

УДК 547.1+541.60+681.3.06

Ивар УГИ

ОБ ИСПОЛЬЗОВАНИИ НОВЫХ ФОРМАЛЬНО-ЛОГИЧЕСКИХ МЕТОДОВ ДЛЯ РЕШЕНИЯ ХИМИЧЕСКИХ ПРОБЛЕМ

(Представил К. Лээтс)

Предсказание с помощью ЭВМ молекулярной структуры соединений и путей протекания химических реакций основывается либо на эмпирических данных, либо на методах формальной логики без детализированной информации [1-4]. Компьютерные программы, разработанные на основе математической модели логической структуры химии, способны генерировать в принципе все возможные решения поставленных задач, в том числе и решения нереальные в химическом смысле, а кроме того, решения без прецедента. В последнем случае происходит освоение новой «химической области». Так как число возможных решений огромно, наиболее важное значение имеют методы и процедуры, уменьшающие потенциально генерируемое количество информации. Поэтому в исследованиях компьютерной химии основное внимание уделяют разработке произвольных и понятных процедур, резко суживающих число возможных результатов [5]. Предъявляемым критериям отвечают только процедуры, базирующиеся на формализме отбора. Процедуры, основывающиеся на эвристических или на различных физико-химических правилах [6], могут быть использованы только после применения формальных логических процедур отбора.

Логическая структура химии

Для представления теории *BE*- и *R*-матриц (bond- and electron- and reaction matrices) в виде алгебраической модели структурной химии [4] необходимо сформулировать некоторые основные понятия.

Химия совокупности атомов $A = \{A_1, \dots, A_n\}$ имеет хорошо дефинированные граничные условия и инварианты. Логическая структура такой химии может быть легко представлена в математических терминах. Когда молекулы образованы как агрегаты совокупности из n атомов $A = \{A_1, \dots, A_n\}$ с эмпирической формулой A , остовы атомов (атомное ядро + внутренние оболочки электронов) связаны между собой общей парой валентных электронов. Некоторые валентные электроны могут находиться в остове и как свободные. Из совокупности A могут быть образованы не только молекулы с эмпирической формулой A , но и ансамбли молекул $EM(A)$ (EM — ensembles of molekules), которые содержат более одной молекулы. Каждый атом из $A = \{A_1, \dots, A_n\}$ содержится в $EM(A)$ только один раз. Множество всевозможных $EM(A)$ составляет семейство изомерных $EM(A) \rightarrow FIEM(A)$ (family of isomeric $EM(A)$). Каждая реакция в $FIEM(A)$ является изомеризацией $EM(B) \rightarrow EM(E)$, где оба, $EM(B)$ и $EM(E)$, — члены $FIEM(A)$. Тогда вся возможная химия $A = \{A_1, \dots, A_n\}$ содержится в $FIEM(A)$. Так как каждую совокупность атомов можно рассматривать как A , теорию $FIEM$ можно считать теорией химии в целом.

По теории BE - и R -матриц, химическая реакция $EM(B) \rightarrow EM(E)$ является реакцией превращения исходного $EM(A)$ (at the beginning) в изомерное $EM(A)$ в конце реакции (at the end) и выражается уравнением

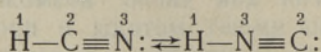
$$B + R = E.$$

Симметричные $(n \times n)$ -матрицы $B = \langle b_{ij} \rangle$ и $E = \langle e_{ij} \rangle$ описывают строение $EM(B)$ и $EM(E)$ следующим образом: атом A_i относится к i -му ряду/столбцу матриц B и E , а элемент $b_{ij} = b_{ji}$ или $e_{ij} = e_{ji}$ ($i \neq j$) показывает кратность ковалентной связи между A_i и A_j ; b_{ii} показывает, сколько свободных электронов относится к A_i . Данную $EM(A)$ можно выразить до $n!$ эквивалентными BE -матрицами. Так как химическая сущность $EM(A)$ не зависит от нумерации атомов, любые две BE -матрицы представляют ту же самую $EM(A)$, если они взаимопревращаемы с помощью перестановки рядов/столбцов в соответствии с перенумерацией атомов.

$(n \times n)$ - R -матрица $R = \langle r_{ij} \rangle$ выражает перераспределение валентных электронов во время превращения $EM(B) \rightarrow EM(E)$; элементы $r_{ij} = r_{ji}$ показывают изменение кратности ковалентных связей, а r_{ii} — изменение положения свободных электронов. В ходе химической реакции никаких валентных электронов не образуется и не исчезает. Поэтому

$$\sum_{ij} (e_{ij} - b_{ij}) = \sum_{ij} r_{ij} = 0.$$

Согласно представленной теории, для химической реакции



уравнение $B + R = E$ описывается следующим образом:

$$\begin{array}{ccc|c} 0 & 1 & 0 & \\ 1 & 0 & 3 & \\ 0 & 3 & 2 & \\ \hline & B & & \end{array} + \begin{array}{ccc|c} 0 & -1 & 1 & \\ -1 & 2 & 0 & \\ 1 & 0 & -2 & \\ \hline & R & & \end{array} = \begin{array}{ccc|c} 0 & 0 & 1 & \\ 0 & 2 & 3 & \\ 1 & 3 & 0 & \\ \hline & E & & \end{array}$$

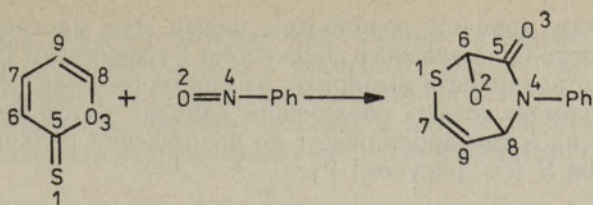
Глобальная геометрическая модель химии, химическая дистанция и иерархия химических реакций

Так как $(n \times n)$ -матрицу $B = \langle b_{ij} \rangle$ можно изобразить точкой $p(B) = (b_{11}, \dots, b_{n1}; b_{21}, \dots, b_{2n}; b_{n1}, \dots, b_{nn})$ или вектором в n^2 -мерном евклидовом пространстве, химию $FIEM(A)$ можно также представить терминами n^2 -мерной геометрической модели, где $EM(A)$ образует решетку точек, соединенных векторами, представляющими их превращения в ходе химических реакций [4, 7].

Химическая дистанция (CD) (chemical distance) — метрика в пространстве L_1 [8] между точками $p(B)$ и $p(E)$, представляющих $EM(B)$ и $EM(E)$, — выражается в виде

$$d(B, E) = \sum_{ij} |e_{ij} - b_{ij}| = \sum_{ij} |r_{ij}|$$

и составляет двойное количество валентных электронов, которые перераспределяются в ходе реакции $EM(B) \rightarrow EM(E)$ или ее обратной реакции. Химическая дистанция между $EM(B)$ и $EM(E)$ зависит от последовательности сочетания атомов (atom-by-atom matching) в двух EM , т. е. если нумерация атомов $EM(B)$ задана, то CD зависит от нумерации $EM(E)$ [9]. Последовательность сочетания атомов изомерных EM также показывает, какие связи будут разорваны/образованы при взаимопревращении этих EM . Примером является реакция [10]:



Разрывающиеся связи: 1=5, 2=4, 3—8, 7=6, 8=9;
 Образующиеся связи: 1—6, 1—7, 2—6, 2—8, 3—5,
 4—5, 4—8, 7—9

Принцип минимальной химической дистанции (ПМХД) [4, 7, 10–12] заключается в том, что в случае одноэтапной химической реакции предпочитается те направления, на которых перераспределяется минимальное количество валентных электронов.

Программа PEMCD (program for determining the exact minima of chemical distance) [7] определяет для пары изомерных *EM* их *CD*-минимум. Возможно существование более одного минимума и соответствующей последовательности сочетания атомов [9, 12, 13].

Но самым важным аспектом последовательности сочетания атомов является иерархическая классификация химических реакций, которая следует из идентификации реакционных центров и разрываемых/образуемых связей. Эта иерархическая классификация служит формальной основой для новой системы документации реакций [14, 15] и для новых интерактивных процедур, которые уменьшают число возможных из комбинарики результатов в компьютерной химии [2, 5, 16].

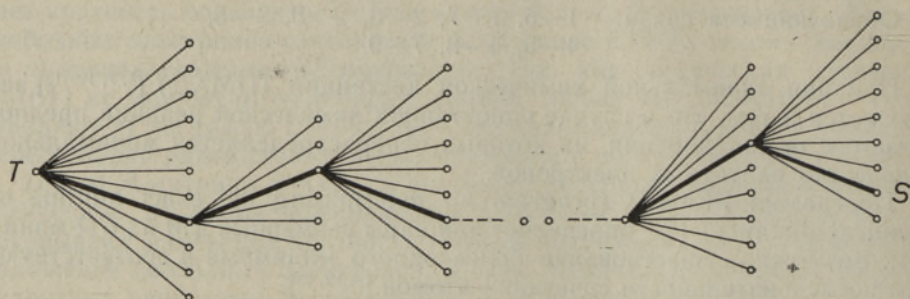
Из последовательности сочетания атомов в *EM(B)* и *EM(E)*, которые взаимопревращаемы при помощи химических реакций, следует матрица реакции $R = E - B$. Ряды/столбцы матрицы *R*, которые содержат ненулевые элементы, соответствуют реакционным центрам в *EM(B)* и *EM(E)*, т. е. тем атомам, которые непосредственно участвуют в процессах разрывов/образований связей; они образуют «ядро реакции».

После исключения всех рядов/столбцов матрицы *R*, которые не содержат ненулевые элементы, получается ее несократимая матрица \hat{R} ; соответственно, сокращенные *BE*-матрицы \hat{B} и \hat{E} образуют «ядро реакции». Несократимая \hat{R} -матрица представляет собой схему перераспределения электронов и т. н. реакций *R*-класса, которые охватывают реакции с одинаковым характером перераспределения электронов и одинаковым числом реакционных центров [14], например, реакции с шестью реакционными центрами и циклическим переносом трех пар электронов — реакции Дильса—Альдера, иенов и перегруппировка Коупа [2]. *R*-класс подразделяется на *RB*-подклассы (*B*: bond), члены которых обычно имеют одинаковые инвариантные ковалентные связи между реакционными центрами. Реакции *RB*-подклассов являются базовыми реакциями. *RB*-подклассы подразделяются на *RA*-подклассы (*A*: atom), которые идентифицируются по химическому элементу в реакционном центре [15]. Индивидуальные реакции занимают нижнюю ступень в иерархической классификации химических реакций.

Генерация и отбор доступных синтезов по дереву решений

Целью ранней компьютерной химии было создание программ, которые могли предложить для данного целевого соединения подходящие исходные материалы и последовательность реакций для получения целевого соединения [17, 18].

К настоящему времени разработан целый ряд программ ретросинтеза, которые исходя из целевого соединения T генерируют большое количество предшественников и выбирают из них одного или несколько. Для выделенных промежуточных соединений находят своих предшественников, отбирают их и так продолжают до нахождения подходящих исходных материалов S (см. рисунок) [1].



Процесс ретросинтеза.

Разные программы ретросинтеза отличаются по способности инновации, что зависит от используемого метода поиска предшественников данной молекулы. Программы, которые базируются на библиотеках реакций, конечно, менее инновативны, чем те, в которых используются эмпирические [19] или совсем формальные методы [1, 3, 20, 21]. Качество конкретного результата зависит не только от качества системы генерации предшественников, но и от системы выбора. Однако надо отметить, что ни одна из доступных процедур выбора не отвечает требованиям, предъявляемым к многоступенчатому процессу выбора.

Алгебраическое решение проблем химии

Фундаментальное уравнение

$$B + R = E$$

теории BE - и R -матриц является уравнением матрицы Диофантина с большим количеством решений при граничных условиях, имеющих химический смысл. Два основных типа решения фундаментального уравнения соответствуют двум типам компьютерных программ для дедуктивного решения проблем химии [22].

1. Если BE -матрица B задана, решениями $B + R = E$ являются пары (R, E) . Эти решения получаются с использованием т.н. генераторов реакции типа 1 ($RG1$) (reaction generators) [12, 22].

2. Пары (B, E) являются решением $B + R = E$, когда R -матрица R задана. Решения получаются с помощью генераторов типа 2 ($RG2$) [12, 16, 22, 23].

Генераторы реакции составляют основную часть компьютерных алгебраических программ. Усовершенствованными генераторами реакции являются генераторы, управляемые таблицами перехода (TRG). Они, в отличие от предыдущих версий RG , сразу при генерации результатов учитывают граничные условия [5].

В уравнении $B+R=E$ каждый ряд/столбец матриц B , R , E предназначен для определенного атома и элемента. Таким образом, каждый ряд/столбец матриц B и E должен соответствовать возможной валентной схеме конкретного химического элемента и любой ряд/столбец матрицы R должен показывать разрешенные изменения в валентной схеме данного элемента.

Для каждого элемента эти условия представляют в виде таблицы перехода. Ряды таблицы показывают валентные схемы данного химического элемента в начале реакции, а столбцы — в конце реакции. Элементы «+» и «-» в пересечении рядов и столбцов показывают, разрешены ли эти переходы в валентных схемах [12, 22].

С помощью TRG1, исходя из BE -матрицы B исходных материалов $EM(B)$, можно генерировать через пары (R, E) все возможные продукты $EM(E)$. Так как $B+R=E$ не имеет направления, то $EM(E)$ являются в то же время возможными исходными материалами для $EM(B)$.

Проблему выделения из большого числа малого количества направлений синтеза легче решить, если дизайн синтеза проводить двухсторонним (билатеральным) методом [24]. Используя этот метод, из списка доступных исходных материалов отбирают с помощью подструктурной корреляции наиболее подходящие в первую очередь для целевой молекулы. Так как такая система имеет два фиксированных конечных узла, решение проблемы по сравнению с открытым деревом решений менее трудоемко.

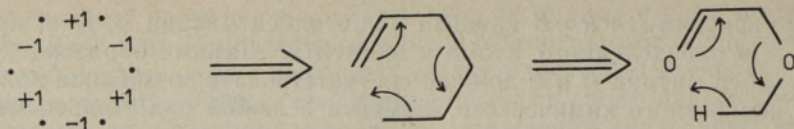
Систему направлений реакции между изомерным EM лучше генерировать таким образом, чтобы дать системе возможность расти с обоих концов $EM(B)$ и $EM(E)$ в направлении друг к другу до закрытия системы, что определяется при помощи алгоритма CANON [25]. Высокая эффективность процедуры обеспечивается тем, что генерируются только такие направления реакции, которые находятся в области перекрывания двух L_1 -гиперсфер вокруг $EM(B)$ и $EM(E)$ с L_1 -радиусами $d(B, E)$, которыми пользователь может варьировать [5].

Программа RAIN (Reaction And Intermediates Network) разработана на основе этой концепции на FORTRAN-77 для персональных компьютеров. RAIN используется не только как функциональный элемент в системе двухстороннего дизайна синтеза, но и как программа для выявления механизмов реакций.

Модификация RAIN2, используя TRG1, может оперировать 32 реакционными центрами [26]. RAIN2 учитывает также требования, выдвинутые в [13, 27]. Программа RAIN неоднократно доказала свою жизнеспособность при генерации механистических направлений для известных реакций, в частности для побочных реакций четырехкомпонентных конденсаций [28, 29]. Во всех случаях ранее установленные механизмы реакций были найдены в сгенерированных системах, а некоторые из новых не противоречат экспериментальным данным [13].

Вторая основная компьютерная программа IGOR (Interactive Generation of Organic Reactions) [2, 23, 30], базирующаяся на TRG2 и дополняющая RAIN, генерирует химические реакции исходя из данной схемы перераспределения электронов [31]. Кроме тех же правил и процедур отбора, что и в программе RAIN, IGOR еще опирается на вышеприведенную иерархическую классификацию химических реакций. Химические реакции в программе IGOR внутренне документированы терминами $B+10E$, где для B и E на основе ПМХД установлена последовательность сочетания атомов [2, 16, 32, 33]. На каждом уровне пользователь программы может проверить промежуточный результат и выбрать подкласс результата на следующем, нижнем уровне.

Примером может служить открытие CO_2 -экструзии из α -формил-оксикетонов [2, 16]:



R-класс:
несократимая
R-матрица *R*

RB-подкласс
или базовая
реакция
 $B + R = E$

RA-подкласс:
подстановка хи-
мических эле-
ментов в реак-
ционные центры

Недавно с помощью IGOR была предсказана новая, не имевшая прецедента и аналога реакция и реализована в лабораторных условиях [33].

Если вводом в IGOR служит нулевая матрица $R=0$, то результатом является $B=E$, и структуры генерируются в соответствии со спецификацией пользователя. Таким образом получили, например, 278 потенциальных 5-членных циклических фосфорилирующих реагентов [33] и 52 аналога сиднонов [16].

Заключение

Первые иллюзии относительно возможностей компьютерной химии рассеялись. Надежды, что экспертные системы и искусственный интеллект смогут решить проблемы химии автоматически и комплексно, не оправдались. На смену пришел здравый смысл. Начинает доминировать мнение, что участие химика-исследователя в компьютерной химии необходимо. Наиболее результативным является максимальное использование потенциала обоих — человека и компьютера. При этом компьютеры применяют для решения задач, которые могут быть описаны терминами алгоритмов. В то же время пользователь, опираясь на свои знания, опыт и интуицию, выбирает наиболее подходящие для решения проблемы пути реакции.

Таким образом, в компьютерной химии наибольший эффект дают такие экспертные системы, в которых используются умение и знания самих исследователей.

Автор статьи признателен сотруднику Института химии АН Эстонии К. Сийрде за участие в составлении данного обзора.

ЛИТЕРАТУРА

1. Ugi, I., Bauer, J., Brandt, J., Friedrich, J., Gasteiger, J., Jochum, C., Schubert, W. Neue Anwendungsgebiete für Computer in der Chemie // *Angew. Chem.*, 1979, **91**, 99—111.
2. Bauer, J., Herges, R., Fontain, E., Ugi, I. IGOR and computer assisted innovation in chemistry // *Chimia*, 1985, **39**, 43—53.
3. Ugi, I., Bauer, J., Baumgartner, R., Fontain, E., Forstmeyer, D., Lohberger, S. Computer assistance in the design of syntheses and a new generation of computer programs for the solution of chemical problems by molecular logic // *Pure and Appl. Chem.*, 1988, **60**, 1573—1582.
4. Dugundji, J., Ugi, I. An algebraic model of constitutional chemistry as a basis for chemical computer programs // *Topics Curr. Chem.*, 1973, **39**, 19—64; Ugi, I. Qualitative mathematical concepts, models and theories for computer assistance in chemistry // *J. Indian Chem. Soc.*, 1985, **62**, 864—868.
5. Ugi, I., Bauer, J., Fontain, E. Transparent formal methods for reducing the combinatorial wealth of conceivable solutions to a chemical problem — computer assisted elucidation of complex reaction mechanisms // *Anal. Chim. Acta*, 1990, **225**, 155—163.
6. Gasteiger, J., Hutchings, M. G., Christophs, G. L., Hiller, C. L., Marsili, M., Saller, H., Yuki, K. A new treatment of chemical reactivity: development expert system for reaction prediction and synthesis design // *Topics Curr. Chem.*, 1987, **137**, 19—73.

7. Wochner, M., Ugi, I. Molecular logic and computer assistance in chemistry // J. Mol. Struct. (Theochem.), 1988, **165**, 229—242; Wochner, M., Brandt, J., Scholley, A. v., Ugi, I. Chemical similarity, chemical distance and its exact determination // *Chimia*, 1988, **42**, 217—225.
8. Dugundji, J. *Topology*. Allyn and Bacon, Boston, 1966.
9. Jochum, C., Gasteiger, J., Ugi, I., Dugundji, J. The principle of minimum chemical distance and the principle of minimum structure change // *Z. Naturforsch.*, 1982, **37B**, 1205—1215.
10. Augelman, G., Fritz, H., Rihs, G., Streith, J. An unexpected addition product of nitrobenzene with pyran-2-thione // *J. Chem. Soc., Chem. Commun.*, 1982, **19**, 1112—1113; Dejoin, A., Augelman, G., Fritz, H., Geoffroy, G., Schmidlin, C., Streith, J. Cycloaddition of 2H-pyran-2-thiones with nitroso derivatives // *Helv. Chim. Acta*, 1985, **68**, 1998—2014.
11. Kolbe, H. Ueber die chemische Konstitution und Natur der organischen Radikalen // *Liebigs Ann. Chem.*, 1850, **75**, 211—239.
12. Fontain, E., Bauer, J., Ugi, I. Computer assisted bilateral generation of reaction networks from educts and products // *Chem. Lett.*, 1987, 37—40; Fontain, E., Bauer, J., Ugi, I. Computerunterstützte mechanistische Analyse der Streith-Reaktion mit dem Programm RAIN // *Z. Naturforsch.*, 1987, **42B**, 889—891.
13. Ugi, I., Wochner, M., Bauer, J., Gruber, B., Karl, R. *Concepts and Applications of Chemical Similarity*. Wiley, New York, 1990.
14. Brandt, J., Gasteiger, J., Jochum, C., Schubert, W., Ugi, I. Computer programs for the deductive solution of chemical problems on the basis of a mathematical model of chemistry // *Computer Assisted Organic Synthesis. Symp. Ser. No. 61*, Amer. Chem. Soc., Washington DC, 1977, 33—59.
15. Brandt, J., Bauer, J., Frank, R. M., Scholley, A. v. Classification of chemical reactions by electron shift patterns // *Chem. Scripta*, 1981, **18**, 53—60.
16. Bauer, J. IGOR 2: a PC-program for generating new reactions and molecular structures // *Tetrahedron Comput. Methodol.*, 1990, **2**, 24—32.
17. Corey, E. J., Wipke, W. T. Computer-assisted design of complex organic synthesis // *Science*, 1969, **166**, 178—192.
18. Gelernter, H., Sridharan, N. S., Hart, H. J., Yen, S. C., Shue, F. W. Discovery of organic synthetic routes by computer // *Topics Curr. Chem.*, 1973, **41**, 113—150.
19. Hendrickson, J. B., Toczku, A. G. Synthesis design logic and the SYNGEN (synthesis generation) program // *Pure and Appl. Chem.*, 1988, **60**, 1563—1572.
20. Koca, J., Kratochvil, M., Kvasnicka, V., Matyska, L., Pospichal, J. *Synthon model of organic chemistry and synthesis design*, Lecture Note Ser. 51, Springer, Heidelberg, 1989.
21. Zefirov, N. S. An approach to systematization and design of organic reactions // *Accounts Chem. Res.*, 1987, **20**, 237—243; Трач С. С., Баскин И. И., Зефуров Н. С. Проблемы молекулярного дизайна и ЭВМ. XIII. Систематический анализ органических процессов, характеризующихся незамкнутыми топологиями перераспределения связей // *Ж. орг. хим.*, 1988, **24**, 1121—1136.
22. Bauer, J., Fontain, E., Ugi, I. Computer-assisted bilateral solution of chemical problems and generation of reaction networks // *Anal. Chim. Acta*, 1988, **210**, 123—134.
23. Bauer, J., Ugi, I. Chemical reactions and structures without precedent generated by computer programs // *J. Chem. Res.*, 1982, (S) 298; (M) 3101—3196.
24. Ugi, I., Bauer, J., Brandt, J., Friedrich, J., Gasteiger, J., Jochum, C., Schubert, W., Dugundji, J. Computer programs for the deductive solution of chemical problems on the basis of mathematical model — a systematic bilateral approach to reaction pathways // *Computational Methods in Chemistry*. Plenum Press, New York, 1980, 275—300.
25. Schubert, W., Ugi, I. Constitutional symmetry and unique descriptors of molecules // *J. Amer. Chem. Soc.*, 1978, **100**, 37—41.
26. Reitsam, K. Ein formaler Reaktionsgenerator auf der Basis der BE-R-Valenz-Transformatoren. // *Dipl. Arbeit*, TU Munich, 1988.
27. Schenk, P. Die relative Komplexität formaler Übergangszustände in Reaktionsnetzwerken. *Dipl. Arbeit*, TU Munich, 1988.
28. Ugi, I., Bauer, J., Fontain, E. Molecular logic and the deductive solution of chemical problems // *Wiss. Z. Techn. Hochsch. Leuna-Merseburg*, 1989, **31**, 9—17.
29. Ugi, I., Lohberger, S., Karl, R. The Passerini and Ugi reactions // *Comprehensive Organic Synthesis: Selectivity for Synthetic Efficiency*, 2. Pergamon, London, 1991, ch. 4.6.
30. Bauer, J., Ugi, I. Chemical reactions and structures without precedent generated by computer programs // *J. Chem. Res.*, 1982, (M) 3201—3260.
31. Knauer, M. Generierung von Reaktionsmatrizen unter bestimmten Randbedingungen. *Dipl. Arbeit*, TU Munich, 1988.

32. Ugi, I., Bauer, J., Fontain, E., Gotz, J., Hering, G., Jacob, P., Landgraf, B., Lemmen, P., Schneiderwind-Stocklein, R., Schwarz, R., Sluka, P., Balgobin, N., Chattopadhaya, J., Pathak, T., Zhou, X.-X. New phosphorylating reagents and protective group techniques for oligonucleotide synthesis, as well as computer assistance in the design of reagents // Chem. Scripta, 1986, 26, 205—215.
33. Ugi, I., Bauer, J., Fontain, E. Reaction pathways on a PC // Personal Computers for Chemists Elsevier, Amsterdam, 1990.

Мюнхенский технический
университет

Поступила в редакцию
28/V 1990

Ivar UGI

UUTE FORMAALLOOGILISTE MEETODITE KASUTAMISEST KEEMIAPROBLEEMIDE LAHENDAMISEL

On üldistatud teoreetilised alused keemiliste ühendite ja reaktsioonide kirjeldamiseks uute formaalloomiliste meetodite abil. On toodud keemia algebralis-loogilise struktuuri põhialused ning selle struktuuri mudelitel baseeruvad võtted ja arvutiprogrammid keemiaprobleemide deduktiivseks lahendamiseks.

Ivar UGI

APPLICATION OF NEW FORMAL-LOGICAL METHODS TO CHEMICAL PROBLEMS

The theoretical foundations of describing chemical compounds and reactions by new formal-logical methods are generalized. The foundations of an algebraic-logical structure of chemistry and methods based on its models, as well as computer programs for deductive solutions of chemical problems are presented.