 535.
[20] Gy. Dömötör, M.I. Bán, L.L. Stachó, J. Comput. Chem. 14 (1993) 1491.
[21] M.I. Bán, Gy. Dömötör, L.L. Stachó, J. Mol. Struct. (Theochem) 311 (1994) 29.
[22] L.L. Stachó, Gy. Dömötör, M.I. Bán, T. Csendes, J. Mol. Struct. (Theochem) 398-399 (1997) 111.
[23] S. Huston, J.W. Ponder, TINKER: Software Tools for Molecular Design, Version 3.5, October 1997 (Copyright Jay William Ponder, 1990-97).
[24] S.S.L. Chiu, J.J.W. McDouall, I.H. Hillier, J. Chem. Soc. Farad. Trans. 90 (1994) 1575.
[25] J.E. Straub, J.K. Choi, J. Phys. Chem. 98 (1994) 10978.
[26] M.A. Collins, Adv. Chem. Phys. 93 (1996) 389.
[27] A. Ulitsky, D. Shalloway, J. Chem. Phys. 106 (1997) 10099.
[28] H.B. Schlegel, in: Modern Electronic Structure Theory, Part I. in Advanced Series in Physical Chemistry, Vol.2, ed. D.R. Yarkony (World Scientific Publ. Co., Singapore, 1995) p.482.
[29] L.L. Stachó, Gy. Dömötör, M.I. Bán, J. Math. Chem., in press.
[30] R. Olender, R. Elber, J. Mol. Struct. (Theochem) 398-399 (1997) 63.
[31] P.R. Surján, M. Kállay, Gy. Dömötör, L. Stachó, M.I. Bán, in: Proc. IWEP'97, Kirchberg in Tirol, Austria, 1997; in: Molecular Nanostructures eds. H. Kuzmany, J. Fink, M. Mehring, S. Roth (World Scientific, Singapore, 1998) p. 147.
[32] W.H. Press, B.P. Flannery, S.A. Teukolsky, W.T. Vetterling, Numerical Recipes: The Art of Scientific Computing (Cambridge University Press, Cambridge, 1986) p. 294.
[33] G.A. Korn, Th.M. Korn, Mathematical Handbook for Scientists and Engineers (McGraw-Hill Book Company, New York, 1961)
[34] K. Müller, L.D. Brown, Theor. Chim. Acta 53 (1979) 75.