Theor Chim Acta (1991) 80: 29–50

Theoretica Chimica Acta

A mathematical model of the logical structure of chemistry. A bridge between theoretical and experimental chemistry and a general tool for computer-assisted molecular design

I. An abstract model

Jaroslav Koča

Department of Organic Chemistry, Faculty of Science, Masaryk University, 611 37 Brno, Czechoslovakia

Received March 19, 1990; received in revised form October 30, 1990/Accepted December 17, 1990

Summary. A mathematical model of the logical structure of chemistry is suggested. The model is based on the phenomenon of convertibility between chemical species which is expressed by the so-called convertibility function Γ. In the center of the model there is the potential energy (hyper)surface, PES. A heuristic modification of the general convertibility function is presented. Several algorithms have been developed for an analysis of PES which is described by paths, and for heuristic obtaining of PES paths. The notion of K-barrier of conformational PES is introduced as well as an algorithm for its computation.

Key words: Mathematical model – Logical structure of chemistry – Potential energy hypersurface – Graph theory

1. Introduction

The power of computers is increased by the quality of mathematical models on which the computer programs are based. One could say that we are very close to the time when a very general and largely applicable chemical program system will be created. Large software packages like CHEMLAB, CHEM-X, MacroModel [1], which are very powerful for experimental chemists, can serve as first examples.

There are several standpoints for the classification of models. Some of them are discussed in the next section. For our purposes, it is reasonable to classify models as "physical" and "non-physical" (or heuristic). Physical models express exactly the physical basis of the process investigated. However, an experimental chemist working a long period of time in a certain field gains experience, and creates his special "internal" model which is usually not exact, but which serves well. Such a model could be called heuristic or logical. The developing of computer science in such fields as expert systems and logical programming are also very important stimulating factors for the creation of heuristic models of chemical reality.

However, the formulation of a mathematical model of logical structure of certain disciplines is very important for the formulation of the main problems of the discipline. The development of models of the logical structure of chemistry was pioneered by Dugundji and Ugi [2] and followed by further authors [3].

This paper is trying to generalize approaches currently used in this field. Potential energy (hyper)surface, PES, an adequate tool for modelling of interconversions between chemical species serves as a basis for the model formulation. The PES modelling consists of two fundamental steps in our approach. The first one expects exact knowledge of PES, and describes it using a graph theoretical model, a graph of PES. The second step starts with an initial set of minima along PES, and creates a similar graph theoretical model as in the first step. But, in this case, the graph of PES is created by a heuristic mechanism.

In the center of our approach there is the so-called general convertibility function Γ. The exact definition of Γ is based on PES. Since the approach presented offers to solve also the opposite task, i.e., a heuristic method of obtaining PES, a heuristic convertibility function and a special algorithm, DAISY, have been developed. However, PES is a mathematical structure which has to be chemically interpreted. Some algorithms utilizable for such an interpretation are described.

2. A static and dynamical part of the model and model classification

From the general point of view, it is difficult to classify models into static and dynamical because they are modelling molecules which are being continuously changed. We will use this classification in order to keep a clearer view. A static description is based on a "fixed" nuclear configuration and can be used for inductive or deductive prediction of features of chemical systems with this "fixed" nuclear geometry, i.e., for example, prediction of energy and optical properties. A dynamical model or dynamical part of a model serves the description of changes of nuclear configurations, i.e., the modelling of conformational, configurational, and reaction resolutions. This part of the model can then be used for the prediction of reaction rates, conditions, yields, etc.

Models of the two above categories are described in the literature. The Schrödinger wave equation is one of the most essential models. In order to make this model more practically usable, some other models have been created. The Born–Oppenheimer approximation [4] can serve as an example. The Schrödinger equation is an exact physical model. For the discussion of physical quantum chemical models see Ref. [5]. These models could be called "microscopic". More close to a "macroscopic" standpoint are heuristic (or logical) models. The Dugundji–Ugi model [2] is the first general model of this category. This model has been further developed and modified [6–10]. Heuristic models do not explicitly describe the physical basis of the chemical system investigated, that is, more or less, included implicitly. However, some models are not exactly physical or heuristic. The common feature for all models "of chemistry" is the fact that they model chemical isomerization. Most of them have been reviewed in Slanina's book about isomerism [11].

The dimension is another classification key. One-, two-, or three-dimensional models are the most frequent. QSAR studies based only on log P values can serve as an example of one-dimensional model. Two parameters as well as topological models can be classified as two-dimensional. Topological models express atoms and their neighborhood without spatial orientation which is, more or less, included implicitly. This class of models can serve the static as well as the dynamical description of a chemical system. A topological model can be visualized by a graph theoretical model and, therefore, graph theoretical models are very often called topological. Note that using a topological or a graph theoretical model, a multidimensional structure can be modelled. Such models of PES can serve as examples [12, 13]. Three-dimensional models study a chemical system as a three-dimensional body. Stereochemistry is an example of utilizing such models, and group theory an example of a suitable mathematical tool for them. As mentioned above, PES is a typical example of a multidimensional model. Of course, PES cannot be, in most cases, fully visualized. Fortunately, basic information can be "extracted" from PES, and represented by models of lower dimension.

3. Potential energy hypersurface – a bridge between physical and heuristic models of chemical reality

PES obtaining is one of the most serious goals of theoretical and computational chemistry. At the same time, although PES are multidimensional bodies, they can be efficiently interpreted by experimental chemists. Therefore, following Mezey [12a], one can say that PES is one of the basic bridges between theoretical and experimental chemistry. PES itself is a model which may be subjected to further modelling. We define a graph theoretical model of PES. Graph theory has been used for modelling PES recently, cf. for example Refs. [13, 14]. The first part of our model, presented in Sect. 3, is based on Mezey's theory of catchment regions and network relations along PES [12b,c]. We assume that PES is an energy function $E(r)$ where r is a fixed nuclear configuration of the investigated chemical structure. Let us suppose that $E(r)$ is twice differentiable.

3.1. A graph theoretical model of PES(A) *– the graph* $G_{\mathrm{PES}(A)}$

Let the (hyper)surface be constructed over the fixed set A of atoms, card $A = N$. We will use two equivalent symbols, PES and PES(A). The latter will usually be used in the case when a connection to set A appears. Let us assume that such a part of PES is studied which does not contain degenerate critical points [12a]. We will distinguish two principally different paths [12a].

(i) The path of a nuclear geometry change which is the path in $3N-6$ dimensional nuclear configuration space R^{3N-6}.

(ii) The corresponding relief path along PES, i.e. in R^{3N-5} dimensional space.

Path (i) will inform us about the progress of nuclear configuration change, and from path (ii) we get some information about energy changes.

In our model, PES will be represented by critical points, i.e., stationary points with zero gradient (all the first derivatives are zeros). We say [12a] that a critical point r is a critical point of index k if the Hessian matrix of the second derivatives has k negative eigenvalues. Specially, critical points of index zero and one are minima (chemical species) and transition states, respectively. In order to keep consistency of the model for pathological and from reality-distant cases, we will consider points with infinite coordinates as critical points. Their classification is based on limiting values of curvature. Let ${}^kV$ denote a set of critical points of PES(A) of indices $0, 1, \ldots, k$. Let ${}^kV = \{v_i\}$, $i = 1, \ldots, n$. Let V_t be the set of all critical points of index t. We may write:

$$^{k}V = \bigcup_{t=0}^{k} V_t. \tag{1}$$

Let us define set E_v for each $v \in V_t$, $t = 1, 2, 3, \ldots, k$, such that E_v is the set of all couples $\{v, v'\}$ such that some steepest descent path along PES(A) starting in v vanishes in v', $v' \in V_{t-1}$. Now, let each couple $\{v, v'\}$ be ordered in an arbitrary but fixed way. So, we get $E_v = \{(v, v') \text{ or } (v', v)\}$. Let us define the set ^{k}E as:

$$^{k}E = \bigcup_{t=1}^{k} E_t. \tag{2}$$

The graph $^{k}G_{\mathrm{PES}(A)}$ is defined as an ordered set:

$$^{k}G_{\mathrm{PES}(A)} = (^{k}V, {}^{k}E, \phi, \{\mathrm{Y}, \mathrm{R}\}) \tag{3}$$

where the set of vertices ^{k}V and edges ^{k}E is defined by Eqs. (1) and (2), respectively, ϕ is mapping $^{k}E \rightarrow \{\mathrm{Y}, \mathrm{R}\}$ defined as:

$$\phi(v, v') = \begin{cases} \mathrm{Y} & \text{if } E(v) \leqslant E(v'), \\ \mathrm{R} & \text{if } E(v) > E(v'). \end{cases}$$

The graph $^{k}G_{\mathrm{PES}(A)}$ is a graph with oriented edges colored yellow (Y) or red (R). The edge (v, v') is oriented by $v \rightarrow v'$. It is clear from the above construction that $^{k}G_{\mathrm{PES}(A)} \subset {}^{k+1}G_{\mathrm{PES}(A)}$. The construction of $^{k}G_{\mathrm{PES}(A)}$ is visualized in Fig. 1.

Sets V and E from Eqs. (1) and (2) are as follows: $^{0}V = \{v_1, v_1', v_1'', v_1'''\}$, $^{1}V = {}^{0}V \cup \{v_2, v_2', v_2'', v_2'''\}$, $^{2}V = {}^{1}V \cup \{v_3, v_3', v_3'', v_3'''\}$, $^{0}E = \varnothing$, $^{1}E = \{(v_2', v_1''), (v_2''', v_3'''), (v_1', v_2'), (v_1', v_2''), (v_1', v_2), (v_2, v_1), (v_2'', v_1'''), (v_1, v_2'''), (v_1''', v_2''')\}$, $^{2}E = {}^{1}E \cup \{(v_3', v_2'''), (v_2, v_3'), (v_2'', v_3'), (v_3'', v_2''), (v_3, v_2), (v_3'', v_2'), (v_3, v_2')\}$. Note that edge orientation in graphs $^{1}G_{\mathrm{PES}(A)}$ and $^{2}G_{\mathrm{PES}(A)}$ is arbitrary. The reason will be seen in Sect. 4.

However, mainly the graphs $^{0}G_{\mathrm{PES}(A)}$ and $^{1}G_{\mathrm{PES}}(A)$ will be of our interest. The graph $^{0}G_{\mathrm{PES}(A)}$ contains isolated vertices, minima representing chemical species. Interconversions between minima can be seen from the graph $^{1}G_{\mathrm{PES}(A)}$.

It should be pointed out that the graph $^{k}G_{\mathrm{PES}(A)}$ is very similar to Mezey's reaction network digraph D [12b]. However, they differ in some points. The essential difference is that only relations between critical points whose indices differ by unit are considered in $^{k}G_{\mathrm{PES}(A)}$. From this point of view, $^{k}G_{\mathrm{PES}(A)}$ could be understood as a subgraph of D. The same difference appears comparing the lower introduced general convertibility function to the relative and absolute E-reachability concept [12b].

3.2. *The general convertibility function*

Following the general idea of the intrinsic reaction coordinate of Fukui [15] we say that a chemical species E is convertible to another chemical species P if there exists a saddle point I along PES(A), and two steepest descent paths joining I such that the first of these vanishes in E and the second one in P. The interconversion path $E \rightarrow P$ is the composition of the first and second path, and point I plays the role of the transition state. We will generalize the above idea

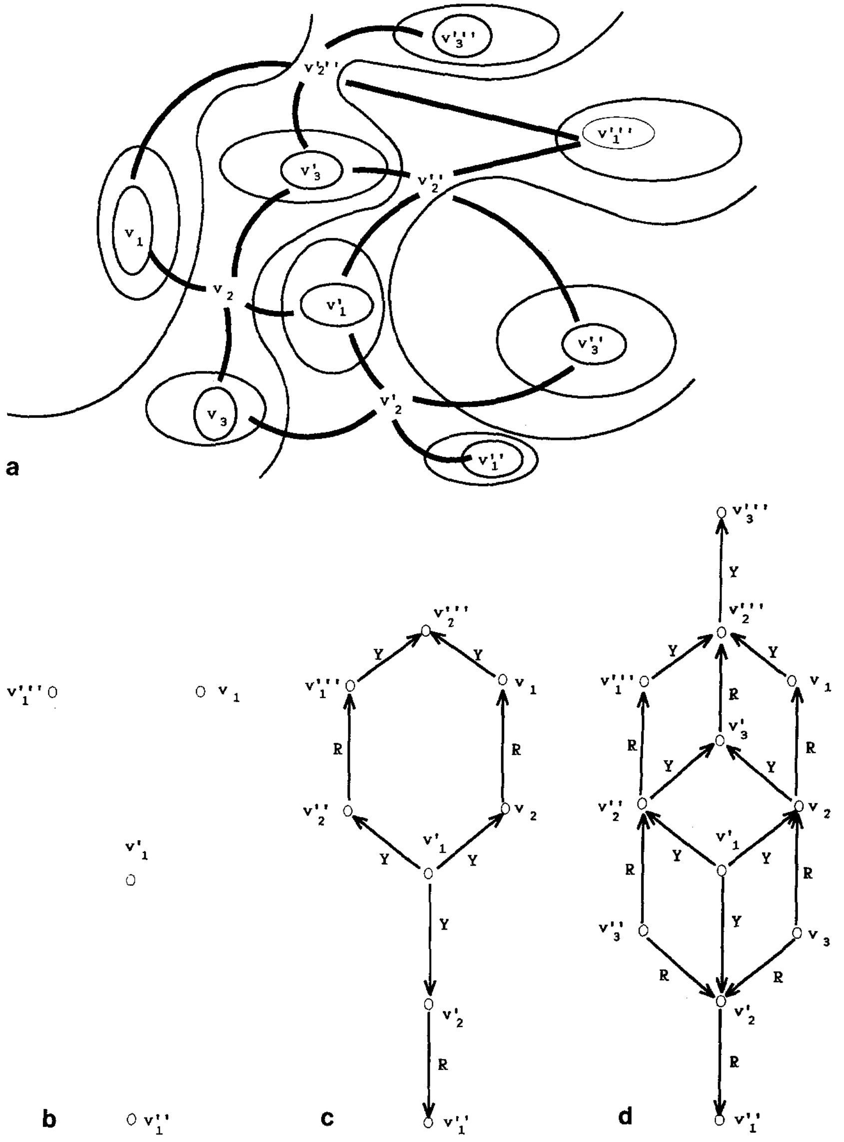

Fig. 1. **a** Part of model surface with critical points v_i, v'_i, v''_i, v'''_i of index $i-1$, and steepest descent paths shown by heavy arcs. **b** Graph ${}^0G_{\mathrm{PES}(A)}$ of the above surface. **c** Graph ${}^1G_{\mathrm{PES}(A)}$ of the surface. **d** Graph ${}^2G_{\mathrm{PES}(A)}$ of the surface

also for critical points of higher indices, and define the general convertibility function Γ for arbitrary $v, v' \in {}^kG_{\mathrm{PES}(A)}$ as follows:

$$\Gamma(v, v') = \begin{cases} 1 & \text{if there exist } q \text{ and } x \text{ such that } v, v' \in V_q,\ x \in V_{q+1}, \\ & (v, x) \in {}^kE \text{ or } (x, v) \in {}^kE \text{ and } (v', x) \in {}^kE \text{ or } (x, v') \in {}^kE, \\ 2 & \text{otherwise.} \end{cases} \tag{4}$$

The values of Γ for the elements of V_0 are of special interest because they inform about the convertibility between chemical species.

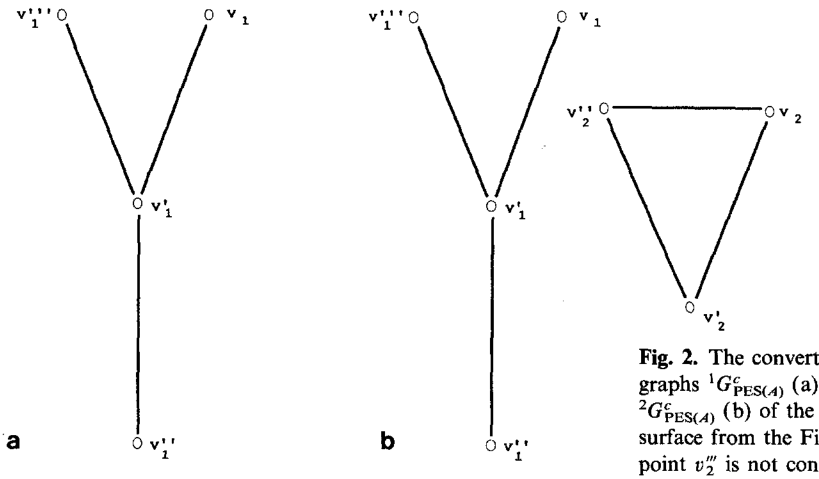

Fig. 2. The convertibility graphs ${}^{1}G^{c}_{\mathrm{PES}(A)}$ (a) and ${}^{2}G^{c}_{\mathrm{PES}(A)}$ (b) of the model surface from the Fig. 1 (the point v_2''' is not considered)

3.3. The convertibility graph of PES(*A*)

The convertibility along PES can be visualized by the convertibility graph ${}^{k}G^{c}_{\mathrm{PES}(A)}$ defined as follows:

$$ {}^{k}G^{c}_{\mathrm{PES}(A)} = ({}^{k-1}V, {}^{k}E^{c}) \tag{5} $$

where ${}^{k-1}V$ is the vertex set of the graph ${}^{k-1}G_{\mathrm{PES}(A)}$ (cf. Eq. (3)), ${}^{k}E^{c} = \{\{v, v'\} \mid v, v' \in {}^{k-1}V \text{ and } \Gamma(v, v') = 1\}$, and the function Γ is defined by Eq. (4).

The edges of ${}^{k}G^{c}_{\mathrm{PES}(A)}$ model the individual interconversions between the nuclear configurations which are symbolized by vertices of graph ${}^{k-1}G^{c}_{\mathrm{PES}(A)}$. The convertibility graphs ${}^{1}G^{c}_{\mathrm{PES}(A)}$ and ${}^{2}G^{c}_{\mathrm{PES}(A)}$ of the part of the model surface from Fig. 1 are shown in Fig. 2. The edge sets E^{c} from Eq. (5) are as follows:

$$ {}^{1}E^{c} = \{\{v_1, v_1'\}, \{v_1', v_1''\}, \{v_1', v_1'''\}\}, \qquad {}^{2}E^{c} = {}^{1}E^{c} \cup \{\{v_2, v_2'\}, \{v_2, v_2''\}, \{v_2', v_2''\}\}. $$

Note that ${}^{k-1}G^{c}_{\mathrm{PES}(A)} \subset {}^{k}G^{c}_{\mathrm{PES}(A)}$.

3.3.1. Energy K-barriers of the graph of conformational PES

Studying the reaction PES, we are usually interested in some domains of PES. There is a different situation if a conformational PES is of our interest. Here, we would like to know all low-energy minima conformations. The heights of the energy barriers for the individual interconversions can be understood as a criterion of experimental evidence, and they are in a close relation to the conformational softness [16] and flexibility of a molecule. The so-called *K*-barrier is a quantitative expression of the flexibility of the molecule.

Let us suppose that the graph ${}^{k}G^{c}_{\mathrm{PES}(A)}$ of a conformational PES is connected, and each interconversion, i.e., the edge of the graph, is described by interconversion barrier which is expressed with respect to the absolute minimum along PES, i.e., with respect to the lowest energy conformation. We define the energy *K*-barrier as the lowest energy barrier over which the graph will be split into *K* components. In other words, if we remove all the edges with the energy value higher than *K*-barrier from the graph ${}^{k}G^{c}_{\mathrm{PES}(A)}$, we get the graph with *K* components. Removing all edges with the highest energy value from the remaining subgraph will split the subgraph at least into $K+1$ components. It is clear from the above definition that the energy K-barrier is always greater than or equal to zero, and it is equal to zero only if ${}^{k}G^{c}_{\mathrm{PES}(A)}$ is composed of isolate vertices, i.e., for $K = \operatorname{card}\, {}^{k-1}V$. The following algorithm has been developed for the computation of the energy K-barriers.

Algorithm 1

0. Initialization: Let ${}^{k}G^{c}_{\mathrm{PES}(A)} = ({}^{k-1}V, {}^{k}E^{c})$ be the convertibility graph of a conformational PES. Let this graph be connected. Let $\psi : {}^{k}E^{c} \to \mathbf{R}$ be a function which describes the conversion barriers for each conversion with respect to the lowest energy conformation. Let ${}^{k-1}V \neq \emptyset, {}^{k}E^{c} \neq \emptyset$.
1. $K := 1$.
2. Let $e \in {}^{k}E^{c}$ be such an edge for which the inequality $\psi(e) \geqslant \psi(e')$ is satisfied for each $e' \in {}^{k}E^{c} - \{e\}$. Put $KB := \psi(e)$; ${}^{k}E^{c} := {}^{k}E^{c} - \{e\}$.
3. If ${}^{k}G^{c}_{\mathrm{PES}(A)}$ has K components then go to step 2.
4. Put energy K-barrier $:= KB$: $K := K + 1$.
5. If ${}^{k}E^{c} \neq \emptyset$ go to step 2.
6. The end of the algorithm.

The algorithm is illustrated in Fig. 3.

Algorithm 1 is implemented as a part of the computer program system DAISY [16] for the analysis of conformational PES's.

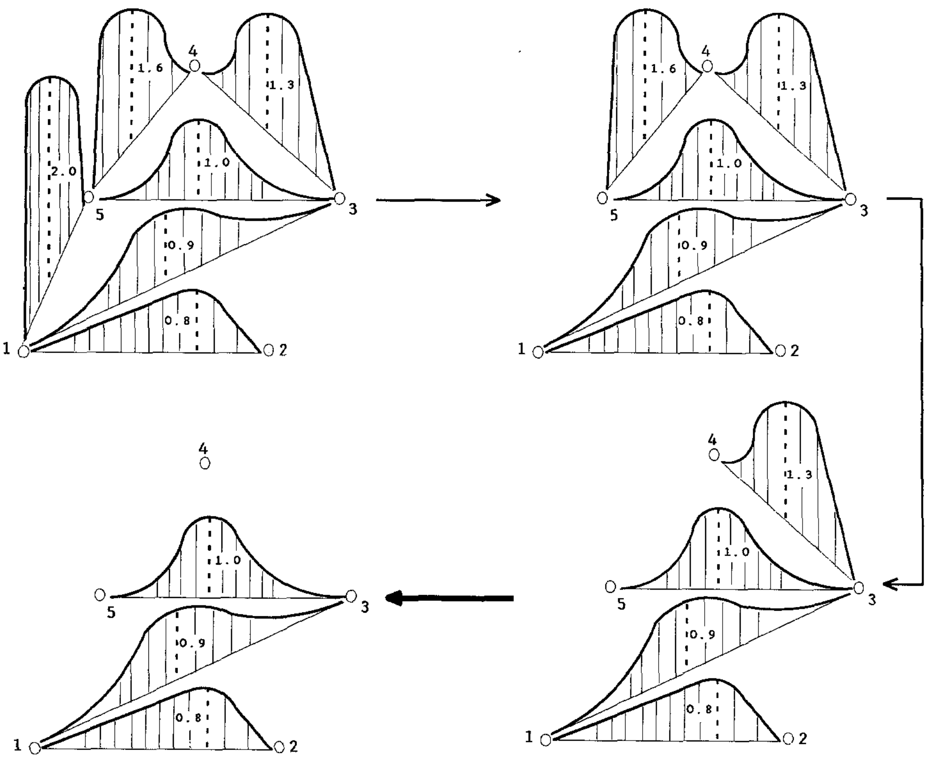

Fig. 3. The application of the algorithm 1 to a graph with 5 conformations and energy barriers in kcal/mol. The third step, shown by heavy arrow, splits the graph into two components (conformation 4 becomes isolated). It means that energy 1-barrier is 1.3 kcal/mol. Analogically, for 2, 3, and 4-barrier we have 1.0, 0.9, and 0.8 kcal/mol, respectively

3.4. The extended graph of PES(*A*)

The graph theoretical model presented above is based on the ideal knowledge of PES(A), i.e., all of its critical points. Let us also extend the model for other points. Consider the set ${}^kV'$ of points of PES(A) such that ${}^kV' \supset {}^kV$, where ${}^kV$ is defined by Eq. (1). Let us define the set E_v for each $v \in {}^kV' - {}^kV$ as the set of all couples $\{v, v'\}$, where the steepest descent path starting in v vanishes in v' ($v' \in {}^kV$), or joins $v' \in {}^kV' - {}^kV$. Let us orient each of the couples $\{v, v'\}$ in an arbitrary but fixed way. So, we get $E_v = \{(v, v')$ or $(v', v)\}$. Now, let us define the set ${}^kE'$ as follows:

$$ {}^kE' = {}^kE \cup \bigcup_{v \in {}^kV' - {}^kV} E_v \tag{6} $$

where ${}^kE$ is defined by Eq. (2).

The extended graph ${}^kG'_{\mathrm{PES}(A)}$ is defined as an ordered set:

$$ {}^kG'_{\mathrm{PES}(A)} = ({}^kV', {}^kE', \phi, \{\mathrm{Y}, \mathrm{R}\}) \tag{7} $$

where ${}^kV', {}^kE'$ are sets defined above and ϕ is mapping defined analogically to that in Eq. (3).

The part of model surface, and corresponding graphs ${}^kG'_{\mathrm{PES}(A)}$ are shown in the Fig. 4. The sets V' and E' from Eqs. (6) and (7) are as follows:

$$ {}^0V' = \{v_1, v_1', v_1'', v_1''', v^{(1)}, v^{(2)}, v^{(4)}, v^{(5)}, v^{(6)}, v^{(7)}\}, $$

$$ {}^1V' = {}^0V' \cup \{v_2, v_2', v_2'', v_2''', v^{(3)}\}, $$

$$ {}^2V' = {}^1V' \cup \{v_3, v_3', v_3'', v_3'''\}, $$

$$ {}^0E' = \{(v^{(2)}, v_1'), (v^{(1)}, v_1), (v^{(5)}, v_1''), (v^{(7)}, v^{(5)}), (v^{(6)}, v^{(5)}), (v_1''', v^{(4)})\}, $$

$$ \begin{aligned} {}^1E' = {}^0E' \cup \{&(v_2', v_1''), (v_1', v_2'), (v_1', v_2''), (v_1', v_2), (v_2, v_1), (v_2'', v_1'''), \\ &(v_1, v_2'''), (v_1''', v_2'''), (v_2, v^{(3)})\}, \end{aligned} $$

$$ \begin{aligned} {}^2E' = {}^1E' \cup \{&(v_3', v_2'''), (v_2, v_3'), (v_2'', v_3'), (v_3'', v_2''), (v_3, v_2), \\ &(v_3'', v_2'), (v_3, v_2'), (v_3''', v_2''')\}. \end{aligned} $$

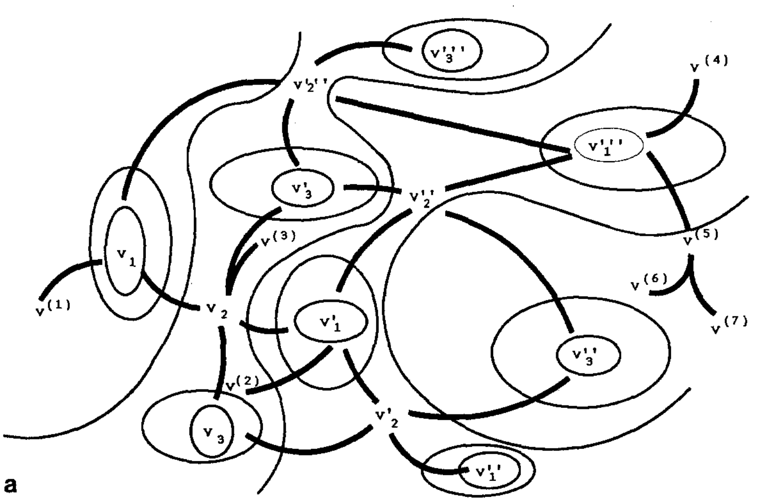

Fig. 4. a

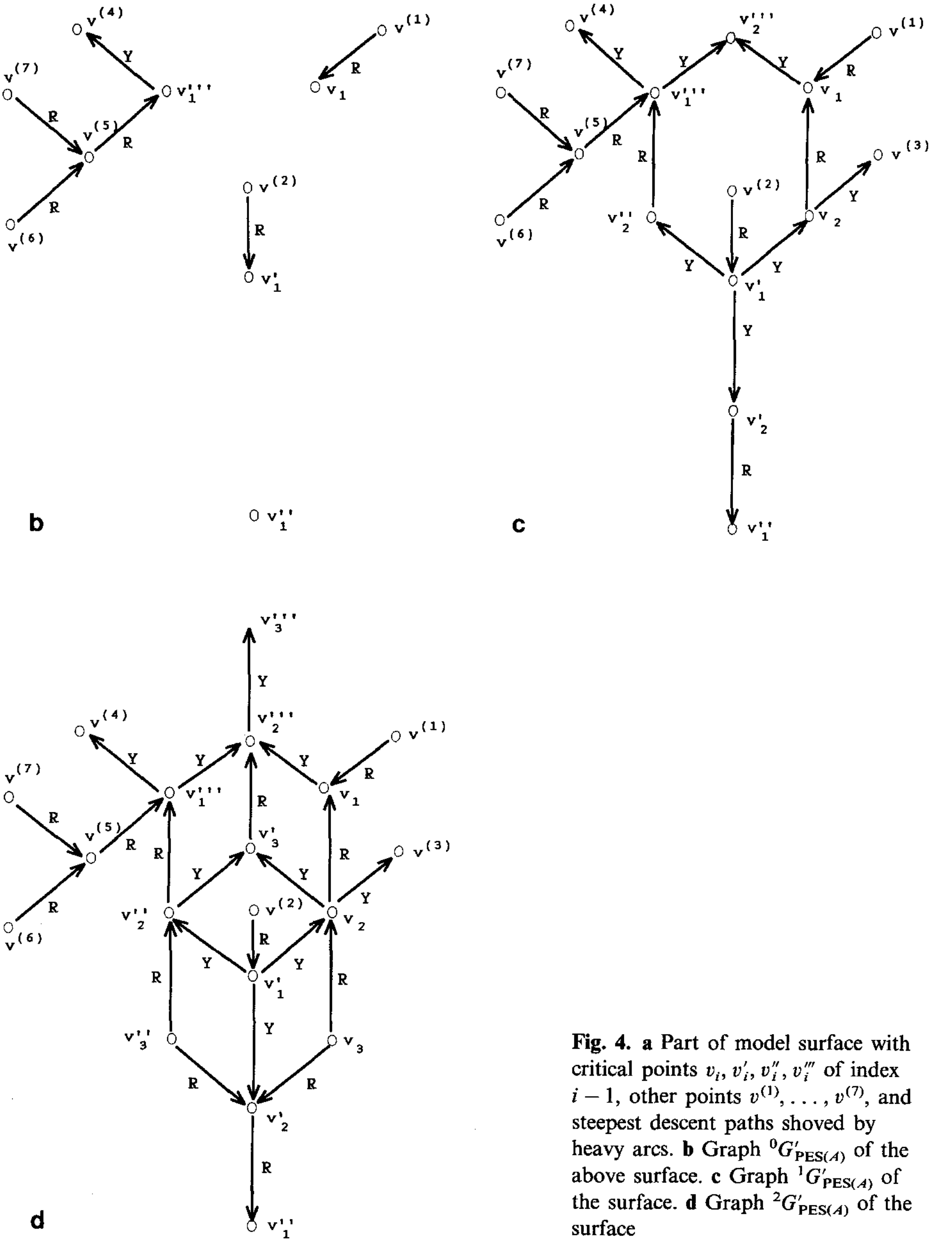

Fig. 4. **a** Part of model surface with critical points v_i, v'_i, v''_i, v'''_i of index $i-1$, other points $v^{(1)}, \ldots, v^{(7)}$, and steepest descent paths shoved by heavy arcs. **b** Graph ${}^{0}G'_{\mathrm{PES}(A)}$ of the above surface. **c** Graph ${}^{1}G'_{\mathrm{PES}(A)}$ of the surface. **d** Graph ${}^{2}G'_{\mathrm{PES}(A)}$ of the surface

3.4.1. *Catchment regions and the graph* ${}^{k}G'_{\mathrm{PES}(A)}$

In order to better understand relations between chemical species, we use the notion of a catchment region. The notion has been, in connection to PES, initially introduced by Mezey [12, 17]. Following him we say that a general point v of PES, which is not a critical point, is in the catchment region of a critical point v' of PES(A) if the steepest descent path joining v vanishes in v'. Chemical species v' can be understood as a set of all nuclear configurations which are in the catchment region of v'.

3.4.2. A classification of the vertices of the graph ${}^kG'_{\mathrm{PES}(A)}$. *An algorithm for the classification of points on* PES

This section is directed to an analysis of PES which is described by paths. It presents some analogies with Mezey's approach [12c] which is oriented to a topological study of boundary properties of catchment regions.

It has been assumed in the above paragraphs that full information about critical points of PES is available. Let us turn our attention to such a case, where the PES is described by certain points and paths between them. This situation is probable when an analytical description of PES is not accessible. In an ideal case, we have a reliable estimation of the graph ${}^kG'_{\mathrm{PES}(A)}$ which contains a good approximation of all critical points. Since exact information about the index of a critical point cannot be obtained without exact knowledge of this critical point, we may assume that almost no information about critical points is available. We introduce the following algorithm 2 for the classification of points along PES which starts with the graph ${}^kG'_{\mathrm{PES}(A)}$ (or better to say, with its reliable estimation), and is based on the set construction. Note that under the term "critical point" is, in this passage, understood rather an approximation of the critical point than formulation of its geometry exactly. The same should be said about paths. The following operations are required from the algorithm:

1. Separation of critical and remaining points of PES.

2. Separation of non-critical points into catchments regions of the critical points.

3. Finding the index of each critical point.

Note that steps 3 and 4 of the algorithm below are implicitly included in step 5. They are introduced separately in order to keep a clearer view.

Algorithm 2

0. Initialization: Let the extended graph ${}^kG'_{\mathrm{PES}(A)}$, defined by Eq. (7), be input structure. Let ${}^kV' := \{v_1, \ldots, v_n\}$, $I := \{1, 2, \ldots, n\}$.
1. Let us define the set K_i for each $v_i \in {}^kV'$ as follows: $K_i = \{v_i\} \cup \{x_0 \in {}^kV' \mid$ there exists a series $x_0, x_1, \ldots, x_m, v_i$ such, that $E(x_t) > E(v_i)$ for $t = 0, 1, \ldots, m$; $E(x_q) > E(x_{q+1})$ for $q = 0, 1, \ldots, m-1$, and $(x_m, v_i) \in {}^kE'$ or $(v_i, x_m) \in {}^kE'$, $(x_q, x_{q+1}) \in {}^kE'$ or $(x_{q+1}, x_q) \in {}^kE'\}$. In other words, K_i is the set of all points of PES from which it is possible to find such a path joining v_i which is composed of the steepest descent paths.
2. Let us choose a new set system K_{i0} from the sets K_i, $i \in I$, in such a way that K_{i0} are not in inclusion. Formally: $i0 \in I_0$, where I_0 is the set of such indices $i0, j0$ that $i0 \neq j0$ implies $K_{i0} \not\subset K_{j0}$ for each $i0$. It is clear that $I_0 \subset I$.
3. Let us construct the sets C_{i0} and I^0 such that

$$C_{i0} := K_{i0} - \bigcup_{\substack{j0 \in I_0 \\ j0 \neq i0}} K_{j0}, \qquad I^0 := \{i \mid v_i \in C_{i0},\ i0 \in I_0\}.$$

4. Let us construct the sets C_{i1} and I^1 such that

$$C_{i1} := K_{i1} - \bigcup_{\substack{j1 \in I_1 \\ j1 \neq i1}} K_{j1}, \qquad I^1 := \{i \mid v_i \in C_{i1}, i1 \in I_1\},$$

where $I_1 \subset I - I^0$ such that $K_{i1} \not\subset K_{j1}$ for each $i1, j1 \in I_1$.

5. Let us construct the sets C_{im} and I^m such that

$$C_{im} := K_{im} - \bigcup_{\substack{jm \in I_m \\ jm \neq im}} K_{jm}, \qquad I^m := \{i \mid v_i \in C_{im}, im \in I_m\},$$

where $I_m \subset I - \bigcup_{j=0}^{m-1} I^j$ such that $K_{im} \not\subset K_{jm}$ for each $im, jm \in I_m$.

6. The end of the algorithm.

It follows from step 5 of the above algorithm that $I_q \subset I^q$, for $q = 0, 1, \ldots$. Since the set ${}^kV'$ is finite, the set I has to be also finite. It implies that there exists such an index m that the set I_m is empty. In the case of the extended graph ${}^kG'_{\mathrm{PES}(A)}$, one can write that $k < m < s + 1$ where s is the dimension of PES(A). The results of algorithm 2 are summarized by the following theorem.

Theorem 1

Let us consider the denotation from algorithm 2. Then

1. The point v_{iq}, $iq \in I_q$, $q = 1, 2, \ldots$, is a critical point of PES(A) of the index q.

2. The point v_{it} is from the catchment region of the point v_{iq} for each $v_{it} \in C_{iq} - \{v_{iq}\}$ and $iq \in I_q$, $q = 1, 2, \ldots$.

Example 1. Let us consider the graph ${}^1G'_{\mathrm{PES}(A)}$ from Fig. 4 with the following numbering of vertices: $1{:}v_1, 2{:}v'_1, 3{:}v''_1, 4{:}v'''_1, 5{:}v_2, 6{:}v'_2, 7{:}v''_2, 8{:}v'''_2, 9{:}v^{(1)}$, $10{:}v^{(2)}$, $11{:}v^{(3)}$, $12{:}v^{(4)}$, $13{:}v^{(5)}$, $14{:}v^{(6)}$, $15{:}v^{(7)}$. The sets K, C, and I from algorithm 2 are constructed as follows:

Step 0: $I = \{1, 2, \ldots, 15\}$.

Step 1: $K_1 = \{v_1, v_2, v'''_2, v^{(1)}, v^{(3)}\} = \{1, 5, 8, 9, 11\}$,
$K_2 = \{v'_1, v_2, v'_2, v''_2, v^{(2)}, v^{(3)}\} = \{2, 5, 6, 7, 10, 11\}$,
$K_3 = \{v''_1, v'_2\} = \{3, 6\}$,
$K_4 = \{v'''_1, v''_2, v'''_2, v^{(4)}, v^{(5)}, v^{(6)}, v^{(7)}\} = \{4, 7, 8, 12, 13, 14, 15\}$,
$K_5 = \{v_2, v^{(3)}\} = \{5, 11\}$, $K_6 = \{v'_2\} = \{6\}$,
$K_7 = \{v''_2\} = \{7\}$, $K_8 = \{v'''_2\} = \{8\}$,
$K_9 = \{v^{(1)}\} = \{9\}$, $K_{10} = \{v^{(2)}\} = \{10\}$,
$K_{11} = \{v^{(3)}\} = \{11\}$, $K_{12} = \{v^{(4)}\} = \{12\}$,
$K_{13} = \{v^{(5)}\} = \{13\}$, $K_{14} = \{v^{(6)}\} = \{14\}$,
$K_{15} = \{v^{(7)}\} = \{15\}$.

Step 2: For instance, the following set inclusions can be found between the sets K_i: $K_5 \subset K_1$, $K_6 \subset K_2$, $K_7 \subset K_4$, $K_8 \subset K_1$, $K_9 \subset K_1$, $K_{10} \subset K_2$, $K_{11} \subset K_2$, $K_{12} \subset K_4$, $K_{13} \subset K_4$, $K_{14} \subset K_4$, $K_{15} \subset K_4$. So, we have: $I_0 = \{1, 2, 3, 4\}$.

Step 3: The sets C_{i0} and the set I^0 are as follows:

$$C_1 = \{v_1, v^{(1)}\}, \qquad C_2 = \{v'_1, v^{(2)}\},$$
$$C_3 = \{v''_1\}.$$
$$C_4 = \{v'''_1, v^{(4)}, v^{(5)}, v^{(6)}, v^{(7)}\},$$
$$I^0 = \{1, 2, 3, 4, 9, 10, 12, 13, 14, 15\}.$$

Step 4: The sets C_{i1}, I_1 and I^1 are constructed in the following manner.

$$C_5 = \{v_2, v^{(3)}\}, \qquad C_6 = \{v'_2\},$$
$$C_7 = \{v''_2\}, \qquad C_8 = \{v'''_2\},$$
$$I_1 = \{5, 6, 7, 8,\}, \qquad I^1 = \{5, 6, 7, 8, 11\}.$$

Step 5: The set I_2 must be a subset of the set $I - \bigcup_{j=0}^{1} I^j$, which is empty set. It implies that I_2 is empty and the algorithm has been terminated. The set I_0 and I_1 is the set of critical points of the index 0 and 1, respectively. The sets $C_1, \ldots, C_8$ contain points from the catchment regions of critical points $1, \ldots, 8$.

Proof of Theorem 1

By induction with respect to q. Let us suppose that $q = 0$.

a1. Let I_0 be defined by step 2 of algorithm 2, $i0 \in I_0$. It follows from step 1 of the algorithm that v_{i0} is the minimal element (in the sense of the energy along steepest descent path) in the set K_{i0}. It follows from step 2 that there is no path which is joining v_{i0}, and which would reach another point of PES(A) with lower energy than v_{i0}. However, if such a path exists then such a set K_{j0} has to exist that $K_{i0} \subset K_{j0}$ which is in contradiction to step 2. It means that v_{i0} is a local minimum along PES(A), a critical point of zero index.

a2. Let us consider a general point $v_i \in C_{i0} - \{v_{i0}\}$. This point cannot be a critical point of zero index because there exists a path composed of the steepest descent parts which connects this point to v_{i0}. This point v_i cannot be a critical point of higher index either. Let us suppose that this point is a critical point of a higher index. It implies that two different paths composed of the steepest descent parts have to exist which connect v_i to two different points v_{i0}, v_{j0}. Step 2 implies that $v_i \in C_{i0}$ and $v_i \in C_{j0}$ which is in contradiction to the definition of the set C_{i0} in step 3. Accordingly, v_i is not a critical point of PES(A). Since $v_i \in C_{i0}$ and, consequently, $v_i \in K_{i0}$, there exists a path connecting v_i and v_{i0}, which is composed of the steepest descent parts. We will show that there does not exist another critical point on this path. Let us suppose that v_j ($v_j \neq v_i$) is a critical point on the path $v_i \to v_{i0}$. The point v_j cannot be of index zero because v_{i0} is of index zero. It implies that there exists a critical point v_{j0} ($v_{i0} \neq v_{j0}$) of index zero such that v_j is connected to v_{j0} by a path composed of steepest descent parts. From this, we have $v_j \in K_{i0} \wedge v_j \in K_{j0}$. Consequently, $v_j \notin C_{i0}$, which is the contradiction, and v_j is an element of the catchment region of v_{i0}.

Let us suppose that the assertion is valid for each $q = 0, 1, \ldots, t$. We will show it for $q = t + 1$.

b1. Let $iq \in I_q$. The point v_{iq} cannot be in any catchment region of any critical point of index less than q because it would be in a set C_{it} ($t < q$). This is not possible because of the construction of the set I_q in step 5. Because of that, and since v_{iq} is the minimal element in the set K_{iq} (and C_{iq}), it has to be a critical point of index q.

b2. Let $v_i \in C_{iq} - \{v_{iq}\}$. In the same way as in a2, we may show that v_i is an element of the catchment region of v_{iq}.

In order to prove Theorem 1, we had to assume, in fact, that we have a very good approximation (or exact knowledge) of critical points. The meaning of the words "very good" here should be, however, subjected to further analysis.

Now, we have got an algorithm which can be used for the classification of critical points of PES(A), and the separation of non-critical points into catchment regions of critical points. Based on the algorithm, we enlarge general convertibility function Γ for the extended graph ${}^kG'_{\mathrm{PES}(A)}$.

3.4.3. General convertibility function on the graph ${}^kG'_{\mathrm{PES}(A)}$

Let us suppose the denotation introduced in the above algorithm 2. We will define the function Γ for two arbitrary points of the set ${}^KV'$ as follows:

$$\Gamma(v, v') = \begin{cases} 0 & \text{if there exists } m \text{ such that } v_i, v_j \in C_{im}, \\ 1 & \text{if there exists } m \text{ such that } v_i \in C_{im},\ v_j \in C_{jm}, \\ & im, jm \in I_m, \text{ and there exists } 1 \in I_{m+1} \text{ such that} \\ & v_1 \in K_{im},\ v_1 \in K_{jm}, \\ 2 & \text{in the remaining cases.} \end{cases} \tag{8}$$

The sets C and I are defined by step 5 of algorithm 2.

Note that the function Γ is an extension of the function from definition 4 for points from the same catchment region.

4. A heuristic description and obtaining of PES

In the above section, we have suggested the algorithm which makes it possible to analyze a PES which is defined by paths. Now, let us turn our attention to the problem of obtaining the paths. The proposed approach is heuristic. A part of that, oriented to conformational PES analysis, has been realized on computer [16]. The approach will be discussed within a graph-theoretical frame.

4.1. The starting set of nuclear configurations – the graph ${}^0G''_{\mathrm{PES}(A)}$

The starting point for obtaining PES paths is the formulation of the starting set A of atoms, and the starting set of nuclear configurations expected as chemical species. Following Eqs. (1, 3, 6), we will denote this set as ${}^0V''$. The graph ${}^0G''_{\mathrm{PES}(A)}$, which is called the heuristic graph of PES(A), is defined as follows:

$${}^0G''_{\mathrm{PES}(A)} = ({}^0V'', {}^0E'', \phi, \{\mathrm{Y}, \mathrm{R}\}) \tag{9}$$

Similarly to the graphs ${}^0G_{\mathrm{PES}(A)}$ and ${}^0G'_{\mathrm{PES}(A)}$, the set ${}^0E''$ is empty, ϕ is defined analogically to that in Eqs. (3) and (7). The part of PES(A) being of interest is determined by the set ${}^0V''$. However, this set is different studying conformational, configurational or reaction PES. If one is interested in conformational or configurational problems then one would like to obtain information about the entire PES(A). Usually only small domains of PES will be of interest studying chemical reaction. Accordingly, the method of gaining the set ${}^0V''$ will be different for each of the above problems.

In conformational analysis, the generation of the starting set is identical with the search for the most stable conformations of the molecule, i.e., search for the lowest energy conformations. Various methods have appeared in this field [18], for example grid search method combined with various heuristic and random approach [19], molecular dynamics as well as expert systems [20].

One of the possible ways how to generate the configurational starting set ${}^0V''$ has been introduced by Ugi et al. [21]. This general approach is based on the group-theory formalism.

The formulation of the set ${}^0V''$ for chemical reaction study may be performed "by hand". It is, for instance, in such a case when we have reactants and products, and we are interested in reaction mechanism, activation barrier, stereochemistry of the reaction, etc. In the simplest cases, the set ${}^0V''$ has two elements. However, from the general (and heuristic) point of view, the above problem is combinatorial as it has been recently discussed in connection with the so-called FIEM(A) (Family of Isomeric Ensembles of Molecules) [2], and FIS(A) (Family of Isomeric Synthons) [8, 9]. However, the sets FIEM(A) and FIS(A) are usually very large, and only parts of them can be used as a starting set ${}^0V''$ for the elucidation of reaction PES.

4.2. Interconversions between elements of ${}^0V''$ – a heuristic convertibility function Γ'

One of the main reasons why to obtain PES is to get a dynamical picture about the investigated system, i.e., the illustration of interconversions between elements of the set ${}^0V''$. The interconversions have been modelled by the convertibility function Γ in the previous section. Now, we introduce a heuristic convertibility function Γ' which is defined intuitively, but similarly as the exact convertibility function Γ. Accordingly, the function Γ' can be understood as an estimation of Γ. It is defined for each v, v' from ${}^0V''$ as follows:

$$\Gamma'(v, v') = \begin{cases} 0 & \text{if both } v \text{ and } v' \text{ are expected to be in the} \\ & \text{catchment region of a critical point } p \text{ of PES}(A), \\ 1 & \text{if } v \text{ is expected to be convertible to } v' \\ & \text{(in the sense of } \Gamma\text{),} \\ 2 & \text{otherwise.} \end{cases} \tag{10}$$

Note that the function Γ' is also defined for such a case when v and v' are in the same catchment region. It is a formally worthless situation for ${}^0V''$. There are two reasons for such a definition. The former is that the set ${}^0V''$ will be extended to ${}^1V'', {}^2V'', \ldots,$ during the path generating process which is described below and, as will be seen in the section describing the algorithm DAISY, the function Γ' must also be applicable for these new sets ${}^1V'', {}^2V'', \ldots,$ of points on PES. The latter is that some points from the starting set ${}^0V''$ may subsequently be recognized as shallow minima which should be called rather unstable intermediates or transition states than chemical species. In other words, some points looking like minima can later be identified in another category. The main aim why Γ' is defined is the need for forming a starting set of conversions which will be used for generating PES paths. The function Γ' is defined by Eq. (10) in a general manner. It is substituted by an actual convertibility function in real situations.

4.2.1. *The stability of the heuristic convertibility function Γ'*

In order to evaluate how the predicted conversions are real, we introduce the notion of stability of Γ' as follows. Let ${}^{k}H'_{\mathrm{PES}(A)} = ({}^{k}V'_{H}, {}^{k}E'_{H}, \phi, \{\mathrm{Y}, \mathrm{R}\})$ be a subgraph of the extended graph of PES(A), ${}^{k}G'_{\mathrm{PES}(A)}$. Let C_{Γ} and $C_{\Gamma'}$ be the set of conversions produced by the function Γ and Γ' on the graph ${}^{k}H'_{\mathrm{PES}(A)}$, respectively. We say that the function Γ' is on the set ${}^{k}V'_{H}$ of vertices:

— strongly stable, if $C_{\Gamma} = C_{\Gamma'}$ (11a)

— stable, if $C_{\Gamma'} \subset C_{\Gamma}$ (11b)

— weakly stable, if $C_{\Gamma} \subset C_{\Gamma'}$ (11c)

— unstable, otherwise. (11d)

Finding a strongly stable convertibility function in some domain of PES(A) would mean that one can predict interconversions in the frame of this domain without any knowledge of PES(A). However, it is an ideal case. Note that stability of Γ' is closely related to the classification of computer programs for synthesis design [9]. From this point of view, informational oriented approaches are based on a stable convertibility function (each proposal can usually be realized). In spite of that, logically oriented programs are based on a weakly stable convertibility function because they can usually find all the solutions of the problem (for example, combinatorially), but it is very difficult to interpret results because of their largeness.

It should be pointed out that the notion of the stability of the heuristic convertibility function Γ' is of a different meaning than that of the stability of minimum energy reaction paths [22]. The former expresses a quality of the heuristic convertibility function Γ' which can, however, predict very different conversions comparing to those derived from PES. The latter is an exact expression, and requires exact knowledge of the domain of PES which is being studied.

4.3. *Verification of heuristic predicted conversions*

The graph ${}^{0}G''_{\mathrm{PES}(A)}$ is a static expression of the investigated chemical system. It includes only minima along PES(A). The used general convertibility function Γ' expresses a heuristic prediction of a dynamical behavior of the system. The main task for Γ' is to reduce the multi-dimensional problem of PES computation. However, the computation of Γ' has to be much faster than the application of some "expensive" and more exact method like the quantum chemical one. This "expensive" method is applied at the second level of the whole approach when the conversions predicted by Γ' are verified. The algorithm DAISY introduced below has been developed for this reason.

4.3.1. *The algorithm* DAISY

DAISY, an acronym for Distance Approach In potential energy (hyper)Surface paths Yield, is a heuristic algorithm for obtaining PES paths. The algorithm requires a starting set of nuclear configurations and a set of predicted interconversions between them. The algorithm is heuristic in such a sense that the paths found really exist, but they need not be optimal. The quality of results is dependent mainly on the general driver used which will be discussed later. The algorithm introduced below is a modification of the recently developed version, which is a part of the program system DAISY [16].

Algorithm 3 – DAISY

Describing the algorithm DAISY we will use the following denotation. Let A be the set of conversions predicted by Γ', i.e., A is the set of couples $\{(v_i, v_j)\}$ such that the conversion $v_i \rightarrow v_j$ has been predicted by Γ'. Let each couple (v_i, v_j) be ordered in such a way that $E(v_i) \leqslant E(v_j)$ (where E is the energy function), i.e., the conversions will be verified starting from the lower energy structure and leading to the higher energy structure. This rule can be changed using one option in the program system DAISY. Let B be the set of produced paths along PES, i.e. B is the set of couples $\{(v, v')\}$ such that the path $v \rightarrow v'$ has been generated by the algorithm DAISY. The function Γ' is applied, if it is not explicitly said differently, within the actual set C_s. The algorithm is described as follows.

0. Initialization of variables and functions for the DRV functor, i.e., selection of functions Γ' and Θ and setting up the step in the variable S, and the constant $E_{\min}$. Put $B := \emptyset$, $s := 0$, $C_s := {}^0V$, $F(v_i, v_j) := 0$ for each (v_i, v_j) from C_s.
1. If A is empty go to step 16, else take the last couple from the set A. Let it be a couple (v_i, v_j), $s := s + 1$, $C_s := C_{s-1}$. Let $t := F(v_i, v_j)$, i.e., t is the index of the set C_s during which the edge $v_i \rightarrow v_j$ has been created.
2. Put $M := 0$, $v := v_i$, $v' := v_i$.
3. Put $M := M + 1$, $v := v'$, $H :=$ FALSE.
4. Put $v' := \mathrm{DRV}(v_i, v', v_j, P, S, \Gamma', \Theta, U, C_t \cup \{v_i\})$, i.e., perform one step of driving from v_i to v_j starting in v'.
5. If $M = 1$ and $P = 1$ put $A := A - (v_i, v_j)$, i.e., delete the predicted conversion from the set A of conversions to be verified, because this conversion is now being verified.
6. If $\Theta(v_i, v') = 0$ (v_i and v' are found to be very similar) go to step 3.
7. If $\Gamma'(v', v_j) = 0$ and $\Theta(v', v_j) \neq 0$ (i.e., v' is in the catchment region of v_j but v' and v_j are not identical) go to step 3.
8. Let D be a set of nuclear configurations v_k from C_s such that $\Theta(v', v_k) = 0$. Put $D := D - \{v_i, v\}$.
9. If $D \neq \emptyset$ then do 9a, 9b and 9c for each v_k from D. Then go to step 1.

9a. If $[E(v') - E(v_i)][E(v_k) - E(v')] < 0$ then set $v_k := v'$.

9b. Set $B := B \cup \{(v_i, v_k)\}$.

9c. If $\Gamma'(v_k, v_j) = 1$ in the set $C_t \cup \{v_k\}$ then put $B := B \cup \{(v_k, v_j)\}$, $F(v_k, v_j) := t$. If $D = \emptyset$ go to step 11.
The condition 9a means that looking for a minimum, we take the lower energy structure, and looking for a maximum, we take the higher energy structure.

10. If $U =$ true, i.e., if DRV cannot drive v' between v_i, v_j any more then put $C_s := C_s \cup \{v\}$, $B := B \cup \{(v_i, v)\}$, and go to step 1.
11. If $[E(v') - E(v)][E(v) - E(v_i)] < 0$ (the energy function E changes monotonousness in v, i.e., v is a local maximum or minimum) set $C := C \cup \{v\} \cup \{v'\}$, $B := B \cup \{(v_i, v), (v, v')\}$, $H :=$ TRUE.
12. If v' still meets the path $v_i \rightarrow v_j$ then if $H =$ FALSE go to 3 else put $A := A \cup \{(v', v_j)\}$, $F(v', v_j) := s$, and go to step 1.
13. Let D be a set of nuclear configurations v_k such that v' is convertible onto v_k, i.e., $\Gamma'(v', v_k) = 1$. Put $D := D - \{v_i, v\}$.

14. If $D = \emptyset$ then save the point with the nuclear configuration v' into the set C of points describing PES, i.e. $C_s := C_s \cup \{v'\}$, $B := B \cup \{(v_i, v')\}$ and go back to step 1.
15. Save all conversions from the point v' onto all points from the set D into the set A of remaining conversions to be verified. Formally, $A := A \cup \{(v', v_k)\}$, $F(v', v_k) := s$, for each v_k from D. Go back to step 1.
16. The end of the algorithm.

Note that the above algorithm is logical but, at the same time, it includes some steps which make it possible to incorporate quantum chemical or other energy computation methods. However, it is not very surprising because energy computation itself needs such methods.

One of the essential parts of the algorithm is a general driver here called $\mathrm{DRV}(v_i, v', v_j, P, S, \Gamma', \Theta, U, C)$. The driver is a functor which performs a verification of the path $v_i \rightarrow v_j$ along PES. The algorithm generates the next point on the PES which should be on the path $v_i \rightarrow v_j$, and which follows the point v' with the step S. The driving is realized on the set C of various nuclear configurations. The logical variable U is set up in such a case when DRV cannot drive any more. The number of expected paths along PES in this case, i.e. number of paths between v_i and v_j, is returned by the variable P. This variable is of a special meaning if $v_i = v'$, i.e. when the starting step of the verification of the path is being realized. Of course, the general driver DRV will be of various forms studying different types of PES. From the general point of view, every approach which is able to follow the verified path (for example, a valley path driver) may be used as a general driver. The approaches normally used in computational chemistry [23] can serve as examples.

It is clear that DRV is dependent on the convertibility function Γ'. Except for this function, the function Θ, which measures an identity of nuclear configurations, is used. From the general point of view, the function Θ should express the principle of uncertainty. The function Θ_1^c, used by current versions of the program system DAISY [16], can serve as an example of a heuristic estimation of Θ. In order to eliminate the conversions with a too low energy barrier, the parameter E_{min} is used by the algorithm. The parameter is not described here. It eliminates all the conversions with energy barrier less than E_{min}.

However, only a primary description of PES, based on paths, is produced by the algorithm. This has to be subjected to further analysis for which some of the above algorithms can be used. The final set B can be understood as the set ${}^kV''$, and the final set C_s as the set ${}^kE''$ of colored and oriented edges. Therefore, the results of the algorithm may be seen as the heuristic graph ${}^kG''_{\mathrm{PES}(A)}$, an analogy of the graph ${}^0G''_{\mathrm{PES}(A)}$ defined by Eq. (9). If we apply algorithm 2 to this graph, we get a heuristic estimation of the value k, critical points of indices $0, 1, \ldots, k$ as well as the separation of "non-critical" points into catchment regions of critical points. By using the function Γ (Eq. (8)) one can get information about convertibility, conversion barriers as well as "transition states". At the same time, algorithm 1 can be applied, and K-barriers obtained.

Note that DAISY reduces a generally exponential problem of computation of PES to the polynomial one. Let us consider that the starting set is composed of n vertices. Let us take the worst case when there is a conversion between each couple of n vertices to be verified. It implies that $\binom{n}{2} = (n^2 - n)/2$ of conversions have to be verified. Each of them will be processed using a driver. Let us consider that the upper bound of driver steps number is m. The computation of energy in each step is polynomial, i.e., it may be expressed, for example, as $P(x)$, where $P(x)$ is a polynomial and x is the number of electrons. The final time of computation can then be estimated as $(m/2) \cdot (n^2 - n) \cdot P(x)$, which is a polynomial. This is very important because the largeness of computed chemical systems which may be computed in polynomial time can be much more increased by the computer speed than if the computational time is expressed exponentially.

The entire approach has been practically realized in several versions of the program system DAISY [16] which can be used for conformational PES paths obtaining with the assistance of molecular mechanics energy computation method (the program MMPMI [24] is used). At the same time, the conformational softness [16] can be computed by DAISY.

5. The generalized frame of the above model

The model can be seen from and used at various levels. One could say that the final aim is to get the graph ${}^kG'_{\mathrm{PES}(A)}$ which contains most of the information we need. However, in real situations, this graph is substituted by the heuristic graph ${}^kG''_{\mathrm{PES}(A)}$ obtained by the algorithm DAISY using various methods of energy computation and geometry optimization. Large chemically important information can be obtained from the first level of this graph, ${}^0G''_{\mathrm{PES}(A)}$, and the heuristic convertibility function Γ'. This function is a heuristic basis of the model which concentrates chemical experience. Note that this function can be substituted, for example, by some program for computer assisted synthesis design. From the general point of view, Γ' includes two levels, the problem of completeness and the practical chemical applicability. The above mentioned mathematical models [2, 6–9] can serve as examples of elucidating the former level problems. The covering of the latter calls for a general heuristic theory of chemical reactivity including solvation and catalysis. The notions of similarity and complementarity may play an important role in this field.

5.1. The similarity and complementarity of chemical systems

Let us turn our attention to two notions, chemical reactivity and activity. The former is one of the central points of entire chemistry. The latter is very important in Computer Aided Molecular Design, CAMD. Both reactivity and activity can be seen from two standpoints. The first one is similarity of chemical systems which has been studied in the literature from both two- and three-dimensional point of view (cf., for example, [25, 26]). It can be seen from many points of view, and used for reactivity predicting on the qualitative as well as quantitative level. The complementarity expresses an affinity of substrate and reagent or substrate and receptor, for example by "docking" [27] based on the theory of "lock and key". The similarity and complementarity seems to be a good formally methodological basis for the development of a heuristic theory of chemical reactivity and activity of chemical species. From the standpoint of the model, both similarity and complementarity can be understood as tools assisting the formulation of the function Γ' and, at the same time, they may serve as general tools for incorporating the model into CAMD.

5.2. Quantum chemistry and the model

The heuristic part of the model plays the role of PES predicting. Note that the predicted paths along PES have to be verified. Accordingly, the model is formulated in such a way that it can be used as a basis for the application of quantum chemical methods which can exactly elucidate small domains of PES. The model is formulated within classical mechanics. Quantum mechanics may be incorporated into the model using, for instance, quantum chemical methods energy computation. At the same time, the first, non-heuristic, part of the model is based on Mezey's theory of PES and, therefore, opened for the topological methods of PES study introduced by him, cf. for example [12]. These methods seem to be able to express the principle of uncertainty, and describe better the elucidated system from the quantum mechanical point of view.

5.3. The model and computer-assisted synthesis design

Computer-assisted synthesis design is one of the most attractive fields in theoretical chemistry. Two main trends have appeared in this area, information oriented programs, and those based on the logical structure [2]. Both above approaches are heuristic, and they do not use quantum chemistry. In theoretical chemistry, some approaches have appeared which are oriented completely non-empirically [28], and which have been called by Ugi et al. computer programs for synthesis design of third generation [29].

From the computer-assisted synthesis design point of view, the presented model can be used at two levels. The first one, the model may be used as a tool for the development of logically oriented computer programs on a heuristic basis using a heuristic convertibility function. From this point of view, the existing programs for computer-assisted synthesis design can be seen as a realized first level of this model. Very close to the philosophy of the model is the program MAPOS [30] which is based on the synthon model of organic chemistry [8, 9, 31].

The realization of the algorithm DAISY is the second level of the model. DAISY is a tool through which quantum chemistry may be incorporated. Mezey has proposed a general algorithm for non-empirical synthesis design [12] based on PES, and this level can serve, for example, as a tool for the definition of the starting set for that algorithm. Note that Mezey's algorithm may be relatively fast if the starting set is well defined, and, at the same time, very slow when the starting set proposal of critical points is not good.

We believe that this model could be a part of a "mosaic" which will serve as a basis for computer-assisted synthesis design programs of a new generation.

6. Conclusions

The main aim of this paper is to propose a mathematical model which would be able to represent the logic of chemical changes on a general level. Therefore, the model is based on convertibility between chemical species, which is expressed by the convertibility function. The model itself is composed of two basic levels. The former is exact, introducing the general convertibility function which supposes exact knowledge of PES. The latter, heuristic, starts with heuristic predicted conversions between chemical species, and generates and elucidates conversion paths along PES. From this point of view, the model is a "synthesis" of exact physical (PES) and heuristic approaches. Therefore, it can be used by theoretical as well as experimental chemistry. The heuristic elucidation of conversion paths along PES is one from the main results of the model. Since PES may be understood as a bridge between theoretical and experimental chemistry, the model could serve as a tool which may contribute to better and easier interpretation of results of theoretical chemistry by experimental chemists. We believe that the model could also serve as a basis for chemically applicable program with this aim.

Because of the included convertibility function and the general driver, the model is flexible. It produces the graph of predicted interconversions between chemical species. This graph itself may be used as a deductive model for convertibility prediction. However, it can be subjected to further analysis, and mainly a heuristic theory of chemical reactivity can be incorporated at this point. Such a theory could be incorporated in a "macroscopic" level as references about the change in a global way, or it can be concentrated into references about elementary interconversions, for example elementary electronic processes, using the theory of reaction mechanisms.

The real using of the model, or computer program system based on it, can be split into four steps.

(a) The formulation of a problem by chemist, i.e., the definition of the starting set ${}^{0}V''$. That can be, for example, conformational or configurational analysis, elucidation of reaction mechanism, the problem of synthesis of a chemical structure, etc.

(b) Solution of the problem on the heuristic level without including any exact method, using a heuristic theory of chemical reactivity. Formulation of a problem for utilizing a more exact method.

(c) Application of an exact method controlled by the heuristic mechanism, realization of the algorithm DAISY.

(d) Recomputation of interesting points of PES by an exact method.

It is easy to see that orientation of the model is determined by the formulation of the convertibility function Γ' and the general driver DRV. One can find several ideas for the formulation of a several-level heuristic convertibility function in the literature. The metric space model by Dugundji, Ugi et al. [3a], their stereochemical model based on group-theory [21], Hendrickson's model [32], the model presented by Barone and Chanon [33], the synthon model of organic chemistry [8–10] etc. can serve as examples. At the same time, the model may be interfaced to Mezey's topological model [12] of PES as well as to his theory of reaction topology [34]. However, a well usable chemical program has to include as a very important part some "interface to neighborhood", i.e., to a chemist. Such a program should be "user friendly". It means that it should "understand", for example, chemical nomenclature. Some parts of such an interface has been described, for instance, by Gordon [35].

The described model has been partially realized as the program system DAISY [16] for studying of the conformational behavior of open chain molecules. The convertibility function and actually used driver is described in ref. [36].

Acknowledgement. The author would like to thank one unknown referee for valuable comments and for pointing out important references.

References

1. CHEMLAB. Molecular Design Ltd. 1986. CHEM-X. Chemical Design Ltd. 1986. MacroModel. Columbia University 1986
2. Dugundji J, Ugi I (1973) Top Cur Chem 39:19
3. (a) Dugundji J, Gillespie P, Marquarding D, Ugi I, Ramirez F (1976) Metric spaces and graphs representing the logical structure of chemistry. In: Balaban AT (ed) Chemical applications of graph theory. Academic Press, New York, p 107
   (b) Kratochvíl M (1981) Chem Listy 75:675
   (c) Kratochvíl M (1983) Chem Listy 77:225
   (d) Plath PJ, Hass EC (1983) Logic of chemical ideas. In: Chemical applications of topology and graph theory. Elsevier, Amsterdam, p 392
4. Born M, Oppenheimer R (1927) Ann Phys 84:457
5. Tomasi J (1988) J Mol Struct (Theochem) 179:273
6. Kvasnička V (1983) Coll Czech Chem Commun 48:2097, 48:2118
   Kvasnička V (1984) Coll Czech Chem Commun 49:1090
7. Kvasnička V, Kratochvíl M, Koča J (1983) Coll Czech Chem Commun 48:2284
8. Koča J (1989) J Math Chem 3:73, 3:91
9. Koča J, Kratochvíl M, Kvasnička V, Matyska L, Pospíchal J (1989) A synthon model of organic chemistry and synthesis design. Lecture notes in chemistry, vol 51. Springer-Verlag, Berlin Heidelberg
10. Kvasnička V, Pospíchal J (1990) Int J Quant Chem 38:253
11. Slanina Z (1986) Contemporary theory of chemical isomerism. Academia and Reidel, Prague and Dordrecht
12. (a) Mezey PG (1987) Potential energy hypersurfaces. Elsevier, Amsterdam
    (b) Mezey PG (1982) Theor Chim Acta (Berl.) 60:409
    (c) Mezey PG (1984) Int J Quant Chem, Quant Chem Symp 18:675
13. Zhuravlev VI, Krivoshey IV, Sleta LA (1975) Zh Strukt Khimii 16:951; Krivoshey IV, Sleta LA (1976) Theor Chim Acta (Berl.) 43:165
14. Sevcenko SM (1983) Theoretical and Experimental Chemistry, No 6:672 (in Russian)
15. Fukui KJ (1970) J Phys Chem 74:4161
16. Koča J, Carlsen PHJ (in preparation)
17. Mezey PG (1981) Theor Chim Acta 58:309; Mezey PG (1981) Int J Quant Chem, Quant Biol Symp 8:185; Mezey PG (1982) Theor Chim Acta 62:133
18. Chang G, Guida WC, Still WC (1989) J Am Chem Soc 111:4379
19. Fusco R, Caccianotti L, Tosi C (1986) Il Nuovo Cimento 8:211
20. Dolata DP, Carter RE (1987) J Chem Inf Comp Sci 27:36
21. Ugi I, Dugundji J, Kopp R, Marquarding D (1984) Perspectives in theoretical stereochemistry. Lecture notes in chemistry, vol 36. Springer-Verlag, Berlin Heidelberg
22. Mezey PG (1980) Theor Chim Acta 54:95
23. Van de Graaf B, Baas JMA, van Veen A (1980) Recueil, J Royl Neth Chem Soc 99:175; Lugovskoy AA, Dashevsky VG, Kitaigorodsky AI (1973) Tetrahedron 29:287; Jørgensen P, Jensen HJA, Helgaker T (1988) Theor Chim Acta 73:55; Berry RS, Davis HL, Beck TL (1988) Chem Phys Let 147:13; Ischtwan J, Collins MA (1988) J Chem Phys 89:2881; Beyer A, Wolschann P (1989) J Mol Struct 196:371
24. Gilbert KE, Gajewski JJ (1987) The Program MMPMI, Indiana University
25. Johnson MA (1989) J Math Chem 3:117
26. Arteca GA, Jammal VB, Mezey PG (1988) J Comput Chem 9:608; Mezey PG (1988) J Math Chem 2:299; Arteca GA, Mezey PG (1989) J Phys Chem 93:4746; Arteca GA, Heal GA, Mezey PG (1990) Theor Chim Acta 76:377
27. Frühbeis H, Klein R, Wallmeier H (1987) Angew Chem Int Ed Engl 26:403

28. Slanina Z (1989) J Mol Struct (Theochem) 185:217
29. Ugi I, Gillespie P (1971) Angew Chem Int Ed Engl 10:914
30. Matyska L, Koča J, J Chem Inf Comp Sci (in press)
31. Koča J (1988) Coll Czech Chem Commun 53:1007
32. Hendrickson JB, Grier DL, Toczko AG (1985) J Amer Chem Soc 107:5228
33. Barone R, Chanon M (1989) MATCH 24:39
34. Mezey PG (1987) J Mol Struct (Theochem) 149:57; Mezey PG (1986) Reaction topology. In: Smith VH, Schaefer HF, Morokuma K (eds) Applied quantum chemistry, Proceedings of the Nobel Laureate Symposium on Applied Quantum Chemistry in Honour of Herzberg G, Mulliken RS, Fukui K, Lipscomb W and Hoffman R, Honolulu, Hawaii, Reidel New York
35. Gordon JE, Brockwell JC (1983) J Chem Inf Comp Sci 23:117; Gordon JE (1984) J Chem Inf Comp Sci 24:81; Gordon JE (1988) J Chem Inf Comp Sci 28:100
36. Koča J (1991) Theor Chim Acta 80