A GRAPH MODEL OF THE *SYNTHON**

Jaroslav KOČA

Department of Organic Chemistry, J. E. Purkyně University, 611 37 Brno

Received December 1st, 1987
Accepted February 14th, 1988

A graph model of the *synthon* (*S*-graph) is devised and a procedure for the construction of the minimal graphs of reactions of *synthons* (*SR*-graphs) is suggested. Chemical examples illustrating the application of this model are given.

An algebraic model of the *synthon* and its conversions has been suggested previously[1]. An important concept introduced in this connection is that of the *reaction distance RD* between two isomeric *synthons*. *RD* cannot be calculated directly as the graph distance in the $G_{FIS(A)}$ graph because of the large extent of the latter[1]. For the calculation of *RD* it is convenient to build up a graph model of the *synthon*; such a model would also be better acceptable by the chemist than the matrix model. Formulation of the graph model of the *synthon* and its conversions is the objective of this paper.

The graph model is based on Kvasnička's graph model of constitutional chemistry[2-4], supplementing it with free valences of the *synthon* to its surroundings. In the model the free valences are conceived between actual atoms of the *synthon* and unspecified, so-called virtual atoms. A static description of the *synthon* is the *synthon* graph (*S*-graph), changes in the *synthon* during the reaction are modelled by the *synthon reaction* graph (*SR*-graph).

THE *Synthon* GRAPH

The definition of the *S*-graph is an extension of that of the molecular graph[2], augmenting it with virtual vertices. The *S*-graph *G* is defined as an ordered 5-tuple

$$G = (V, E, L, \varphi, \curlyvee) , \qquad (1)$$

where *V* is the set of vertices corresponding to the atoms of the *synthon*, *E* is the set of edges corresponding to the bonds in the *synthon*, *L* is the set of loops each cor-

* Part XII in the series Mathematical Model of Organic Chemistry; Part XI: Collect. Czech. Chem. Commun. *53*, 1007 (1988).

responding to a lone valence electron, $\mathscr{V}$ is the vocabulary of chemical symbols of elements, augmented with the symbol ε (empty vertex) for a virtual vertex, and φ is $V \to \mathscr{V}$ mapping, assigning to the vertices their chemical value.

The *S*-graph of a *synthon* can be construed as the union of two components, viz. of the internal *S*-graph of the *synthon* G^I and the external *S*-graph of the *synthon* G^E, defining these as

$$G^I = (V^I, E^I, L, \varphi, \mathscr{V}) \tag{2}$$

$$G^E = (V^E, E^E, \emptyset, \varphi, \mathscr{V}), \tag{3}$$

where

$$V^I = \{v \in V \mid \varphi(v) \neq \varepsilon\}$$

$$E^I = \{e = \{x, y\} \in E \mid x \in V^I \text{ and } y \in V^I\}$$

$$E^E = E - E^I$$

$$V^E = \{v \in V \mid v \text{ is incident with } e \text{ where } e \in E^E\}.$$

In general, then, $V^I \cap V^E \neq \emptyset$.

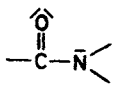
It is a straightforward consequence of the definition of G^I and G^E that $G = G^I \cup G^E = (V^I \cup V^E, E^I \cup E^E, L, \varphi, \mathscr{V})$; E and L are multisets, i.e., two different elements of these sets can formally appear as identical (e.g., two edges joining the same two vertices – multiedges).

For avoiding tedious drawing of one-electron loops, the symbol

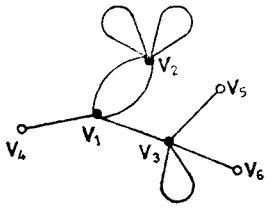
will be used for a lone electron pair, and the symbol

for a one-electron loop (unpaired electron).

Example 1. Consider the *synthon*



Denoting virtual vertices as ○, nonvirtual as ●, we can represent its *S*-graph as having $V^I = (v_1, v_2, v_3)$, $E^I = \{\{v_1, v_2\}, \{v_1, v_2\}, \{v_1, v_3\}\}$, $\varphi(v_1) = C$, $\varphi(v_2) = 0$,



$\varphi(v_3) = N$, $\mathscr{V} = \{C, O, N, \varepsilon\}$, $L = \{\{v_2, v_2\}, \{v_2, v_2\}, \{v_2, v_2\}, \{v_2, v_2\}, \{v_3, v_3\}, \{v_3, v_3\}\}$, $V^E = \{v_1, v_3, v_4, v_5, v_6\}$, $E^E = \{\{v_1, v_4\}, \{v_3, v_5\}, \{v_3, v_6\}\}$, $\varphi(v_4) = \varphi(v_5) = \varphi(v_6) = \varepsilon$. Then, $G = G^I \cup G^E = (V^I \cup V^E, E^I \cup E^E, L, \varphi, \mathscr{V})$.

The *Synthon* Reaction Graph

The process of conversion of a *synthon* $S(A)$ into another, isomeric *synthon* $S'(A)$ can be modelled by the *SR*-matrix[1] or the *synthon reaction* graph (*SR*-graph). The *synthon reaction* graph G_R is defined as the union of the internal and external reaction graphs G_R^I and G_R^E, respectively.

The Internal SR-*Graph*

The internal *SR*-graph G_R^I is defined analogously as the reaction graph in the graph model of constitutional chemistry[4], only the loops each model a single electron rather than two electrons as in ref.[4]

Be G and G' the *S*-graphs of two isomeric *synthons* $S(A)$ and $S'(A)$, respectively, where $A = (A_1, A_2, \ldots, A_n)$ is a set of atoms. In terms of Eqs (*2*) and (*3*), assume that $G = G^I \cup G^E$, $G' = G'^I \cup G'^E$. Given $G^I = (V^I, E^I, L, \varphi, \mathscr{V})$ and $G'^I = (V^I, E'^I, L', \varphi, \mathscr{V})$, we define G_R^I as

$$G_R^I = (v_R^I, E_R^I, L_R, \psi^I, \omega, \{-1, 1\}), \tag{4}$$

where $E_R^I = E^I \setminus E'^I$, $L_R = L \setminus L'$, $\psi^I: E_R \to \{-1, 1\}$, $\psi^I(e) = -1$ for $e \in E^I$ and 1 for $e \in E'^I$, $\omega: L_R \to \{-1, 1\}$, $\omega(l) = -1$ for $l \in L$ and 1 for $l \in L'$. $V_R^I \times V$ is the set of all vertices in G^I and G'^I graphs that are incident with some edge from E_R^I or with some loop from L_R. Operation $\setminus$ denotes symmetrical difference of sets, hence

$$A \setminus B = \{x \mid (x \in A \wedge x \notin B) \vee (x \notin A \wedge x \in B)\} = (A - B) \cup (B - A). \tag{5}$$

External SR-*Graphs*

Whereas internal reaction graphs do unambiguously exist (they are defined on the same vertex set V^I), with external *SR*-graphs the situation is different, the bijective mapping between V^E and V'^E sets being not defined in advance. Generally, such mapping may not exist because V^E and V'^E sets can be different. For constructing external *SR*-graphs, we would have to complete the one of the V^E and V'^E sets containing fewer elements so that the cardinality of the two sets be identical (say, m), and then, to construct all the $m!$ bijections between the two sets, for each bijection to construct the G_R^E graph in the same manner as we constructed the G_R^I graph, and from among all of the graphs to pick out those that are not isomorphous and contain the minimal number of edges. The problem is illustrated by Table I. The first pair of *synthons* in this table illustrates the necessity to choose G_R^E graphs with the minimal number of edges, the second pair illustrates the non-uniqueness of these graphs (the so-called minimal *SR*-graphs).

Although the problem in question can be solved, using additional constraints, in less than $m!$ steps, we suggest a different, considerably more effective, approach to establishing all non-isomorphous external *SR*-graphs. The effectivity of this approach consists in the fact that all correct solutions are found and none has to be rejected.

The procedure starts from minimization of the external *SR*-graph on the individual atoms. This minimization is performed as follows. We introduce the notion of the external valence state of atom A_i in the *synthon* $S(A)$ (or $S'(A)$) as the valence state comprising only the external valences of this atom in the *synthon* $S(A)$ (or $S'(A)$). Now, describe this valence state by the vector[5,6] $\boldsymbol{v} = (0, v_2, v_3, v_4)$ (or $\boldsymbol{u} = (0, u_2, u_3, u_4)$) and put

$$\boldsymbol{w} = (w_1, w_2, w_3) = (u_2 - v_2,\ u_3 - v_3,\ u_4 - v_4)\,. \tag{6}$$

Vector $\boldsymbol{w}$ can be expressed in the x, y, z basis,

$$x = (-1, 0, 0) \tag{7a}$$

$$y = (1, -1, 0) \tag{7b}$$

$$z = (0, 1, -1)\,, \tag{7c}$$

of space $E^{(3)}$ where vectors x, y, z model vanishing of a single, a double and a triple bond, respectively; hence,

$$\boldsymbol{w} = (w_1, w_2, w_3) = a(-1, 0, 0) + b(1, -1, 0) + c(0, 1, -1)\,. \tag{8}$$

From this, the coordinates a, b, c are

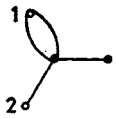
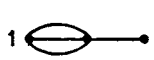
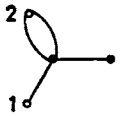
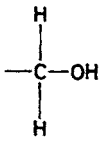
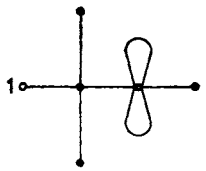
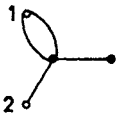
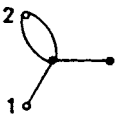
$$a = -w_1 - w_2 - w_3 \tag{9a}$$

$$b = -w_2 - w_3 \tag{9b}$$

$$c = -w_3 . \tag{9c}$$

TABLE I

Forms of external reaction graphs for various numbering of virtual vertices of S-graph. Virtual vertices are labelled ○, non-virtual vertices, ●

Starting *synthon*	Numbering of vertical vertices of S-graph	Product *synthon*	Numbering of virtual vertices of S-graph	External SR-graph G_R^E
				
				
				
				

The minimized number of forming external bonds N_+ and the minimized number of vanishing external bonds N_- at atom A_i during the $S(A) \rightarrow S'(A)$ conversion can be written as[5]

$$N_+ = F(-a) + F(-b) + F(-c) \tag{10a}$$

$$N_- = F(a) + F(b) + F(c)\,, \tag{10b}$$

where $F(x) = x$ for $x > 0$ and $F(x) = 0$ for $x \leqq 0$. Since no more than three edges can form or vanish between a non-virtual and a virtual vertex, all the conceivable combinations of a, b, c coordinates can be readily established combinatorially. These combinations are given in Table II. There exist only 12 combinations of a, b, c coordinates for all the possible edges (or multiedges) in the subgraph of graph G_R^E for atom A_i and any fixed virtual vertex. In general, however, a change in the edges of atom A_i at several virtual vertices can take place. All such changes are included in coordinates a, b, c. Now, let us seek for all expressions of vector (a, b, c) by means of vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{12}$ from Table II. Mathematically, this is expressed by Eq. (*11*):

$$(a, b, c) = \sum_{i=1}^{12} t_j \mathbf{s}_j\,, \tag{11}$$

where t_j are suitable parameters. This equation must have at least one solution $t_1, t_2, \ldots, t_{12}$ because, for instance, vectors $\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$ form a basis of space $E^{(3)}$. It is therefore possible to describe the vanishing or forming bonds by vectors $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_6$ and $\mathbf{s}_7, \mathbf{s}_8, \ldots, \mathbf{s}_{12}$, respectively. For atom A_i we have

$$t_1 + t_2 + t_3 + 2t_4 + 2t_5 + 3t_6 = N_- \tag{12a}$$

$$t_1 + t_2 + t_3 + 2t_4 + 2t_5 + 3t_6 = N_+\,. \tag{12b}$$

TABLE II

Combinatorial analysis of the relation of coordinates a, b, c and edges in the G_S^E subgraph for pairs of virtual (○) and nonvirtual (●) vertices

a, b, c, vector	s_1	s_2	s_3	s_4	s_5	s_6	s_7	s_8	s_9	s_{10}	s_{11}	s_{12}
a	1	0	0	1	0	1	−1	0	0	−1	0	−1
b	0	1	0	1	1	1	0	−1	0	−1	−1	−1
c	0	0	1	0	1	1	0	0	−1	0	−1	−1
G_R^E graph edges												

By combining Eqs (*11*) and (*12a, b*), a system of five equations in twelve unknowns is obtained for A_i. This system is solvable under the condition that $N_- = N_+ + a + b + c$, which is always satisfied (cf. Eq. (*10a, b*)). Thereby, the system transforms into one of four linearly independent equations in twelve unknowns. This system, without constraints, has an infinite number of solutions $t_1, t_2, \ldots, t_{12}$, which are dependent on eight parameters. The solutions can be expressed, for instance, in the form

$$t_5 = 2b - a - c + t_1 - 2t_2 + t_3 - t_4 - (3t_7 - 3t_8 + 3t_9 + 3t_{12} - N_+)/2 \quad (13a)$$

$$t_6 = a - b + c - t_1 + t_2 - t_3 + t_7 - t_8 + t_9 + t_{12} \quad (13b)$$

$$t_{10} = c - b + t_2 - t_3 + t_4 - t_8 + t_9 \quad (13c)$$

$$t_{11} = b - c - t_2 + t_3 - t_4 - (t_7 - t_8 + 3t_9 + 3t_{12} - N_+)/2 \,. \quad (13d)$$

The $t_1, t_2, \ldots, t_{12}$ values must be natural numbers and must meet the following conditions:

$$t_1 \leqq N_- ; \quad t_7 \leqq N_+ ; \quad t_2 \leqq N_- - t_1 ; \quad t_8 \leqq N_+ - t_7 ;$$

$$t_3 \leqq N_- - t_1 - t_2 ; \quad t_9 \leqq N_+ - t_7 - t_8 ;$$

$$t_4 \leqq (N_- - t_1 - t_2 - t_3)/2 ; \quad t_{12} \leqq (N_+ - t_7 - t_8 - t_9)/3 ;$$

$$t_1 \leqq v_1 ; \quad t_7 \leqq u_1 ; \quad t_2 + t_4 \leqq v_2 ; \quad t_8 + t_{10} \leqq u_2 ;$$

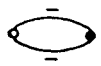
$$t_3 + t_5 \leqq v_3 ; \quad t_9 + t_{11} \leqq u_3 \,. \quad (14)$$

The last six constraints are a formal expression of the fact that the number of vanishing bonds cannot be higher than the number of initially present bonds, similarly as the number of forming bonds cannot be higher than the number of bonds present after the reaction.

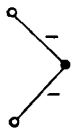
For each particular solution (*13a–d*) there is a single form of subgraph of the external *SR*-graph for atom A_i (contains a vertex corresponding to atom A_i and all edges incident with it), which is minimized. The number of virtual vertices in this subgraph is given by the sum $t_1 + t_2 + \ldots + t_{12}$. The number of edges and their evaluation are given in Table III, t_j expressing the number of virtual vertices that will be joined with A_i by an edge (multiedge) corresponding to t_j according to Table III.

Example 2. Consider the transformation $\diagup\!\!\diagdown C^I\!-\!\underset{|}{C}\!- \rightarrow -C^I\!\equiv\!C-$. For atom C^I, we have for the vectors of the external valence states: $\mathbf{v} = (0, 1, 1, 0)$, $\boldsymbol{u} = (0, 1, 0, 0)$. From Eq. (*6*), $\boldsymbol{w} = (0, -1, 0)$ and from Eqs (*9a–c*), $a = b = 1$, $c = 0$. From Eqs (*13a–d*) and the associated conditions for t_j's we obtain two solutions, $\mathbf{t}^1 =$

$= (0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0)$, $\mathbf{t}^2 = (1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$. From Table III, the subgraph of *SR*-graph G_R^E for atom C^1 is in the form



for solution $\mathbf{t}^1$ and



for solution $\mathbf{t}^2$. The former graph corresponds to detachment of substituent bonded by the double bond, the latter, to addition of the double bond and detachment of substituent bonded by the single bond.

All minimal external *SR* graphs G_R^E for the $S(A) \rightarrow S'(A)$ conversion are obtained by all the possible combinations of all subgraphs of the external *SR*-graphs of each atom A_i. Finding all minimal external *SR*-graphs is in principle the same problem as finding all maximal common subgraphs[7]. Here the problem was solved algebraically.

A single *SR*-graph G_R then corresponds to each minimal *SR*-graph G_R^E:

$$G_R = G_R^I \cup G_R^E = (V_R, E_R, L_R, \psi, \omega, \{-1, 1\}) = \\ = (V_R^I \cup V_R^E, E_R^I \cup E_R^E, L_R, \psi, \omega, \{-1, 1\}), \qquad (15)$$

where G_R^I is defined by Eq. (*4*) and mapping ψ is defined as $\psi(e) = \psi^I(e)$ for $e \in E_R^E$ and $\psi(e) = \psi^E(e)$ for $e \in E_R^E$. All G_R^E graphs are generated by the GEN-SRG algorithm described below.

TABLE III

Correspondence between solution $t_1, t_2, \ldots, t_{12}$ from Eq. (*13a–d*) and edges in the external *SR*-graph incident with the vertex corresponding to atom A_i

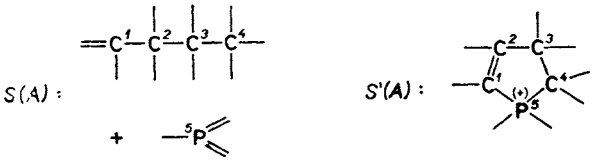
Unit parameter value	t_1 t_2 t_3	t_4 t_5	t_6	t_7 t_8 t_9	t_{10} t_{11}	t_{12}
G_R^E graph edges						

GEN-SRG *Algorithm*

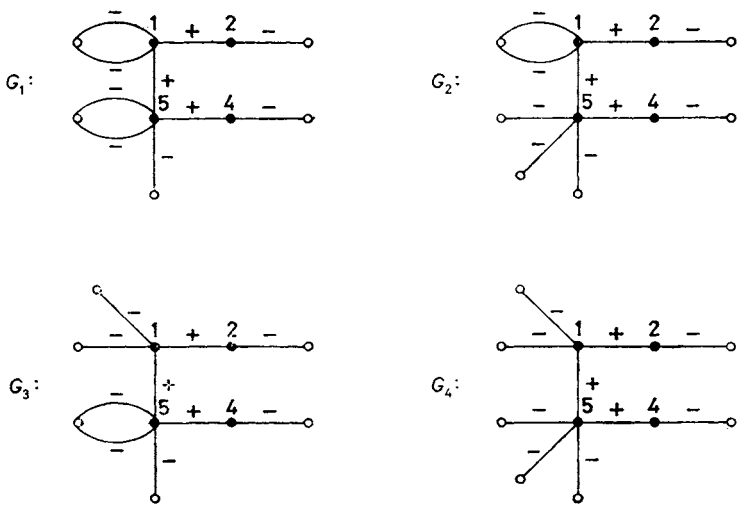
This algorithm generates external *SR*-graphs G_R^E in the form $G_R^E = (V_R^I, E_R^E, \emptyset, \psi^E, \{-1, 1\})$. Introduce the following symbols: $T^i = \{T_1^i, T_2^i, \ldots, T_m^i\}$ be the set of all solutions (*13a–d*) for atom A_i and transformation $S(A) \to S'(A)$; $n = \text{card}\, A$, $m = n$, $\mathfrak{R} = \emptyset$, $m_i = \text{card}\, T^i$; U be the set $U = (U_1, U_2, \ldots, U_z)$, where $U_j = (u_1^j, u_2^j, \ldots, u_n^j)$, $z = \text{card}\, U$, $1 \leqq u_k^j \leqq m_k$, $1 \leqq j \leqq z$, $1 \leqq k \leqq n$, $V_R^E = E_R^E = \emptyset$. The algorithm is as follows:

1. Cycle for $i = 1$ to z
2. Cycle for $j = 1$ to n
3. Be $\mathbf{t} = (t_1, t_2, \ldots, t_{12}) = T_{u_{jt}}^j$
4. Cycle for $k = 1$ to 12
5. If $t_k = 0$ go to 12
6. Cycle for $s = 1$ to t_k
7. $m = m + 1$, $V_R^E = V_R^E \cup \{v_m\} \cup \{v_j\}$
8. $e = \{v_j, v_m\}$, $E_R^E = E_R^E \cup \{e\}$, if $k \leqq 6$ then $\psi^E(e) = -1$ else $\psi^E(e) = 1$
9. If $(k > 3 \wedge k < 7) \wedge (k > 9)$ then $E_R^E = E_R^E \cup \{e\}$
10. If $(k = 6) \wedge (k = 12)$ then $E_R^E = E_R^E \cup \{e\}$
11. End of cycle s
12. End of cycle k
13. End of cycle j
14. $\mathfrak{R} = \mathfrak{R} \cup G_R^E$, $V_R^E = E_R^E = \emptyset$
15. End of cycle i
16. End of algorithm

Example 3. Consider the transformation $S(A) \to S'(A)$ where



Using the GEN-SRG algorithm, four external *SR*-graphs were obtained which with the internal *SR*-graph give the following four *SR*-graphs G_1, G_2, G_3, G_4:



The crucial role in this transformation is played by atoms C^1 and P, from which the highest numbers of edges start in the graph. According to the *SR*-graph, the following four processes can occur at these atoms.

Process 1. Substitution by the phosphorus atom for the substituent initially bonded to C^1 by the double bond. The substituent bonded to C^1 by the double bond thus is eliminated.

Process 2. Substitution by the phosphorus atom for the substituent initially bonded to C^1 by the single bond. The substituent bonded to C^1 by the single bond thus is eliminated.

Process 3. Addition of the C^1 and C^4 atoms to the double bonds of the phosphorus atom and elimination of the substituent initially bonded to the latter by a double bond, or addition of one of the atoms C^1, C^4 to one of the double bonds at the phospohorus atom, substitution by the remaining of the C^1, C^4 atoms for the substituent initially bonded to the phosphorus atom by the double bond and addition of some virtual atom to the other double bond at the phosphorus atom. Which of the events takes place is determined by analysis of the particular mechanism with particular substituents. The substituent initially bonded to the phosphorus atom by a double bond thus is eliminated.

Process 4. Addition of the C^1 and C^4 atoms to the double bonds at the phosphorus atom elimination of the substituent initially bonded to the latter by the single bond, or addition of one of the C^1, C^4 atoms to one of the double bonds at the phosphorus atom, substitution by the remaining of the C^1, C^4 atoms for the substituent initially bonded to the phosphorus atom by the single bond and addition of some virtual

atom to the other double bond at the phosphorus atom. Which of the events takes place is determined by analysis of the particular mechanism with particular substituents. The substituent initially bonded to the phosphorus atom by the single bond thus is eliminated.

All of the G_1 through G_4 graphs include double bond rearrangement from the C atom be between the C^1 and C^2 atoms. Graph G_1 includes, in addition, a combination of processes 1 and 3, graph G_2, a combination of processes 1 and 4, graph G_3, a combination of processes 2 and 3, and graph G_4, a combination of processes 2 and 4.

The above example demonstrates that the *SR*-graph concept can serve as a convenient tool for the study of reaction mechanisms. In our approach, generation of *SR*-graphs is necessary for the calculation of the *reaction distance*. This problem is the subject of the next paper in this series.

The author wishes to thank Associate Professor M. Kratochvíl (Institute of Pure Chemicals, Lachema, Brno), Prof. V. Kvasnička and Dr J. Pospichal (Slovak Institute of Technology, Bratislava) for stimulating discussions concerning the problem of the graph model of the synthon.

REFERENCES

1. Koča J.: Collect. Czech. Chem. Commun. *53*, 1007 (1988).
2. Kvasnička V.: Collect. Czech. Chem. Commun. *48*, 2097 (1983).
3. Kvasnička V.: Collect. Czech. Chem. Commun. *48*, 2118 (1983).
4. Kvasnička V., Kratochvíl M., Koča J.: Collect. Czech. Chem. Commun. *48*, 2284 (1983).
5. Koča J., Kratochvíl M., Kunz M., Kvasnička V.: Collect. Czech. Chem. Commun. *49*, 1247 (1984).
6. Kratochvíl M., Koča J., Kvasnička V.: Chem. Listy *79*, 807 (1985).
7. Baláž V., Koča J., Kvasnička V., Sekanina M.: Cas. Pest. Matem. *111*, 431 (1986).

Translated by P. Adámek.