A gradient-only algorithm for tracing a reaction path uphill to the saddle of a potential energy surface

Wolfgang Quapp

Mathematisches Institut, Universität Leipzig, Augustus-Platz, D-04109 Leipzig, Germany

Received 15 December 1995; in final form 2 February 1996

Abstract

We propose a procedure to follow the weakest ascent along a valley using gradient-only calculations. The device allows one to search for saddles starting somewhere near a minimum provided that the saddle is connected with the minimum by a valley. We define a local criterion for a minimum energy path by comparison of gradients. The algorithm is a predictor–corrector method using two parameters: step length and tolerance. There is no need for a guess of the saddle point region. We calculate valley pathways on 2D test examples, on the HCN surface, and on a 12D potential of an argon 4-cluster.

1. Introduction

A quarter of a century ago, a mathematical definition of the chemical reaction path (RP) on the potential energy surface (PES) of reacting molecules was given [1–3]. The RP concept is to follow the steepest descent from the saddle point to the reactant or product minimum by tracing the negative gradient of the PES downhill. Since then, the concept of the RP of PES has been of continual importance in theoretical chemistry [4–9]. It is basically a curve in the configuration space of the atoms forming the chemical system. The curve connects two minimizers of the PES. The energy profile over the reaction path should be a 'valley floor' leading through the point of highest energy, the saddle point (SP) of index 1. This point corresponds to the transition structure of 'transition state theory'.

The fundamental problem in handling n-dimensional hypersurfaces, $E = E(x^1, \ldots, x^n)$, is the dimension. Molecules with more than $N = 4$ atoms cause an overwhelming number of net points in the dimension $n = 3N$. The RP concept is promising, because it reduces the problem to finding an algorithm for one-dimensional curves of the PES, without knowledge of the whole PES. Any parametrization s of the RP $x(s) = (x^1(s), \ldots, x^n(s))^T$ is called the reaction coordinate. Different mathematical definitions of the RP may produce different lines, outside the stationary points. From the point of view of a chemical application, this is not disturbing, if only the energy of the SP, $E(x_{sp})$, is of interest as it is in transition state theory.

In theoretical chemistry, the so-called minimum energy path (MEP) is identified with the steepest descent from the SP, calculated in mass-weighted coordinates [1–3]. The path is also called the intrinsic reaction coordinate (IRC) [2,4,10]. Despite its success in theoretical chemistry, the MEP has two imperfections which are of particular interest.

(i) There exists no possibility of turning the direction of the search in order to go uphill from the minimum to the saddle point by gradient-only steps. (If we have a guess of the SP, there are some methods, see Ref. [11], and if we have a guess of the product region, see Refs. [8,12] and references therein for further methods. Eigenvector-following devices and tracing of gradient extremals [13–19] use no lower than second derivatives of the energy functional.)

(ii) Steepest descent works well downhill for steep slopes. However, along the bottom of the valley, it shows a numerical disadvantage, the so-called zigzagging across the valley ground line, cf. Ref. [20].

In this Letter we propose an algorithm which allows us to go uphill along a valley bottom in the direction of the weakest ascent by using a controlled zigzagging procedure. We should realize regarding this idea that the direction of the scaled gradient vector of the point x_0 pointing to x_1 does not agree with the direction of the gradient vector in x_1 in the general case. However, sometimes this happens, see Fig. 1. In this case we presume that we are on a valley pathway! Controlling a coincidence in the two gradients is achieved by means of the scalar product of the normalized vectors.
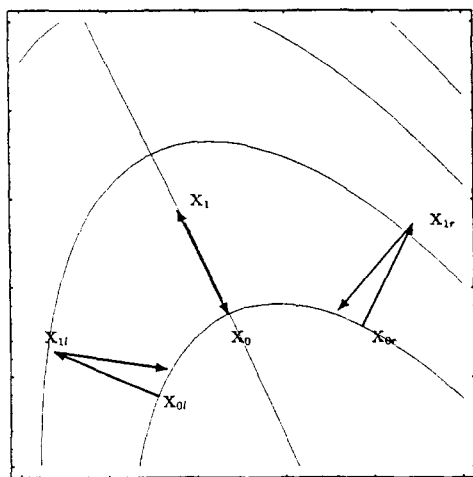

Fig. 1. Scheme of three different points on and near an MEP and their gradient vectors: in x_{0l}, x_0, and x_{0r} the gradients are drawn; however, in x_{1l}, x_1, and x_{1r}, the negative gradients are drawn. The vectors -grad(x_{1r}) and -grad(x_{1l}) are used for corrector steps for an MEP search uphill, see text.

2. The algorithm

We use the normalized gradient vector, grad(x) $= \nabla E(x)/|\nabla E(x)|$. It is the vector of the first derivatives of the PES functional of unit length. We start with a new definition of a valley pathway.

Definition. A point x belongs to a q-minimum energy path (qMEP) if the vector equation holds

$$\text{grad}(x) = \text{grad}(x_q), \quad \text{where } x_q = x + q\,\text{grad}(x),$$

(1)

and $q > 0$ is a step length parameter. (The definition is also satisfied by points on a 'q-ridge'.)

This definition contains a local characterization of certain points x. Thus, the definition does not need an initial condition of the pathway, as opposed to the IRC, which does need a SP [2,10]. The definition compares differences of gradient vectors. However, it does not use higher derivatives, as is necessary using the local definition of gradient extremals [16–18]. If q is sufficiently small, and if the point x fulfills the condition of Eq. (1) then we are near the so-called asymptotic steepest descent path [21], a line defined by the confluence of many steepest descent lines from the right and from the left-hand side into the streambed of the valley ground. Note: the asymptotic steepest descent path may be curvilinear. The qMEP will follow its curvature. Points next to the qMEP are shown in Fig. 1. Gradients in these points have a characteristic pattern. If the point x_{l0} is to the 'left' of the qMEP, then the negative gradient of x_{l1} points a step to the right. Vice versa, if the point x_{r0} is displaced to the 'right' then the negative gradient of x_{r1} points to the left. Thus, these negative gradients may be used as corrector steps. We propose the following procedure.

Gradient-only algorithm

(i) Choose step length q and parameter $\alpha < 1$ with $\alpha \ll q$, and set $i := 0$.

(ii) Choose the initial point x_0 with $|\nabla E(x_0)| \neq 0$.

(iii) Predictor step: set $x_{i+1} := x_i + q\,\text{grad}(x_i)$.

(iv) If $|\nabla E(x_{i+1})|$ is small enough then STOP (SP region is reached).

(v) Calculate scal $:= \text{grad}(x_{i+1}) \cdot \text{grad}(x_i)$. Backwards control: If (scal $\geq 1 - \alpha$) set $i := i + 1$ and Goto (iii), else

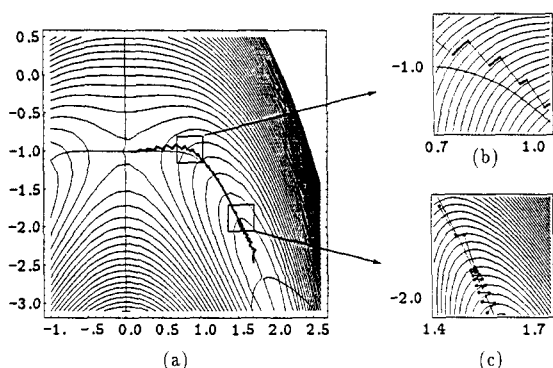(vi) Corrector step: $x_{i+1} := x_{i+1} - q\,\text{grad}(x_{i+1})$ and Goto (iii) with $i := i + 1$.

Fig. 2. (a) Test potential Eq. (2), from Ref. [9], demonstrating the use of the saddle point search algorithm for $q = 0.1$, $\alpha = 0.001$, number of steps: 130. The steps are given by black dots connected by a line. (With the refined corrector (vi') we need 28 steps.) The thin lines are equipotential lines and gradient extremal lines. Right: enlargements of difficult regions of Fig. 2a: (b) region of the curvilinear transition of the qMEP to the SP plateau, and (c) region of the transition from the curvilinear to the more straight region of the deep MEP valley.

The algorithm needs two parameters: step length q and tolerance α. The algorithm works well if α is two or more powers of ten smaller than q. The cut-off test (iv) may also be combined with q, for example $| \nabla E | \leqslant q$.

3. Tests and refinement of the algorithm

Results are given for a test potential of fourth order [9]

$$E(x, y) = 2y + y^2 + (y + 0.4x^2)x^2. \qquad (2)$$

In Fig. 2 we show the case for $q = 0.1$ and $\alpha = 0.001$ giving an approximation of the qMEP by 130 steps when starting at point $(1.77, -2.5)$ and stopping the search with $| \nabla E(x)| < 0.1$ in step (iv) assuming that the SP region is reached. The thin lines which cross the level lines are the gradient extremals defined elsewhere [9,16,17]. They depict valley ground lines, cirques and ridges of the PES. They are given for comparison. An increase of α is tested up to 0.014 in order to approximate the qMEP. Smaller values of α give a better approximation. A larger step length q may cause an ineffective increase of short corrector zigzag steps; for example, $q = 0.25$, $\alpha = 0.001$ increases the number of steps to 600. With $q = 0.1$, $\alpha = 0.015$ the algorithm breaks

down. It does not always jump into the corrector loop and then goes up by erratic predictor steps. If q is sufficiently small then it is the ratio of q and α that decides whether the algorithm works or not.

The example of Eq. (2) is chosen to illustrate the general task of a search procedure uphill: The algorithm has to work in three quite different situations. First, in a deep and curved valley ground, second, in a moderately deep and straight valley, and third on a flat SP region where again the qMEP is curvilinear. In Fig. 2c, we enlarge a conglomeration of corrector steps around the gradient extremal line. The algorithm zigzags only by corrector steps which always shorten themselves uphill along the true path, up to a point where the tolerance is satisfied. After that happens, the algorithm does a pure gradient step uphill – the predictor step. The new point is also near the true qMEP, but it still requires one corrector step. The last step in Fig. 2c is again the gradient step uphill, nearly parallel to the true pathway.

To avoid such a conglomeration of corrector steps, we refine the algorithm. In point (vi) we may use a shortened version of the corrector step by the scalar product scal. This factor scal makes the algorithm more efficient and prevents some of the unnecessary steps. If the valley is deep, we could further reduce the corrector steps, for example, by a constant factor of 0.9. However, this does not work in the SP region. Fig. 2b shows the enlargement of the SP region: we also find short corrector steps. This necessitates a further refinement of the corrector (vi): if it were too small, we may enlarge the corrector step from x_i to direction x_{i+1}. We have used the following refined corrector step with scal from step (v), and $2.5 < f < 5$:

(vi') Corrector step: $x_{i+1} := x_{i+1} - q$ scal grad(x_{i+1});

if distance $(x_i, x_{i+1}) < q/10$ and if scal $> 1 - 10\alpha$ set $x_{i+1} := x_i - f(x_i - x_{i+1})$;
goto (iii) with $i := i + 1$.

By this refinement, we can reduce the number of steps tracing the qMEP on test potential Eq. (2) from 130 to 28 steps using $q = 0.2$ and $\alpha = 0.002$.

4. Further results

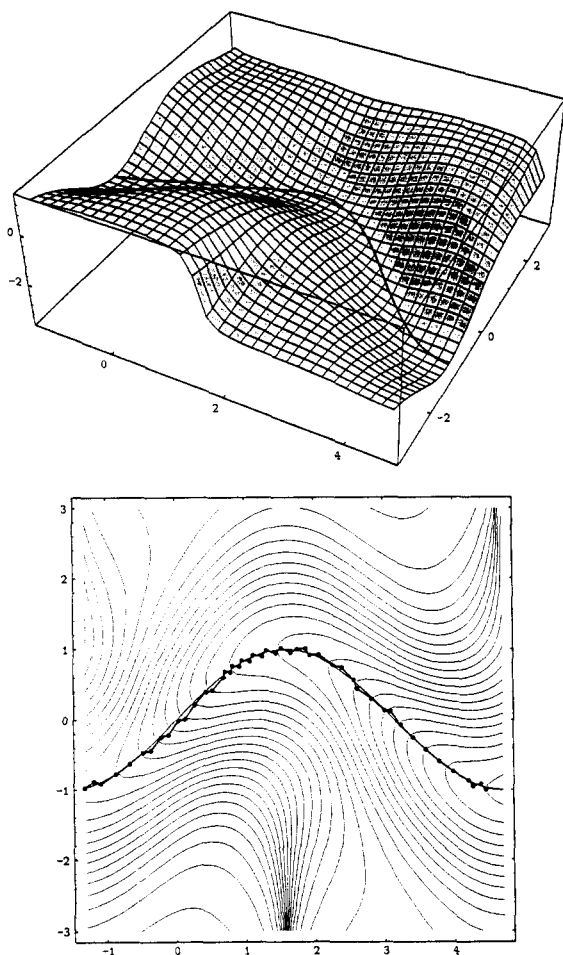We report tests with the usage of 'optimal' parameters, i.e. the algorithm finds the SP with a

Fig. 3. Test potential Eq. (3), from Ref. [22], below the approximated qMEP with a minimal number of 50 steps. These are given by black dots connected by a line. The MEP is demonstrated by a bold line calculated 'point by point' with a small q.

ber of steps because we are always on a deep lying curvilinear valley floor path. This test potential is of particular interest because it has side valleys, or more exactly, it has beginning side valleys at an altitude a little above the MEP, cf. Ref. [21]. Thus, there are side valleys, but the MEP does not bifurcate. Normally, if we use sufficiently poor parameters in the algorithm, we leave the valley floor pathway somewhere and reach a ridge. Here, for $q = 0.5$, $\alpha = 0.01$, and a corrector step shortened by the product of q and scal and 0.9, we may jump uphill accidentally to the northern side valley characterized by $x = \pi/2$ and y positive.

Remark: in the case of valley bifurcation (see test PESs in Refs. [6,9,15,16]) the proposed gradient-only algorithm, i.e. a first-order method, cannot find this peculiarity of the PES, cf. Refs. [9,23]. The calculation of a qMEP goes uphill accidentally to one of the valleys depending on the initial point of the search. However, if the search meets the bifurcation point itself, the path following can also change to the ridge between the two valleys. A procedure for the determination of bifurcation points is given in Ref. [19] using the theory of gradient extremals. A difference between qMEP and gradient extremals can be visualized in a further example. We treat the 2D-PES section of the Morse potential of the HCN molecule in a fit of Carter, Mills and Handy [24], see Fig. 4.
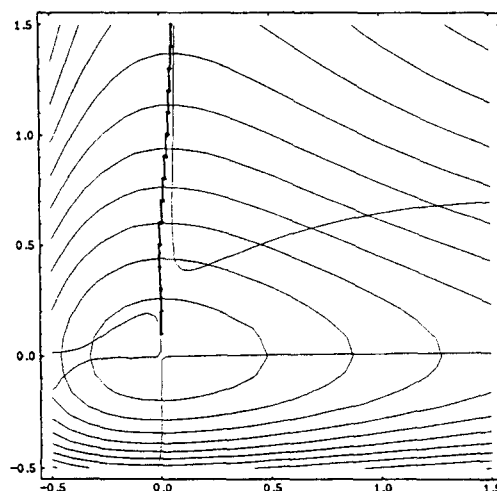
minimal number of steps using these parameters regardless of a crude qMEP approximation. In order to demonstrate the algorithm, we treat the standard 'streambed' test potential of Gonzales and Schlegel [22], see also Ref. [21]

$$E(x, y) = \text{arccot}[-e^y \cot(x/2 - \pi/4)]$$
$$- 2 \exp[-(y - \sin x)^2/2]. \quad (3)$$

We start at $(4.7, -1.0)$ using $q = 0.25$ and $\alpha = 0.05$. A resulting approximation of the qMEP with a number of 50 steps uphill is shown in Fig. 3. In this case, the refined corrector (vi') does not reduce the num-

Fig. 4. Fitted PES of HCN stretching modes [24] with gradient extremal lines (thin) and the calculated qMEP along the CN stretching valley. It is $q = 0.1$ and $\alpha = 0.0001$.

The x axis is the CH change (in Å) of the geometric distance coordinate of the bond, the y axis is the change of the CN distance. The geometry of HCN is linear in the equilibrium structure, and also for stretchings. The bending dimension is suppresed for illustration. The gradient extremals (the thin lines which cross the equipotential lines) create an unsatisfactory gap: some valley ground gradient extremals are separated by turning points which hide the valley structure of the PES. The qMEP nicely bridges this gap and goes up to a SP of the potential.

If in step (iv) the cut-off for reaching the SP region is too strong in comparison to step length q, then the search normally does not stop. The path turns 90° around the SP and goes up further along one of the ridges. A corner on the computed pathway may indicate such a bypassed SP. A possibility to test the index of the arrived SP is a downwards IRC calculation.

The chosen step length q is always a compromise between the convergence of the corrector and the number of steps to reach the SP region. In the neighborhood of the SP, a reduction of q may be used refining the SP search, see the next example. Alternatively, the exact determination of the SP may be done by the more effective second-order Newton–Raphson method. Chekmarev [11] has shown that a final stage of the SP search by Newton–Raphson typically saved about 75% of the steps of a gradient method and provided approximately the same gain of computer time.

Finally, we treat the 12D task of a cluster of four argon atoms using the Lennard-Jones potential

$$E = 4\epsilon \sum_{i<j} \left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \qquad (4)$$

for pairwise interatomic interactions [25]. ϵ is the pair potential well depth, $\sigma = 3.4$ Å is the separation appropriate for argon where the pair interaction goes through zero. r_{ij} is the separation of atoms i and j. The energy unit is $\epsilon = 1.671 \times 10^{-14}$ erg [26]. The minimum structure is a tetrahedron at -6.0ϵ, the SP1 of index 1 is a plain diamond with side length 3.80680 Å and with the two diagonals of 3.81638 and 6.58802 Å. The short diagonal also is an edge of the cluster. The SP region is extremely flat. The eigenvalue of the decomposition mode (out of plane)
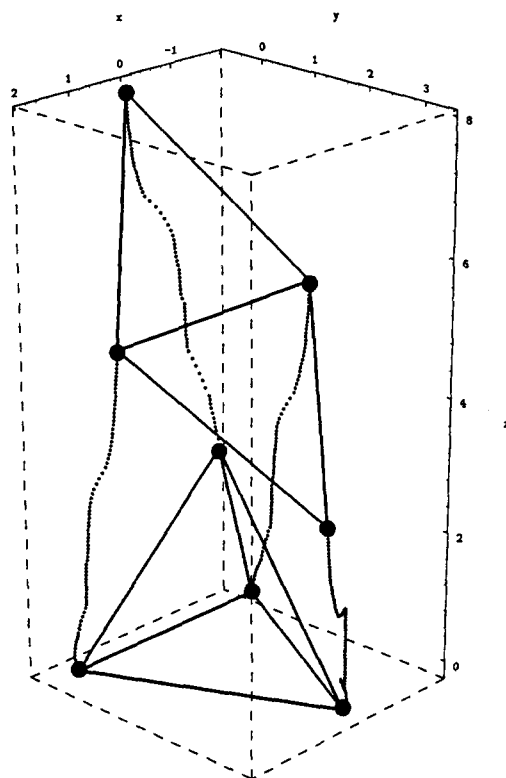

Fig. 5. Visualization of a saddle point search on a cluster of four argon atoms using a Lennard-Jones PES (configuration space with $n = 12$). The change of the cluster geometry is shown in the 3D Cartesian space. The tetrahedron below is the minimum structure, and the diamond above is the SP of index 1. The calculated valley bottom of the PES is depicted by dots, where every 25th point of the calculation is shown. For better illustration, the structures are shifted by a translation in the z direction, in each step.

is only -0.04. Thus, this potential of Eq. (4) is an extreme case in order to test the proposed gradient-only method. In the plane of the diamond, the diamond may change into a square of side length 3.78292 Å. The square is on SP2 of index 2. The energy difference between the minimum tetrahedron and the SP1 is 0.92658ϵ, and the difference between SP1 and SP2 is 0.59280ϵ. The lower structure in Fig. 5 is the tetrahedron. Shown by dots is the weakest ascent path arriving at an approximation of the rhombus in the upper part of Fig. 5. For better illustration, we shifted step by step the computed path by a small distance in the z direction. We used a three-stage approximation of the SP1 with increasingly strong search parameters given in Table 1.

Table 1
Parameters used in the SP search of an argon 4-cluster, and number of steps of the calculation

$|\nabla E|$	q	α	Steps
0.025	0.005	0.0005	1800
0.007	0.001	0.0001	4500
0.001	0.0001	0.00001	10880

Fig. 5 shows the result of the first stage, i.e. the first line in Table 1. The second stage still increases the exactness of the plain structure of the diamond by ≈ 0.2 Å, however, the third stage changes the diamond only by the thickness of the lines used. This demonstrates the well-known slowness of gradient procedures for an exact determination of extremizers of the PES [27].

5. Conclusion

This work demonstrates that the algorithm may work for an MEP following uphill on an n-dimensional potential energy hypersurface by means of gradient-only calculations of the PES functional. We define a new qMEP by comparison of gradients about a fixed step length q. The procedure is a local working predictor-corrector algorithm using 'controlled' zigzagging of gradient steps uphill along the qMEP. It does not need initial information of the SP region, or of the product region of the PES. The algorithm is simple. It works in any coordinate system, in Cartesian or internal coordinates. All degrees of freedom are automatically taken into account. The algorithm will not be disturbed by zero eigenvalues because it does not need a matrix inversion.

In program packages of theoretical chemistry, gradient calculations are standard. Thus, the proposed algorithm can be used to search a saddle point region if the SP is connected with the initial minimum by a valley at all.

The algorithm needs a high number of steps to find the exact localization of the SP in comparison with methods of second order. Thus, as all other gradient methods do, the proposed method suffers from slow convergence to the stationary point [27]. It

is well-known that this may be greatly improved by incorporating a method of higher order [11,28].

References

[1] R.A. Marcus, J. Chem. Phys. 49 (1968) 2610, 2617.
[2] K. Fukui, J. Phys. Chem. 74 (1970) 4161.
[3] D.G. Truhlar, A.J. Kupperman, J. Am. Chem. Soc. 93 (1971) 1840.
[4] A. Tachibana, K. Fukui, Theoret. Chim. Acta 49 (1978) 321.
[5] W. Quapp and D. Heidrich, Theoret. Chim. Acta 66 (1984) 245.
[6] D. Heidrich, W. Kliesch and W. Quapp, Lecture notes in chemistry, vol. 56. Properties of chemically interesting potential energy surfaces (Springer, Berlin, 1991).
[7] J.-Q. Sun and K. Ruedenberg, J. Chem. Phys. 99 (1994) 5257.
[8] M.I. Ban, Gy. Dömötör and L.L. Stacho, J. Mol. Struct. THEOCHEM 311 (1994) 29.
[9] W. Quapp, O. Imig and D. Heidrich, in: The reaction path in chemistry: current approaches and perspectives, ed. D. Heidrich (Kluwer Academic Press, Dordrecht, 1995), p. 137.
[10] K. Fukui, in: The world of quantum chemistry, eds. R. Daudel and B. Pullman (Reidel, Dordrecht, 1974) p. 113.
[11] S.F. Chekmarev, Chem. Phys. Letters 227 (1994) 354.
[12] I.V. Ionova and E.A. Carter, J. Chem. Phys. 103 (1995) 5437.
[13] S. Pancîř, Coll. Czech. Chem. Comm. 40 (1975) 1112.
[14] M.V. Basilevski and A.G. Shamov, Chem. Phys. 60 (1981) 347.
[15] C.J. Cerjan and W.H. Miller, J. Chem. Phys. 75 (1981) 2800.
[16] D.K. Hoffman, R.S. Nord and K. Ruedenberg, Theoret. Chim. Acta 69 (1986) 265.
[17] W. Quapp, Theoret. Chim. Acta 75 (1989) 447.
[18] J.-Q. Sun and K. Ruedenberg, J. Chem. Phys. 98 (1993) 9707.
[19] O. Imig, D. Heidrich and W. Quapp, submitted to J. Chem. Phys.
[20] B.L. Garret, R. Steckler, D.G. Truhlar, K.K. Baldridge, D. Bartol, M.V. Schmidt and M.S. Gordon, J. Phys. Chem. 92 (1988) 1476.
[21] K. Ruedenberg and J.-Q. Sun, J. Chem. Phys. 100 (1994) 5836.
[22] C. Gonzales and H.B. Schlegel, J. Chem. Phys. 95 (1991) 5853.

[23] W. Quapp, J. Chem. Soc. Faraday Trans. 90 (1994) 1607.
[24] S. Carter, I.M. Mills and N.C. Handy, J. Chem. Phys. 99 (1993) 4379.
[25] J.E. Lennard-Jones, Proc. Roy. Soc. A 106 (1924) 463.
[26] D.J. Wales and R.S. Berry, J. Chem. Phys. 92 (1990) 4283.

[27] H. Schwetlick, Numerische Lösungen nichtlinearer Gleichungen (Deutscher Verl. Wiss., Berlin, 1979).
[28] C.Y. Peng, P.Y. Ayala, H.B. Schlegel and M.J. Frisch, J. Comput. Chem. 17 (1996) 49.