

УДК 547.1 : 681.142.2

КОМПЬЮТЕРНЫЙ СИНТЕЗ

Зефирова Н. С., Гордеева Е. В.

Изложено состояние, проблемы и перспективы новой области компьютерной химии — планирование и поиск путей синтеза органических соединений с помощью ЭВМ (компьютерный синтез). Рассмотрены основные программные разработки, непосредственно использующиеся химиками-органиками. Проведен сравнительный анализ существующих программ компьютерного синтеза, выявлены общие черты и различие более 20 компьютерных систем, в том числе созданной авторами системы «ФЛАМИНГО».

Библиография — 123 ссылки.

ОГЛАВЛЕНИЕ

I. Введение	1753
II. Постановка задачи. «Дерево синтеза»	1754
III. Основные программы компьютерного синтеза	1755
IV. Представление и анализ химических структур	1756
V. Эмпирический и неэмпирический компьютерный синтез	1757
VI. Представление и анализ химических реакций	1758
VII. Критерии отбора	1761
VIII. Универсальная компьютерная система «ФЛАМИНГО» для решения проблем синтеза и установления механизмов реакций в органической химии	1768
IX. Перспективы развития компьютерного синтеза	1770

I. ВВЕДЕНИЕ

Одной из наиболее важных научных разработок по применению компьютеров в химии является поиск с помощью ЭВМ возможных путей синтеза органических соединений, получивший в литературе название «компьютерный синтез» (Computer-Assisted Synthesis), сокращенно КС. Начиная с 1969 г., когда была опубликована статья Кори и Уипке [1], содержащая описание первой программы компьютерного синтеза, данная область развивалась довольно быстрыми темпами, и в настоящее время можно говорить о КС не как о совокупности отдельных разработок, а как о большом научном направлении (см., например, обзорные статьи [2—7]).

Цель данного обзора — охватить основные аспекты проблемы КС. В то же время мы не включаем в обзор общие вопросы стратегии и планирования синтеза, поскольку эти вопросы уже получили достаточное освещение в литературе [8, 9]. Напротив, в данном обзоре мы будем по возможности учитывать компьютерную специфику проблемы поиска путей синтеза с помощью ЭВМ. Мы сочли более целесообразным выделить именно ключевые моменты в создании программного обеспечения для решения задач КС, не вдаваясь в подробности алгоритмизации и машинной реализации конкретных программ, поскольку число программ КС достаточно велико и постоянно растет. Мы полагаем, что в рамках выбранной структуры изложения можно будет более четко определить сходство и различие между существующими программами КС, выявить принципиальную новизну того или иного подхода.

В заключительной части обзора мы рассматриваем основные принципы построения разработанной нами компьютерной системы «ФЛАМИНГО», которая, в частности, может применяться и для решения задачи планирования синтеза.

Следует отметить, что, как всякая новая область науки, КС не обладает устоявшейся, универсальной терминологией, что существенно затрудняет обзорное изложение материала. В данном обзоре предпринята

попытка представить основные аспекты КС в более унифицированной терминологии, вобравшей в себя самые удачные и точные, на наш взгляд, формулировки.

II. ПОСТАНОВКА ЗАДАЧИ. «ДЕРЕВО СИНТЕЗА»

Пусть задано некоторое соединение, для которого необходимо найти возможные пути синтеза из других соединений-предшественников. В этом случае мы решаем задачу планирования синтеза в обратном (ретро-синтетическом) направлении [1]. Можно сформулировать и противоположную задачу прямого (синтетического) планирования синтеза, т. е. задачу поиска возможных путей реагирования и продуктов реакций заданного соединения или совокупности соединений [1]. Заданное соединение (набор соединений) будем в дальнейшем называть химической системой (ХС); как правило, в случае ретро-синтетического подхода в ХС входит лишь одно соединение, а при прямом планировании синтеза — одно или несколько соединений.

Результатом работы программы КС как при прямом, так и при обратном направлении поиска является построение «дерева синтеза» [1], т. е. графа, вершины которого соответствуют исходным, промежуточным и конечным соединениям, а ребра изображают химические реакции, в ходе которых осуществляется взаимопревращение этих соединений. Если компьютерный поиск проводится в обратном (ретро-синтетическом) направлении, то «дерево синтеза» представляет собой схематичное изображение возможных путей синтеза заданной ХС из исходных соединений через промежуточное образование предшественников заданной ХС. Если поиск проводится в прямом (синтетическом) направлении, то «дерево синтеза» можно интерпретировать как «дерево реакций», т. е. как схематичное изображение возможных направлений реагирования заданной ХС.

Построение «дерева синтеза» в ходе работы программ КС осуществляется следующим образом. В программу вводится структура соединения, входящего в заданную ХС. Программа генерирует набор предшественников заданной ХС, опираясь на эмпирические или логические инструкции, заложенные в программу. Данная процедура может быть повторена для всех или части структур, генерированных на первой стадии, и т. д., пока не будут выполнены некоторые заданные условия окончания построения «дерева синтеза».

Самым важным и принципиальным моментом в КС является формализация и кодирование человеческого мышления, т. е. тех представлений и рассуждений, которыми оперирует химик-органик при выборе или оценке химических превращений. Все существующие в настоящее время компьютерные системы планирования синтеза представляют собой лишь первые и, вероятно, достаточно робкие шаги в направлении решения этой фундаментальной проблемы. Действительно, пока не создана строгая общая теория органической химии, невозможно не только точное предсказание результата некоторой реакции, но и априорная оценка возможности протекания этой реакции [10]. Тем не менее, попытки воспроизвести с помощью ЭВМ все те логические, эмпирические и интуитивные соображения, которыми руководствуется химик-органик при планировании синтеза, представляются очень полезными как основа для создания системы искусственного интеллекта в органической химии.

Пользуясь введенной терминологией, можно определить задачу выбора и оценки путей синтеза как задачу сокращения «дерева синтеза» [11], т. е. исключения из него химических превращений и соединений, не удовлетворяющих тем или иным критериям отбора. Критерии отбора позволяют значительно сократить «дерево синтеза», а также упорядочить генерацию предшественников заданной ХС таким образом, чтобы в первую очередь программа генерировала самые вероятные или самые интересные пути синтеза. Критерии отбора наиболее эффективны, если они действуют на ранних стадиях построения «дерева синтеза».

Построение «дерева синтеза» может осуществляться в различных режимах: а) неинтерактивно, когда программа после введения некоторой необходимой исходной информации осуществляет поиск без вмешательства пользователя, не требуя дополнительной информации; б) интерактивно, когда программа в определенных ситуациях не может «самостоятельно» принять решение и обращается за помощью к человеку. В идеале программа КС должна генерировать оптимальное «дерево синтеза» без участия пользователя, однако реальная ситуация в теоретической органической химии позволяет решать в автоматическом режиме лишь довольно узкие проблемы или требует вводить в программу очень жесткие ограничения, что неизбежно приводит к потере интересных результатов. В настоящее время КС развивается по пути увеличения «самостоятельности» программ, хотя наиболее распространенные компьютерные системы предусматривают также возможность интерактивного поиска.

III. ОСНОВНЫЕ ПРОГРАММЫ КОМПЬЮТЕРНОГО СИНТЕЗА

Как уже говорилось, цель настоящего обзора — охватить основные аспекты данной научной области как целого. Специфика задачи КС определяет основные проблемы, на которых следует сосредоточить внимание: а) представление и анализ химических структур в программах КС; б) представление и анализ химических реакций (трансформаций) в программах КС; в) критерии отбора в программах КС. Для облегчения понимания изложенного ниже материала мы приводим краткий сводный перечень основных программ КС с указанием руководителей проекта, страны и литературных ссылок (звездочкой отмечены наиболее оригинальные компьютерные разработки).

«АНМОС»* (Automatisierte Heuristische Modellierung Organisch-chemischer Synthesen); Вайзе (ГДР) [12—15].

«ASSOR» (Allgemeines Simulations-System Organischer Reaktionen); Шуберт (ФРГ) [16, 17].

«CAMEO»* (Computer-Assisted Mechanistic Evaluation of Organic reactions); Джорнсон (США) [18—26].

«CASP» (Computer-Assistierten SynthesePlanung); разработка фирмы (ФРГ, Швейцария) [6].

«CHIRP» (Chemical engineering Investigation of Reaction Paths); Агнихотри, Мотард (США) [27].

«CICLOPS» (Computers In Chemistry, Logic Oriented Planung of Syntheses); Уги, Гастайер (ФРГ) [28].

«EROS»* (Erzeugung von Reaktionen für die Organische Synthese); Гастайер (ФРГ) [29, 30].

«FLAMINGOES»* (Formal-Logical Approach to Molecular Interconversions. Non-empirical Generation, Orientation and Evaluation of Syntheses); Зефилов (СССР) [31—34].

«LHASA»* (Logic and Heuristic Applied to Synthetic Analysis); Кори (США) [35—41].

«MASSO»; Морья (Франция) [42].

«MATCHEM» (MAThematical model of constitutional CHEMistry); Уги, Гастайер (ФРГ) [28].

«OCSS» (Organic Chemical Simulation of Synthesis); Кори и Уипке (США) [1].

«PASCOP» (Programme d'Assistance a la Synthese des Composes OrganoPhosphores); Кауфман, Шоплин (Франция) [43—48].

«REACT» Кархарт, Смит, Джераси (США) [49, 50].

«REACT» (REACTion path synthesis program for the petrochemical industry); Говинд, Пауэрс (США) [51].

«SECS»* (Simulation and Evaluation of Chemical Synthesis); Уипке (США) [11].

«SOS» (Simulation of Organic Synthesis); Барон, Шанон (Франция) [52—54].

«SYNCHEM»* (SYNthetic CHEMistry); Гелернтер (США) [55—58].

«SYNGEN»* (SYNthesis GENeration); Хендриксон (США) [59—61].
«TOSCA»* (TOpological SYNthesis design by Computer Application);
Зандер (ФРГ) [62].

Программа Берсона* (Канада) [63—65].

Программа Уитлока (США) [66].

«ЭПОС»* (Эмпирическое Планирование Органического Синтеза);
Коптюг, Смирнов, Пиоттух-Пелецкий (СССР) [67].

Многие из перечисленных программ не являются полностью оригинальными разработками, а представляют собой модификации ранее разработанных программ, специально ориентированные на решение более узких задач, чем программы-прототипы. Так, программа Уитлока, системы «SOS» и «PASCOP» базируются на тех же принципах, что и программы «LHASA» и «SECS», однако предназначены для исследования более узких областей органической химии, таких как синтез ациклических структур, химия гетероциклических соединений и химия фосфорорганических соединений. Программа «SECS» составила основу также и для компьютерной системы «CASP». В дальнейшем изложении мы будем опираться в основном на оригинальные разработки (отмечены звездочкой), опуская непринципиальные модификации программ и устаревшие версии, такие как «OCSS» (см. более новые разработки «LHASA» и «SECS»), «CICLOPS», «MATCHEM» (см. программу «EROS»), «MASSO» (см. программу «SYNGEN»).

Особое место среди перечисленных компьютерных систем занимают программы «REACT», системы «CHIRP», «ASSOR» «AHMOS» и «CAMEO». Программа «REACT» (Кархарт, Смит, Джерасси) является вспомогательной частью системы «CONGEN», ориентированной на установление структуры неизвестного соединения по данным спектроскопических и химических исследований. Система «REACT» (Говинд, Пауэрс) во многом сходна с программами «LHASA» и «EROS», однако предназначена специально для изучения химико-технологических процессов.

Программы «AHMOS», «CAMEO» и «ASSOR» представляют «механистическое» направление в КС. Системы «AHMOS» и «CAMEO» осуществляют поиск исключительно в прямом (синтетическом) направлении, генерируя не пути синтеза химических соединений, а возможные направления реагирования заданной ХС, с учетом механизма превращений и закономерностей физической органической химии. Программа «ASSOR» имеет много общих черт с системой «EROS» (они отпочковались от общего корня — системы «MATCHEM» — «CICLOPS»), однако «ASSOR» отличается более детальное, поэтапное рассмотрение трансформаций, генерируемых в заданной ХС. Машинный поиск в системе «ASSOR», в отличие от программ «AHMOS» и «CAMEO», может осуществляться как в синтетическом, так и в ретро-синтетическом направлении. О системах «FLAMINGOES» («ФЛАМИНГО») и «ЭПОС» см. гл. VIII и VI соответственно.

IV. ПРЕДСТАВЛЕНИЕ И АНАЛИЗ ХИМИЧЕСКИХ СТРУКТУР

Как уже говорилось, функционирование любой программы КС начинается с ввода некоторых начальных данных, основу которых составляет информация о структуре заданной ХС.

Практически все программы КС используют в настоящее время возможности ввода структуры заданной ХС в виде рисунка с помощью специальных графических устройств (графический дисплей, световое перо и т. д.) [68]. Ввод рисунка осуществляется, как правило, в самом привычном для пользователя виде. Полученная графическая информация преобразуется в некоторое внутреннее представление структуры в программе.

Рассмотрим основные принципы кодирования молекулярной структуры в компьютерных программах, поскольку уже на этом этапе можно определить, какие именно аспекты строения молекул являются наиболее важными для решения задач КС.

Любая программа КС в первую очередь учитывает типы атомов в заданной ХС и связи между атомами, в том числе кратности связей. Иначе говоря, общим для всех программ является представление структуры в виде химического мультиграфа и его матрицы смежности [69]. В ряде программ КС для описания структуры химического соединения используются так называемые «таблицы связности» [1, 70], основанные на матрице смежности соответствующего мультиграфа. Таблицы связности могут содержать и дополнительную информацию о заданной ХС, например, описывать стереохимические особенности структуры [11, 24, 40, 71—75], заряды на атомах [18, 39, 40, 53]. В программе «SYNGEN» описание заданной ХС включает только атомы скелета и связи между ними. Кроме того, для каждого атома скелета в таблице связности предусмотрена графа «функциональность», указывающая на характер связей данного атома с теми или иными функциональными группами (ФГ) [61, 76].

Программа «EROS» хранит описание структуры в виде *BE*-матрицы, которая указывает порядки связей между атомами заданной ХС (недиагональные элементы) и число свободных электронов на внешней валентной оболочке каждого атома (диагональ матрицы) [29].

Программа «SYNCHEM» [55] использует как обычное представление структур в виде таблиц связности, так и в виде линейных кодов Висвессера (WLN). Описание заданной ХС в виде таблицы связности используется при анализе и трансформациях структур, в то время как нотация Висвессера удобна для хранения и составления структурных банков данных и библиотек. Кроме того, программа связана с библиотекой доступных веществ, составленной на основе каталога фирмы Aldrich, в котором реагенты описаны именно с помощью нотации Висвессера.

В ходе анализа структур, входящих в состав заданной ХС, программы КС извлекают дополнительную информацию, необходимую для дальнейшей работы, а именно, особенности строения скелета (число и взаимосвязь циклов, цепей), число и вид ФГ, симметрия структуры, стереохимия и т. д. Анализ структуры осуществляется, как правило, в автоматическом режиме, в специальном блоке распознавания [77—87].

В. ЭМПИРИЧЕСКИЙ И НЕЭМПИРИЧЕСКИЙ КОМПЬЮТЕРНЫЙ СИНТЕЗ

Прежде чем перейти к описанию различных способов представления химических реакций в программах КС, следует, отметить, что существуют различные способы доступа к этим формализованным описаниям. Именно на данном этапе можно выделить два принципиально различных подхода: эмпирический КС, когда трансформации заданной ХС осуществляются на основе закодированных сведений об известных органических реакциях, и неэмпирический КС, когда трансформации генерируются логико-комбинаторным путем, без привлечения фактических сведений. В первом случае химические реакции должны быть заранее собраны в виде некоторой библиотеки, во втором случае для поиска трансформаций в заданной ХС используется набор неких логических инструкций и/или комбинаторный алгоритм, непосредственно генерирующий трансформации.

Преимущество эмпирического направления заключается в том, что программа предсказывает, как правило, вполне правдоподобные пути синтеза, а большинство критериев отбора в неявном виде содержатся в описании каждого конкретного превращения. С другой стороны, такие программы привязаны к конкретной библиотеке превращений и не способны предложить принципиально новый синтетический путь или найти новую реакцию. Программы неэмпирического КС лишены этого недостатка, однако требуют включения в программу строгих критериев отбора, чтобы избежать генерирования малореальных или химически неинтересных результатов. В настоящее время многие программы неэмпирического КС включают и взаимодействие с банками данных, обраще-

ние к которым позволяет производить химическую оценку полученных результатов, т. е. позволяет реализовать эмпирический критерий отбора [13, 18, 61].

VI. ПРЕДСТАВЛЕНИЕ И АНАЛИЗ ХИМИЧЕСКИХ РЕАКЦИЙ

Основным этапом при компьютерном планировании синтеза является поиск реакций, осуществляющих превращения в заданной ХС. Существование, что для решения задачи ретро-синтеза необходимо рассматривать ретро-реакции («антитетические реакции» или «трансформации» [1]), т. е. превращения, обратные тем реакциям, в результате которых образуется заданная ХС. В литературе используется различная терминология для обозначения таких превращений. Строго говоря, термин «перераспределение связей» точнее всего отражает формализованный характер представления химических реакций в программах КС, однако в данном обзоре мы будем употреблять термин «трансформация», как наиболее распространенный.

Большинство программ КС относится к эмпирическому направлению и использует в своей работе библиотеки трансформаций (БТ) [1, 11, 40, 44, 50, 51, 53, 57, 64, 66]. Можно выделить следующие общие черты всех БТ. Во-первых, описание трансформации включает перечисление структурных фрагментов, которые должны присутствовать в заданной ХС, для того чтобы данная трансформация могла осуществиться. Во-вторых, описание содержит набор различных тестов, которые определяют принципиальную возможность осуществления данного превращения в конкретной ХС и/или определяют приоритетность («рейтинг») описываемой трансформации. Кроме того, описание включает также перечень структурных изменений, которые нужно произвести, чтобы генерировать структуры, соответствующие результату применения данной трансформации в заданной ХС. Наконец, обязательно присутствует идентификатор каждой трансформации.

Перечисленные аспекты присущи всем БТ, описанным в литературе [11, 40, 55, 88], однако, в зависимости от специфики конкретной программы КС, описание трансформаций может насыщаться дополнительной информацией. Например, в программах «LHASA» и «SECS» каждой трансформации приписывается «характер», т. е. сжатое определение структурных изменений в заданной ХС, которые вызывает данная трансформация (образование или разрыв связи, замыкание или раскрытие цикла, введение, удаление или изменение ФГ и т. д.) [11, 40]. Кроме того, в БТ программ «LHASA» и «SECS» содержатся сведения об условиях реакций (температура и вспомогательные реагенты), а также литературные ссылки. В системе «REACT» (Говинд, Пауэрс) [51] также уделено большое внимание описанию технологических условий для успешного осуществления конкретных процессов.

Описание трансформаций в программах Берсона включает указание выхода соответствующей синтетической реакции и дополнительные тесты, анализирующие корректность получающихся в результате данной трансформации структур [88].

Самыми объемными и высокоорганизованными БТ располагают программы «LHASA» [39—41] и «SECS» [11]. Трансформации в этих библиотеках записаны на специальных языках (CHMTRN в программе «LHASA» и ALCHEM в программе «SECS»), соединяющих в себе строгие логические конструкции алгоритмического машинного языка и понятную химику лексику английского языка.

Гелернтер использует другую форму описания трансформаций, более близкую к машинной интерпретации [55]. Набор тестов, которые должны быть выполнены на структуре заданной ХС, представлен в виде булевых функций [89]. В результате последовательного действия этих функций формируется специальный код, позволяющий программе сделать вывод о применимости трансформации, приписать ей соответствующий рейтинг или объявить о необходимости предварительного обращения к другой трансформации.

Система «ЭПОС» — первая в СССР программа эмпирического КС; разрабатывается в Научно-информационном центре по молекулярной спектроскопии Новосибирского ИОХ СО АН СССР [67]. Программа рассчитана на осуществление ретро-синтетического планирования органического синтеза с использованием библиотеки трансформаций; по форме организации обращения к БТ данная система имеет сходство с программой Гелернтера.

В отличие от программ эмпирического КС, в программах неэмпирического направления трансформации заданной ХС осуществляются не на основе схем перераспределения связей, полученной из БТ, а в результате применения некоторых логических инструкций [13, 16, 18, 27, 28—34, 42, 61, 62].

Программы «EROS» и «TOSCA» используют наборы «генераторов реакций», т. е. инструкций, которые описывают в самом общем виде перераспределение связей в ходе химических реакций [29, 30, 62]. В последней версии программы «EROS» [30] используется пять генераторов реакций, которые в совокупности описывают большинство органических реакций: 1) $X: + I - J \rightarrow I - X - J$; 2) $I - X - J \rightarrow X: + I - J$; 3) $I - J + K - L \rightarrow I - K + J - L$; 4) $I - J + K - L + M - N \rightarrow N - I + J - K + L - M$; 5) $I - J + K - L + X: \rightarrow J: + I - K + L - X$, где I, J, K, L, M, N — реакционные центры, т. е. атомы, связи между которыми изменяют свой порядок на 1. Разрыв «связи» между центрами I и J в генераторе реакции может означать как разрыв, так и уменьшение порядка связи в заданной ХС: аналогично, образование «связи» в генераторе реакции может соответствовать как реальному образованию связи, так и увеличению кратности уже имеющейся связи. Центр X: соответствует атому, изменяющему в ходе реакции свою валентность на 2 единицы, например, карбенному центру в реакции присоединения карбена. Функционирование программы «TOSCA» опирается на генераторы реакций, аналогичные генераторам 3) и 4). Используется также дополнительная трансформация вида $I - J - K + L - M \rightarrow I - K + L - J - M$ [62].

Очевидно, для различных центров и связей в ХС один и тот же генератор реакции будет порождать различные химические превращения. Например, генератор № 3 может соответствовать как присоединению по двойной связи, так и реакции замещения. Таким образом, генераторы реакций в применении к конкретной ХС могут порождать, наряду с хорошо известными превращениями, совершенно новые трансформации.

Описание механизма превращения составляет специфику программы «ASSOR» [16, 17], которая оперирует всего четырьмя трансформациями (основными реакциями). Эти трансформации соответствуют процессам окисления-восстановления ($I^+ + J \rightleftharpoons I + J^+$ и $I^- + J^+ \rightleftharpoons I + J^-$), и процессам диссоциации-ассоциации ($I - J \rightleftharpoons I^+ + J^-$ и $I - J \rightleftharpoons I^- + J^+$), представленным в самом общем виде. Такое представление трансформаций во многом аналогично генераторам реакций в программе «EROS», однако система «ASSOR» больше внимания уделяет механизму превращения, что ведет к увеличению числа стадий КС. Например, результату действия генератора № 3 (одна стадия машинного поиска в программе «EROS») будут соответствовать четыре стадии поиска в программе «ASSOR».

Система «АНМОС» [12, 13] использует формализованное представление органических реакций, также во многом аналогичное представлению реакций в программе «EROS». Основные отличия заключаются в том, что, во-первых, трансформации в программе «АНМОС» более конкретны, чем генераторы реакций, поскольку содержат указание на тип реакционного центра (электрофил, нуклеофил и т. д.). Во-вторых, трансформации в программе «АНМОС» соответствуют именно элементарным стадиям органических реакций, а не суммарному результату процесса. Вайзе [13] рассматривает трансформации, соответствующие следующим процессам:

- 1) присоединение: $E + Nu \rightarrow E - Nu^+$
- 2) замещение: $EFABG - NFABG + Nu \rightarrow EFABG - Nu^+ + -NFABG$

- 3) диссоциация: $\text{EFABG} - \text{NFABG} \rightarrow {}^+\text{EFABG} + {}^-\text{NFABG}$
- 4) протонирование: $\text{X}-\text{H} + \text{Nu} \rightarrow \text{X}^- + \text{Nu}^+ - \text{H}$
- 5) секстетная перегруппировка: $\text{Y}-\text{X}-\text{Z} \rightarrow \text{X}^+ - \text{Z}^- - \text{Y}$
- 6) поляризация: $\text{X}^- - \text{Z} \rightarrow \text{X} = \text{Z}^-$ или $\text{X}^- - \text{Z} = \text{Y}^+ \rightarrow \text{X} = \text{Z} - \text{Y}$
- 7) электрофильное замещение: $\text{E} + \text{H}-\text{C}(\text{Ar}) \rightarrow \text{E}^- - \text{C}(\text{Ar}) + \text{H}^+$
- 8) элиминирование: $\text{Nu} + \text{H}-\text{X}-\text{EFABG} - \text{NFABG} \rightarrow \text{X} = \text{EFABG} + \text{H} - \text{Nu}^+ + {}^-\text{NFABG}$

Здесь E — электрофил, Nu — нуклеофил, EFABG и NFABG — соответственно электрофугная и нуклеофугная уходящие группы, X, Y, Z — другие реакционные центры (Z — секстетный атом).

Программа неэмпирического КС «ФЛАМИНГО» [31–34] может оперировать как многостадийными химическими реакциями (произвольной степени сложности), так и отдельными стадиями механизма. Универсальность компьютерной системы «ФЛАМИНГО» по сравнению с другими программами неэмпирического КС достигается за счет использования не отдельного, весьма ограниченного набора генераторов реакций, а общего комбинаторного алгоритма, в ходе которого могут быть генерированы самые различные химические реакции и/или стадии реакций, в том числе совершенно новые или малоизученные процессы (подробнее о системе «ФЛАМИНГО» см. гл. VIII).

В основу программы «SYNGEN» [60, 61] положена оригинальная классификация органических реакций по Хендриксону [90–95]. Эта система основана на утверждении о том, что большинство органических реакций можно представить как изменение непосредственного окружения при одном или нескольких атомах углерода при переходе от исходных соединений к продуктам (или, при ретро-синтезе, от заданной ХС к предшественникам). Все многообразие возможных окружений атома углерода предлагается выразить четырьмя типами связей:

- 1) тип H: σ -связь с атомом водорода или атомом, менее электроотрицательным, чем водород;
- 2) тип R: σ -связь с другим атомом углерода;
- 3) тип П: π -связь с другим атомом углерода;
- 4) тип Z: σ - или π -связь с атомом, более электроотрицательным, чем водород.

Число связей каждого типа обозначается соответственно h , σ , π , z ; очевидно, что $h + \sigma + \pi + z = 4$. Сумму $f = \pi + z$ Хендриксон называет функциональностью атома углерода. В данной системе любая реакция будет определена, если будет указано изменение окружения на всех атомах углерода при переходе от одной ХС к другой. Для отдельного атома углерода это изменение обозначается двумя буквами: первая характеризует тип образующейся связи, вторая — тип разрывающейся связи. Таким образом, всего возможно $4 \cdot 4 = 16$ типов структурных изменений на атоме углерода.

Хендриксон разделяет все множество органических реакций на три класса. К первому классу относятся конструктивные реакции, т. е. реакции, в результате которых происходит изменение углеродного скелета органического соединения. Конструктивные реакции включают три типа структурных изменений: RH, RP, RZ. Ко второму классу относятся реакции перефункционализации (введение, удаление или изменение ФГ), включающие 9 типов структурных изменений: HH, HZ, ZH, ZZ, PH, PZ, HP, ZP, PP. Остальные 4 типа изменения окружения на атомах углерода относятся к деструктивным реакциям, которые не используются в программе «SYNGEN».

Из всех конструктивных реакций программа «SYNGEN» использует только такие, в ходе которых образуется одна σ -связь C—C, а изменение окружения происходит не более, чем на трех атомах углерода в цепи по обе стороны от образующейся связи. Для всех таких конструктивных реакций, а также для реакций перефункционализации в программе хранятся в закодированном виде списки инструкций, которые обеспечивают выполнение необходимых трансформаций в заданной ХС; эти инструкции также называются полуреакциями. С одной стороны, составление

таких инструкций не связано с использованием фактического материала органической химии, но в то же время набор возможных полуреакций хранится в памяти ЭВМ подобно библиотекам трансформаций в программах эмпирического КС. Следовательно, подход Хендриксона является в некоторой степени промежуточным между эмпирическим и неэмпирическим КС.

VII. КРИТЕРИИ ОТБОРА

В предыдущих главах было показано, как формализуется и кодируется химическая информация в известных компьютерных системах. Однако, как уже говорилось, самая важная проблема в КС — это создание и формализация критериев отбора, оперирование которыми как раз и придает программам КС черты «искусственного интеллекта».

В программах эмпирического КС критерии отбора могут использоваться на трех основных стадиях поиска: 1) выбор определенной «стратегии синтеза», которая ограничивает число и вид превращений, доступных в БТ; 2) оценка или отбраковка конкретных трансформаций до их применения в заданной ХС; 3) оценка или отбраковка конкретных предшественников заданной ХС, полученных в результате трансформации.

«Стратегии синтеза», которые позволяют направить машинный поиск в заданном направлении, получили наибольшее развитие в программах «LHASA» и «SECS». Перечислим основные типы стратегий [11, 35, 39—41].

1. «Стратегия применимости», т. е. поиск в БТ всех превращений, которые применимы к заданной ХС. Основной недостаток этой стратегии заключается в том, что большие классы трансформаций, такие как введение, удаление или изменение ФГ, применимы почти всегда, что приводит к значительному разрастанию «дерева синтеза» [11, 35].

2. Стратегии, ориентированные на трансформации, т. е. поиск в БТ известных, достаточно надежных реакций, с помощью которых можно осуществить элегантный синтез заданной ХС [41]. К числу таких реакций относятся, например, следующие процессы: реакция Дильса — Альдера, аннелирование по Робинсону, альдольная циклизация, сигматропные перегруппировки. Другая разновидность стратегии, ориентированной на трансформации, заключается в построении с помощью простейших, «механистических» трансформаций (присоединение протона, снятие протона и т. д.) возможных механизмов образования заданной ХС [41].

3. Структурно-ориентированные стратегии, т. е. поиск трансформаций, приводящих именно к таким предшественникам заданной ХС, которые содержат определенные структурные фрагменты [11, 41].

4. Топологические стратегии, т. е. поиск трансформаций, приводящих к разрыву одной или нескольких «стратегических» связей, в результате которого может произойти значительное упрощение структуры исходной ХС [11, 41, 96].

5. Стереохимические стратегии, т. е. поиск трансформаций, с помощью которых можно осуществить синтез заданной ХС с определенной конфигурацией всех стереоцентров [41].

6. Стратегии, ориентированные на функциональные группы, т. е. поиск трансформаций, осуществляющих введение, удаление, защиту или изменение ФГ. Обычно данная стратегия предвещает применение других стратегий, непосредственное применение которых часто бывает затруднено именно из-за наличия в заданной ХС конкретных функциональных групп [11, 41, 97—99].

Выбор определенной стратегии может осуществляться автоматически или с участием пользователя. Применение нескольких стратегий позволяет, как правило, найти весьма эффективные и элегантные пути синтеза [40, 41].

В программах Берсона [63, 64] и Гелсритера [55—57] обращение к БТ осуществляется на основе данных, получаемых в ходе анализа структуры заданной ХС, иначе говоря, воплощается «стратегия применимости». В программе Берсона применение этой стратегии строго упорядо-

чено. Во-первых, все структурные фрагменты, найденные ЭВМ в блоке распознавания, получают различные рейтинги. Наивысший рейтинг имеют структурные фрагменты, при образовании которых происходит одновременно замыкание наибольшего числа связей, как, например, при образовании трехчленного цикла в реакции Симмонса — Смита и шестичленного цикла в реакции Дильса — Альдера.

Во-вторых, превращения в каждом разделе БТ также расположены в порядке уменьшения приоритетности соответствующих синтетических реакций. Наивысший рейтинг имеют реакции введения ФГ; затем следуют реакции построения молекулярного скелета, реакции изомеризации, реакции введения защиты ФГ, удаления защиты ФГ, изменения ФГ, удаления ФГ. Реакции фрагментации имеют самый низкий рейтинг, поскольку соответствующие им трансформации усложняют структуру заданной ХС [64].

Один из самых принципиальных и нерешенных до настоящего времени вопросов органической химии — вопрос априорной оценки вероятности протекания реакции — решается в программах КС с помощью формальных и эмпирических критериев отбора. Трансформация считается формально возможной, если в заданной ХС присутствует структурный фрагмент, формально необходимый для осуществления данной трансформации. Эмпирические критерии отбора реализуются на основе более глубокого анализа структуры заданной ХС, условий реакции, реагентов и т. д. Окончательная оценка (рейтинг) трансформации определяется в результате выполнения ряда тестов, содержащихся, как правило, в самом описании трансформации [11, 39, 40, 55, 88]. Именно на этой стадии наиболее активно используется информация, полученная в блоке распознавания. В целях более детального анализа превращений может привлекаться и дополнительная информация.

Дополнительные критерии отбора наиболее развиты в программе «SECS» [11]. Так, применимость того или иного превращения из БТ определяется не только по результатам тестов, содержащихся в описании трансформации, но и в соответствии с заданным «списком целей», который содержит указания о требуемых изменениях скелета и ФГ. В соответствии с этим списком происходит отбраковка превращений из БТ, если «характер» трансформации не соответствует ни одной из «целей», перечисленных в списке. Если превращение удовлетворяет данному критерию, но структурный фрагмент, указанный в описании трансформации, отличается от структурного фрагмента заданной ХС лишь наличием и/или видом ФГ, то перед непосредственным применением этой трансформации программа «SECS» проводит необходимую модификацию структуры (в программе «LHASA» аналогично действует стратегия, ориентированная на ФГ).

Уже в первой версии программы «SECS» большое внимание было уделено учету стерических и электронных эффектов при оценке конкретного превращения. Подпрограмма «SYMIN» конструирует в памяти ЭВМ трехмерную модель заданной ХС, используя расчеты по методу молекулярной механики. Для оценки вероятности реакций с участием сопряженных систем используются расчеты молекулярных орбиталей по методу Хюккеля. В зависимости от полученных результатов может измениться рейтинг трансформации, выбранной из БТ.

Важным критерием применимости выбранного превращения является проверка, не вызовут ли условия, благоприятные для протекания соответствующей синтетической реакции, нежелательных изменений в заданной ХС. Программа «SECS» анализирует чувствительность различных ФГ в условиях данной реакции, подбирает защитные реагенты. Если условия реакции и строение заданной ХС не позволяют осуществить удовлетворительную защиту ФГ, трансформация считается неприемлемой.

В отсутствие информации о симметрии заданной ХС применение одной и той же трансформации к структуре, обладающей несколькими элементами симметрии, неизбежно приведет к дублированию результатов.

Программа «SECS» определяет полную группу симметрии трехмерной модели заданной ХС, что позволяет генерировать в дальнейшем только неэквивалентные результаты.

Программы Берсона [63, 64] и Гелернтера [55—57] работают неинтерактивно, генерируя без вмешательства пользователя пути синтеза заданной ХС. Следовательно, критерии отбора, заложенные в эти программы, должны быть более строго формализованы по сравнению с критериями в программах «LHASA» и «SECS». Так, в программе Берсона трансформации внутри одного класса приоритетности оцениваются специальным рейтингом, который тем выше, чем выше выход соответствующей синтетической реакции [88].

В более ранних работах Берсон и сотр. [63] использовали дополнительные критерии для оценки отдельных стадий синтеза. Каждая трансформация получала свой рейтинг, который характеризовался в условных единицах отношением «упрощение»/«стоимость». Значение величины «упрощение» увеличивалось, если в результате трансформации уменьшалось число циклов, число ФГ и неводородных атомов. «Стоимость» зависела от многих факторов, в том числе и от выхода реакции.

В программе «SYNCHM» [55—57] каждая трансформация строго оценивается в соответствии с набором тестов в виде булевых функций. Эти тесты воплощают многие из эвристических правил, которыми руководствуется химик при оценке химической реакции. На основании результатов этих тестов начальный рейтинг трансформации может быть изменен (например, повышен, если присутствует сопряженная активирующая группа, или понижен, если осуществлению трансформации мешают стерические препятствия). В результате последовательного действия булевых функций может быть частично изменена и схема самой трансформации. Например, если обнаружена ФГ, чувствительная к реагенту, используемому в ходе реакции, то программа может ввести другой реагент или обратиться к процедуре защиты ФГ.

Наряду с оценкой трансформаций, в программах эмпирического КС проводится и оценка самих предшественников. Например, в программе «SECS» [11] отбраковка вариантов производится с помощью следующих структурных критериев отбора: нарушение правила Бредта; наличие антиароматической системы; наличие кумулированной системы; тройная связь в малом цикле; *транс*-олефин в малом цикле; циклическая система с *транс*-расположенным мостиком; *транс*-конденсированные 3-членные циклы; неправильное значение валентности атома; два одинаковых по знаку заряда на разных атомах; нестабильная комбинация ФГ. Аналогичные критерии отбора структур применяются в большинстве программ КС, как эмпирического, так и неэмпирического направления.

Помимо критериев, проверяющих формальную корректность и «химический смысл» генерированных структур, в ряде неинтерактивных программ КС, работающих в направлении ретро-синтеза [55, 56, 60, 61, 63, 64], используется очень эффективный критерий, определяющий окончание построения цепочки синтеза заданной ХС. Для реализации этого критерия программа должна быть дополнена специальным банком данных — библиотекой (каталогом) доступных веществ (БДВ). Если структура очередного предшественника, генерированного в ходе машинного поиска, совпадает с одной из структур, хранящихся в БДВ, построение одного из возможных путей синтеза считается законченным. Учитывая, что реакции введения, удаления или замещения ФГ играют второстепенную роль по сравнению с реакциями построения или модификации скелета, можно считать совпадение с точностью до числа и вида ФГ достаточным для удовлетворения данного критерия [60, 61].

Основная проблема, с которой сталкиваются программы неэмпирического КС, заключается в том, чтобы отобрать из всего множества формально возможных путей синтеза наиболее вероятные и интересные с химической точки зрения. В программах неэмпирического направления можно выделить следующие подпроблемы, для решения которых применяются критерии отбора: ограничение типов трансформаций, генерируе-

мых в конкретной заданной ХС; ограничение применимости трансформаций в заданной ХС; оценка или отбраковка генерированных предшественников заданной ХС. Критерии отбора предшественников [18, 62] во многом сходны с аналогичными критериями в программах эмпирического КС, поэтому центр тяжести обсуждения мы перенесем на специфику критериев отбора первых двух типов в программах неэмпирического КС.

Возможность выбора типов трансформаций, которые могут генерироваться в заданной ХС, в явном виде присутствует в программе «EROS» [29, 30], где, во-первых, предусмотрена ситуация, когда пользователь исключает из списка генераторов реакций, хранящихся в памяти ЭВМ, генераторы, отвечающие тем процессам, которые, по его мнению, не могут протекать в заданной ХС. По умолчанию пользователя принимается, что все пять генераторов реакций могут применяться для поиска предшественников заданной ХС. Во-вторых, имеется возможность наложить определенные ограничения на вид и размер «дерева синтеза» [30]. В частности, указать максимальное число уровней «дерева синтеза» (число стадий КС), число предшественников на каждой стадии, максимальное число предшественников, генерируемых для одной ХС-, максимальное число предшественников, которые будут рассматриваться как заданные ХС на следующей стадии КС.

Основной принцип, по которому осуществляется выбор типов трансформаций в программе «SYNGEN» [60, 61], заключается в том, что планирование синтеза должно опираться в первую очередь на конструктивные реакции, т. е. реакции, в ходе которых формируется скелет заданной ХС. Реакции перефункционализации, в ходе которых изменяются ФГ, но не затрагивается скелет молекулы, рассматриваются как вспомогательные.

Как было показано в гл. VI, описание любой трансформации в программах неэмпирического КС в общем виде можно представить как набор реакционных центров с указанием, каким образом перераспределяются связи между ними. Следовательно, критерий отбора, который может ограничить применимость тех или иных трансформаций, заключается в выборе потенциальных реакционных центров или реакционноспособных связей из всего множества атомов и связей в заданной ХС.

В программе «EROS» [29, 30] данный критерий реализуется при анализе структуры заданной ХС, в ходе которого определяются связи, могущие, в принципе, разрываться в ходе дальнейших превращений. Распознавание таких связей может осуществляться как в автоматическом режиме, так и с участием пользователя. В автоматическом режиме реакционноспособными объявляются кратные связи, связи C—X, N—X, X—X (X — гетероатом), а также смежные с ними связи. Ароматические связи не считаются реакционноспособными. Пользователь может произвольным образом скорректировать список реакционноспособных связей, первоначально сформированный ЭВМ. Кроме того, пользователь может задать оптимальное значение энтальпии реакции и диапазон допустимых значений энтальпии. Для каждого превращения, генерированного ЭВМ, значение энтальпии вычисляется по схеме Аллена [100—103]. Авторы [103] утверждают, что такой, в общем-то нестрогий, подход дает удовлетворительное совпадение с экспериментальными данными и может служить для грубой оценки вероятности протекания реакции. Если вычисленное значение энтальпии реакции попадает в диапазон допустимых значений, превращение считается вероятным; в противном случае найденный ЭВМ предшественник отбраковывается. Заданное оптимальное значение энтальпии позволяет более детально оценить вероятность протекания процесса, предсказанного ЭВМ. Чем ближе вычисленное значение энтальпии реакции к заданной величине, тем выше рейтинг трансформации, и соответствующие предшественники будут в первую очередь исследоваться на следующей стадии КС. Отметим, что родственная системе «EROS» программа «CHIRP» использует вычисленные значения свободной энергии Гиббса для оценки вероятности протекания трансформации в заданной ХС [27].

Гастайгером и сотр. ведутся также работы по созданию неэмпирических критериев выбора реакционноспособных связей, критериев, основанных на количественном определении частичных зарядов на атомах [104, 105], индуктивных [106], резонансных эффектов [107], дипольных моментов [107] и эффектов поляризуемости [108]. В работах [104, 105] описаны итерационные процедуры, которые позволяют определить распределение частичных зарядов в простых малоатомных соединениях, содержащих гетероатомы. На основе полученных данных можно проводить вычисления энергий связи C—X (X—гетероатом) [105], pK_a [105], $J(^{13}\text{C}-\text{H})$ [109], сродства к протону [106, 108, 110, 111]. Можно надеяться, что в дальнейшем своем развитии метод частичного уравнивания орбитальной электроотрицательности [105, 112] действительно внесет существенный вклад в создание общей теории реакционной способности органических соединений, однако в настоящее время выбор реакционноспособных связей по топологическим и структурным характеристикам является более общим, более быстрым и надежным методом.

В программе «TOSCA» [62] используется интересная и оригинальная методика ограничения применимости трансформаций в заданной ХС. Как и программы эмпирического КС, система «TOSCA» в своей работе опирается на выбранную стратегию синтеза. Программа использует две стратегии: консонантную и диссонантную. Консонантная стратегия заключается в поиске реакций нуклеофильного и электрофильного замещения и присоединения. Диссонантная стратегия включает поиск процессов окисления, восстановления, реакций, приводящих к изменению полярности. В отличие от программ эмпирического направления, в программе «TOSCA» выбор стратегии осуществляется именно на этапе выбора потенциальных реакционных центров и реакционноспособных связей. Все атомы, входящие в заданную ХС, начиная с гетероатомов, помечаются знаком «+» или «-» в соответствии с их электрофильными или нуклеофильными свойствами. Если в заданной ХС все связи — консонантные, т. е. нет ни одной пары смежных атомов с одинаковыми зарядами, предпочтение отдается консонантной стратегии.

Для заданной ХС, содержащей диссонантные связи (смежные атомы с одинаковыми знаками), может быть выбрана как диссонантная, так и консонантная стратегия. При консонантной стратегии на действие трансформаций накладывается ограничение: связи могут образовываться только между атомами с противоположными знаками, а диссонантные связи объявляются нереакционноспособными. Диссонантная стратегия, напротив, допускает комбинацию атомов с одинаковыми знаками.

Логика программы «SYNGEN» [60, 61] подчинена идее «идеального синтеза» как последовательности конструктивных реакций [10]. Поэтому в качестве реакционноспособных связей выбираются исключительно связи в скелетах молекул, входящих в заданную ХС. Сначала программа «SYNGEN» осуществляет поиск всех возможных путей разбиения скелета на составные части — синтоны. В общем случае число таких разбиений может быть очень велико, поэтому необходимы дополнительные критерии оценки эффективности того или иного набора синтонов для синтеза заданной ХС. Хендриксон, опираясь на представление об «идеальном синтезе», вводит критерий конвергентности [10, 60, 95]. В конвергентном синтезе фрагменты целевой структуры формируются независимо и объединяются лишь в конце синтеза. В соответствии с этим требованием, расщепление скелета заданной ХС на первой стадии должно приводить к двум примерно одинаковым по числу атомов синтонам, на второй стадии — к четырем, на третьей — к восьми и т. д.

В ходе однократного разбиения скелета допускается разрыв одной-двух связей и размыкание максимум одного цикла. Предпочтение отдается разбиениям, для которых процесс обратной «сборки» молекулы описывается эффективными синтетическими реакциями, например, реакцией Дильса — Альдера. Необходимо выделять ситуацию, когда синтоны, полученные в результате расщепления скелета, оказались одинаковыми, так как в этом случае схема синтеза значительно упрощается. Кроме

совпадения или сходства синтонов между собой проверяется и наличие каждого из найденных синтонов в БДВ. Для синтонов, которые отсутствуют в БДВ, операция разбиения повторяется [60, 61].

Каждая последовательность разбиений скелета заданной ХС, приводящая к синтонам, для которых найдены соответствия в БДВ, порождает так называемый «упорядоченный набор связей» («ordered bondset») [60]. Этот набор состоит из связей, которые нужно в определенном порядке размыкать, чтобы получить заданное разбиение на синтоны. В том же порядке поочередно рассматриваются связи, входящие в данный набор, и к каждой связи применяются все подходящие (исходя из наличия в заданной ХС конкретных ФГ) конструктивные полуреакции, или, если это необходимо, реакции перефункционализации. В результате становятся известны не только скелеты исходных соединений, но и функциональные группы, присутствующие в этих соединениях.

Как уже говорилось, особое место среди программ КС занимают системы «АНМОС» [12—15] и «САМЕО» [18—26]. По характеру представления структур и трансформаций эти программы близки к системам «ЕРОС» и «LHASA» соответственно. Программу «АНМОС» принято относить к неэмпирическому направлению КС, в то время как система «САМЕО», опирающаяся на несколько небольших банков данных, занимает промежуточное положение между неэмпирическим и эмпирическим направлениями. Мы вынесли специальное обсуждение программ «АНМОС» и «САМЕО» в главу, посвященную критериям отбора, поскольку именно оригинальные критерии, основанные на закономерностях физической органической химии, составляют неотъемлемую особенность этих программ.

В основу системы критериев отбора, которую использует программа «АНМОС» [12, 13], положена классификация кислот и оснований по Пирсону. Вайзе рассматривает шесть типов реакционных центров: жесткий электрофил, мягкий электрофил, жесткий нуклеофил, мягкий нуклеофил, нуклеофуг и электрофуг. В программе имеется небольшой банк данных, в котором хранятся индексы реакционной способности различных ФГ, количественно отражающие нуклеофильные и электрофильные свойства этих групп. На основе введенной шкалы оценок, все ФГ, найденные программой в блоке распознавания, получают свой рейтинг — индекс реакционной способности. В дальнейшем программа использует эти рейтинги для вычисления другой количественной характеристики — индекса «тенденции к реагированию» для различных способов взаимодействия электрофилов и нуклеофилов, предусмотренных программой (см. гл. VI). Полученные значения сравниваются с заданными пороговыми значениями и таким образом определяются трансформации, осуществление которых в заданной ХС наиболее предпочтительно. Программа осуществляет приоритетные трансформации, генерируя тем самым возможные продукты реагирования заданной ХС.

Критерии для оценки и отбора возможных механизмов органической реакции особенно развиты в работах Джоргенсена и сотр. [18—26]. На основе этих критериев создана программа «САМЕО» для предсказания продуктов реакций, если заданы исходные вещества и условия реакции. Следует отметить, что программа осуществляет поиск в прямом (синтетическом) направлении, и предназначена не столько для генерации целых последовательностей реакций, сколько для того, чтобы оценивать возможность протекания одной и той же реакции в разных направлениях. Программа имеет несколько блоков, каждый из которых ориентирован на работу с определенным классом реакций. В целом программа «САМЕО» охватывает следующие процессы: взаимодействие электрофилов и нуклеофилов в условиях основного катализа [18], включая некоторые реакции металлоорганических соединений [19], реакции кремнийорганических соединений [19], перициклические реакции [21, 22], нуклеофильное и электрофильное замещение в ароматических соединениях [23—25]. Для каждого из этих классов реакций выработаны спе-

циальные критерии, оценивающие вероятность протекания процесса в том или ином направлении.

Рассмотрим подробнее критерии отбора трансформаций, интерпретируемых в программе «САМЕО» как реакции электрофилов и нуклеофилов в условиях основного катализа [18]. Выбор потенциальных реакционных центров осуществляется следующим образом. С помощью небольшого банка данных и специального алгоритма [26] программа определяет самые кислые протоны и генерирует трансформации, соответствующие переносу протонов к основанию. Атомы, с которыми были связаны протоны, рассматриваются как потенциальные реакционные центры — нуклеофилы. Для металлоорганических соединений [19] учитывается также возможность протекания конкурирующих реакций присоединения и обмена. Доминирующими считаются процессы, приводящие к более слабому основанию.

Потенциальные электрофильные центры определяются по структурным характеристикам (например, кратная связь), а также по значению pK_a уходящей группы, с учетом заданных условий реакции. С помощью набора эвристических правил и данных о структуре реагентов и характере ФГ, полученных в блоке распознавания, определяется возможность протекания конкурирующих процессов присоединения, замещения и элиминирования с участием нуклеофильных и электрофильных реакционных центров.

В дальнейшем этот блок программы «САМЕО» претерпел существенные изменения и в настоящее время включает также распознавание и исследование реакций с участием кремнийорганических реагентов [20], винильных и ароматических электрофилов [23].

Функционирование блока для исследования реакций электрофильного замещения в ароматических системах [25] основано на определении потенциальных реакционных центров, участвующих в процессе. Самой важной процедурой в этом блоке является вычисление относительной реакционной способности различных центров в ароматической системе. В зависимости от условий реакции и типа ароматической системы (замещенный бензол, гетероцикл, полициклическая система) ей присваивается начальный индекс относительной реакционной способности. Используя данные, полученные при анализе заданной ХС в блоке распознавания, и набор эвристических правил, программа учитывает влияние заместителей. Было выведено уравнение для вычисления поправки к начальному индексу, в зависимости от числа и вида электронодонорных и электроноакцепторных заместителей. После вычисления индекса относительной реакционной способности ароматических систем в целом, программа приступает к определению различных реакционных центров в каждой системе, руководствуясь набором эвристических правил. Учитывается также стерический эффект заместителей.

При исследовании перициклических реакций [21] большую роль играют стереохимические критерии отбора. Так, диены считаются не-реакционноспособными, если стерические препятствия не позволяют им достичь *S-цис*-конфигурации. Кроме стереохимических критериев, для отбора реакционноспособных диенов, диснофилов, 1,3-диполей и дипольрофилов активно используются различные эвристические правила. Основным критерий отбора перициклических реакций реализуется при сравнении энергий граничных орбиталей. Считается, что чем меньше разница этих энергий, тем легче протекает циклоприсоединение. Кроме того, по результатам расчета энергий граничных орбиталей программа делает предсказания относительно регио- и стереоселективности процесса. Определение энергий осуществляется на основе значений потенциалов ионизации, взятых из специального банка данных программы «САМЕО».

В дальнейшем в программу был включен анализ реакций [2+2]-циклоприсоединения, [2,3]- и [3,3]-сигматропных перегруппировок, [1,*j*]-сдвигов σ -связей ($j=3, 5, 7$) [22]. Для оценки приоритетности того или иного направления реакции использовались величины энергии актива-

ции, вычисленные на основе анализа большого количества экспериментальных кинетических данных для реакций перечисленных типов. Вопреки ожиданиям авторов [21], оказалось, что огромное разнообразие скоростей перидиклических реакций, влияние на скорость реакции слишком многих факторов, — все эти причины не позволяют в настоящее время выработать общие принципы для оценки вероятности протекания перидиклических реакций.

VIII. УНИВЕРСАЛЬНАЯ КОМПЬЮТЕРНАЯ СИСТЕМА «ФЛАМИНГО» ДЛЯ РЕШЕНИЯ ПРОБЛЕМ СИНТЕЗА И УСТАНОВЛЕНИЯ МЕХАНИЗМОВ РЕАКЦИЙ В ОРГАНИЧЕСКОЙ ХИМИИ

Компьютерная система «ФЛАМИНГО» («FLAMINGOES» — Formal-Logical Approach to Molecular Interconversions. Non-Empirical Generation, Orientation and Evaluation of Syntheses) представляет собой оригинальную разработку в области применения ЭВМ в органической химии, созданную авторами настоящего обзора [31—34]. Для формализации химической информации, используемой в данной системе, применяется формально-логический подход, разработанный ранее Зефиоровым и Трачом [113—119]. Обобщенное рассмотрение химических процессов, основанное на этом подходе, позволило создать данную универсальную компьютерную систему, пригодную для решения любых проблем органической химии, формализуемых как конструктивное перечисление перераспределений связей в химических структурах. Основные типы задач, для решения которых может применяться система «ФЛАМИНГО» — это, во-первых, компьютерный синтез, во-вторых, изучение перегруппировочных процессов, в-третьих, предсказание механизмов сложных многоступенчатых реакций.

Кратко рассмотрим представление и анализ химических реакций в системе «ФЛАМИНГО». Наиболее важные типы органических реакций формально описываются как результат циклического перераспределения связей (ЦПС) в исходной ХС, в результате которого образуется конечная ХС [113]. Следовательно, альтернативным подходом к решению задачи КС (по сравнению с использованием БТ или набора генераторов реакций) является систематическая неповторная генерация всех возможных ЦПС в заданной ХС, которая позволяет найти полный набор возможных предшественников заданной структуры (в случае ретро-синтеза) или полный набор продуктов возможных реакций в исходной ХС (в случае прямого КС). Если полученные результаты представляются перспективными для дальнейшего исследования, процедура генерации трансформаций (в виде ЦПС) может быть распространена на несколько стадий.

Система «ФЛАМИНГО», основанная на систематической неповторной генерации ЦПС, не имеет прямых аналогий среди зарубежных разработок. Сохраняя все достоинства неэмпирического подхода (генерация трансформаций независимо от банков данных, возможность оперирования новыми или малоизученными реакциями), система «ФЛАМИНГО» по широте охвата химического материала приближается к программам эмпирического направления, поскольку ~90% известных химических реакций можно описать как процессы с ЦПС, и, следовательно, данные процессы могут быть генерированы системой «ФЛАМИНГО» в ходе решения задач компьютерного синтеза.

Рассмотрим основные типы критериев отбора, используемых в системе «ФЛАМИНГО». Во-первых, введен формальный критерий, проверяющий соблюдение правильных значений валентности атомов и порядков связей.

Во-вторых, программа использует следующий эмпирический критерий: для того, чтобы ограничить число формально корректных, но химически неинтересных решений, система «ФЛАМИНГО» может быть ориентирована на генерацию не всех ЦПС, возможных в заданной ХС, но

только тех, которые отвечают заданным ограничениям на размер ЦПС, а также ограничениям на число образующихся связей и число разрывающихся связей. Система «ФЛАМИНГО» является интерактивной, и химик может ввести эти ограничения в программу в режиме диалога. Следует отметить, что допустимые размеры ЦПС, а также числа образующихся и разрывающихся связей определяют набор допустимых трансформаций в заданной ХС [113].

В-третьих, компьютерная система «ФЛАМИНГО» использует критерий эквивалентности. Генерация ЦПС в химической системе, для которой соответствующий граф обладает несколькими автоморфизмами, неизбежно приводит к повторяющимся результатам. Действие критерия эквивалентности организовано в системе «ФЛАМИНГО» таким образом, что позволяет избежать дублирования результатов еще на стадии генерации ЦПС; для реализации данного критерия отбора используется информация о группе автоморфизмов графа, соответствующего исходной ХС. Следует подчеркнуть, что в программе «ФЛАМИНГО» все наиболее важные критерии отбора действуют на максимально ранних стадиях генерации, задолго до построения конечных ХС в явном виде.

Система программ «ФЛАМИНГО» написана на языке БЭЙСИК и ориентирована на широкодоступный отечественный персональный компьютер «Искра-226» (64 К). Настоящая версия системы «ФЛАМИНГО» включает ведущую программу, девять основных подпрограмм и занимает 60 КБайт компьютерной памяти. Среднее время генерации составляет 3—5 с на результат. Рассмотрим основные программные блоки, составляющие систему «ФЛАМИНГО».

1. Ввод исходной ХС. В данном блоке химик вводит структуру исходной ХС в виде рисунка на графическом дисплее. Для представления структуры в программе используется матрица смежности мультиграфа, соответствующего исходной ХС. В данном блоке пользователь задает также число стадий компьютерного синтеза.

2. Построение группы автоморфизмов. В ходе выполнения данной подпрограммы автоматически генерируются все подстановки группы автоморфизмов графа, соответствующего исходной ХС. Информация о симметрии графа необходима для генерации только неэквивалентных ЦПС, т. е. для реализации соответствующего критерия отбора. Блок построения группы автоморфизмов основан на фортранном комплексе «АВТОГРАФ» [120].

3. Ввод информации, необходимой для реализации эмпирических критериев отбора:

а) ограничение применимости трансформаций в заданной ХС (ср. с. 1364): в настоящей версии программы «ФЛАМИНГО» гетероатомы и атомы, связанные кратной связью, автоматически объявляются как потенциальные реакционные центры. Затем химику предоставляется возможность в режиме диалога произвольным образом скорректировать набор центров, предложенный ЭВМ. Кроме того, химик выбирает из множества потенциальных реакционных центров те центры, которые могут изменять свою валентность. Для каждого из выбранных центров необходимо указать допустимый вариант изменения валентности;

б) ограничение типов трансформаций, генерируемых в заданной ХС (ср. с. 1364). Химик имеет возможность ввести ограничения на допустимые размеры ЦПС и на числа образующихся и разрывающихся связей.

Следует отметить, что в программе предусмотрен режим «поиск без ограничений», в ходе которого осуществляется генерация всех формально возможных ЦПС (с размером цикла от 3- до 8-членного). Такой режим позволяет выявить все принципиально возможные стратегические пути синтеза заданного соединения, в особенности для малоатомных структур, для которых число решений будет не слишком велико.

4. Бесповторная генерация ЦПС. Данная подпрограмма представляет собой наиболее важный комбинаторный блок в компьютерной системе «ФЛАМИНГО». Быстродействие этого блока определяет быстродействие программы в целом. В модуле А реализован лексикографический обзор

«дерева перебора», в ходе которого генерируются все возможные ЦПС, с учетом действия формальных и эмпирических критериев отбора. Модуль В реализует критерий эквивалентности на основе информации о группе автоморфизмов графа.

5. Вывод конечных ХС. Вывод структур конечных ХС может (по желанию пользователя) осуществляться на печатающее устройство (в виде цифровых кодов для соответствующих ЦПС), на графический дисплей и принтер-плоттер (в виде рисунков).

6. Выбор результатов для исследования на следующей стадии (многостадийный КС). Из множества генерированных результатов химик может выбрать конечные ХС, которые представляют интерес для дальнейшего исследования. Выбранные ХС хранятся в специальном массиве внутри самой программы; когда возможности для неповторной генерации ЦПС в заданной ХС исчерпаны, программа поочередно выбирает структуры, записанные в этот массив, и вводит их как исходные ХС на следующей стадии компьютерного поиска.

Система «ФЛАМИНГО» обладает широкими возможностями для решения разнообразных проблем синтетической органической химии и установления механизмов органических реакций. Исходная ХС, которая задается химиком в режиме диалога (см. п. 1), может содержать до 32 центров, причем центрами являются отдельные атомы или целые атомные группировки, вводимые как единое целое. С учетом возможностей персональной ЭВМ «Искра-226», число стадий компьютерного синтеза, которые могут одновременно храниться в оперативной памяти ЭВМ, не превышает шести, однако с использованием внешних запоминающих устройств (например, гибкого магнитного диска) число стадий КС может быть произвольно велико. Система «ФЛАМИНГО» позволяет работать с любыми органическими структурами, включая сложные многоатомные соединения, гетероциклические, металлоорганические соединения, ионные и радикальные частицы.

С целью демонстрации возможностей системы «ФЛАМИНГО» был проведен компьютерный поиск и исследование путей синтеза ряда каркасных соединений, перегруппировок в семействах изомеров, возможных механизмов органических реакций с участием ионных и радикальных частиц [31—34]. Проведенное исследование продемонстрировало хороший уровень предсказательной способности системы «ФЛАМИНГО». При этом в ходе работы программы были не только воспроизведены практически все литературные данные по изучаемым превращениям, но и предсказаны новые, оригинальные пути синтеза, перегруппировки и механизмы реакций исследованных органических соединений [121, 122].

IX. ПЕРСПЕКТИВЫ РАЗВИТИЯ КОМПЬЮТЕРНОГО СИНТЕЗА

В настоящее время уже нет никаких сомнений в том, что для решения задач планирования синтеза, прогнозирования направления реакции, изучения перегруппировочных процессов необходимо применение компьютеров. Более того, формализованное представление химической информации в ЭВМ позволяет осуществить стратегическое планирование химического эксперимента с использованием новых или малоизученных процессов.

Как сказано выше, компьютерный синтез развивается в двух направлениях, а именно как эмпирический КС и неэмпирический КС. В начале своего развития программы компьютерного синтеза должны были выполнять роль не только аппарата для планирования синтеза, но, в большей степени, автоматизированного справочного пособия по синтетическим методам органической химии. Естественно, с этой ролью могли справиться только системы эмпирического КС. Однако в 1980-х гг. появились высокоорганизованные базы данных по химическим реакциям, например «REACCS» [123], специально ориентированные на поиск известных реакций по заданным структурным фрагментам, условиям реакции и т. д. Поэтому в настоящее время при решении задачи планирования синтеза

на первый план выдвигается проблема поиска оригинальной стратегии синтеза, проблема, исчерпывающее решение которой под силу именно программам неэмпирического направления, системам «искусственного интеллекта». Поэтому, несмотря на то, что эмпирическое направление несравненно более развито, будущее, как нам представляется, за комбинацией этих подходов, иначе говоря, за «полуэмпирическими» программами КС. С этой точки зрения особенно перспективным кажется обогащение неэмпирических программ эмпирическими критериями отбора, позволяющими сочетать достоинства исчерпывающего, беспристрастного комбинаторного перебора с системой экспертных оценок, выработанных всем опытом органической химии.

ЛИТЕРАТУРА

1. Corey E. J., Wipke W. T. // Science. 1969. V. 166. P. 178.
2. Bersohn M., Esack A. // Chem. Rev. 1976. V. 76. P. 269.
3. Gasteiger J. // Chim. Ind. 1982. V. 64. P. 714.
4. Костилов Р. Р., Мандельштам Т. В., Разин В. В. // Современные проблемы органической химии. 1982. Вып. 7. С. 4.
5. Huggin J. // Chem. Eng. News. 1983. V. 61. № 9. P. 7.
6. Zeigler E. // Computer in der Chemie. Berlin: Springer, 1984. S. 280.
7. Овчинников А. А., Болдырев А. И. // Успехи химии. 1986. Т. 55. С. 539.
8. Winter J. H. // Chemische Syntheseplanung. Berlin: Springer, 1982. S. 215.
9. Маки Р., Смит Д. // Путеводитель по органическому синтезу. М.: Мир, 1985. С. 352.
10. Hendrickson J. B. // J. Amer. Chem. Soc. 1977. V. 99. P. 5439.
11. Wipke W. T., Braun H., Smith G. et al. // ACS Symp. Ser. 1977. V. 61. P. 97.
12. Weise A. // Z. Chem. 1973. Jg. 13. S. 155.
13. Weise A. // Ibid. 1975. Jg. 15. S. 333.
14. Westphal G., Klebsch A., Weise A. et al. // Z. Chem. 1977. Jg. 17. S. 295.
15. Weise A., Scharnow H. G. // Ibid. 1979. Jg. 19. S. 49.
16. Schubert W. // MATCH. 1979. V. 6. P. 213.
17. Schubert W., Ugi I. // Chimia. 1979. Jg. 33. S. 183.
18. Salatin T. D., Jorgensen W. L. // J. Org. Chem. 1980. V. 45. P. 2043.
19. Salatin T. D., McLaughlin D., Jorgensen W. L. // Ibid. 1981. V. 46. P. 5284.
20. Peishoff C. E., Jorgensen W. L. // Ibid. 1983. V. 48. P. 1970.
21. Burnier J. S., Jorgensen W. L. // Ibid. 1983. V. 48. P. 3923.
22. Burnier J. S., Jorgensen W. L. // Ibid. 1984. V. 49. P. 3001.
23. Peishoff C. E., Jorgensen W. L. // Ibid. 1985. V. 50. P. 1056.
24. Peishoff C. E., Jorgensen W. L. // Ibid. 1985. V. 50. P. 3174.
25. Bures M. G., Roos-Kozel R. L., Jorgensen W. L. // Ibid. 1985. V. 50. P. 4490.
26. Gushurst A., Jorgensen W. L. // Ibid. 1986. V. 51. P. 3513.
27. Agnihotri R. B., Motard R. L. // ACS Symp. Ser. 1980. V. 124. P. 193.
28. Brandt J., Friedrich J., Gasteiger J. et al. // ACS Symp. Ser. 1977. V. 61. P. 33.
29. Gasteiger J., Jochum C. // Top. Curr. Chem. 1978. V. 74. P. 93.
30. Marsili M., Gasteiger J., Carter R. E. // Chim. Oggi. 1984. № 9. P. 11.
31. Гордеева Е. В. // Тезисы докл. Всесоюз. конф. по молекулярным перегруппировкам. Ереван, 1985. С. 25.
32. Гордеева Е. В. Материалы конференции молодых ученых Химфака МГУ. Москва, 1985 г. Ч. 1. С. 53—56. — Деп. в ВИНТИ. 1985, № 8373-В.
33. Зефирова Н. С., Гордеева Е. В., Трач С. С. // Тезисы докл. VII Всесоюз. конф. «Использование вычислительных машин в химических исследованиях и спектроскопии молекул». Рига, 1986. С. 19.
34. Гордеева Е. В. Автореф. дис. ... канд. хим. наук. М.: МГУ им. М. В. Ломоносова, 1986.
35. Corey E. J. // Quart. Rev., Chem. Soc. 1971. V. 25. P. 455.
36. Corey E. J., Cramer R. D. III., Howe W. J. // J. Amer. Chem. Soc. 1972. V. 94. P. 440.
37. Corey E. J., Howe W. J., Pensak D. A. // Ibid. 1974. V. 96. P. 7724.
38. Stolow R. D., Joncas L. J. // J. Chem. Educ. 1980. V. 57. P. 868.
39. Corey E. J., Pensak D. A. // ACS Symp. Ser. 1977. V. 61. P. 1.
40. Long A. K., Rubenstein S. D., Joncas L. J. // Chem. Eng. News. 1983. V. 61. № 9. P. 22.
41. Corey E. J., Long A. K., Rubenstein S. D. // Science. 1985. V. 228. P. 418.
42. Moreau G. // Nouv. J. Chim. 1978. V. 2. P. 187.
43. Choplin F., Laurencio C., Marc R. et al. // Ibid. 1978. V. 2. P. 285.
44. Choplin F., Bonnet P., Zimmer M. H., Kaufmann G. // Ibid. 1979. V. 3. P. 223.
45. Zimmer M. H., Choplin F., Bonnet P., Kaufmann G. // J. Chem. Inf. Comp. Sci. 1979. V. 19. P. 235.
46. Choplin F., Goundiam S., Kaufmann G. // Ibid. 1983. V. 23. p. 26.
47. Laurencio C., Villien L., Kaufmann G. // Tetrahedron. 1984. V. 40. P. 2721.
48. Laurencio C., Villien L., Kaufmann G. // Ibid. 1984. V. 40. P. 2731.
49. Varkony T. H., Smith D. H., Djerassi C. // Ibid. 1978. V. 34. P. 841.
50. Varkony T. H., Carhart R. E., Smith D. H., Djerassi C. // J. Chem. Inf. Comp. Sci. 1978. V. 18. P. 168.

51. Govind R., Powers G.//ACS Symp. Ser. 1977. V. 61. P. 81.
52. Barone R., Chanon M., Metzger J.//Tetrahedron Lett. 1974. P. 2761.
53. Barone R., Chanon M., Cadiot P., Cense J. M.//Bull. Chem. Soc. Belg. 1982. V. 91. P. 333.
54. Barone R., Chanon M., Contreras M. L.//Nouv. J. Chim. 1984. V. 8. P. 311.
55. Gelernter H. L., Sridharan N. S., Hart A. J. et al.//Top. Curr. Chem. 1973. V. 41. P. 113.
56. Gelernter H. L., Sanders A. F., Larsen D. L. et al.//Science. 1977. V. 197. P. 1041.
57. Agarwal K. K., Larsen D. L., Gelernter H. L.//Comput. and Chem. 1978. V. 2. P. 75.
58. Gelernter H. L., Bhagwat S. S., Larsen D. L., Miller G. A.//Anal. Chem. Symp. Ser. 1983. V. 15. P. 35.
59. Hendrickson L. B., Braun-Keller E.//J. Comput. Chem. 1980. V. 1. P. 323.
60. Hendrickson J. B., Braun-Keller E., Toczko A. G.//Tetrahedron. 1981. V. 37. Suppl. № 1. P. 359.
61. Hendrickson J. B., Grier D. L., Toczko A. G.//J. Amer. Chem. Soc. 1985. V. 107. P. 5228.
62. Doenges R., Groebel B., Nickelsen H., Sander J.//J. Chem. Inf. Comp. Sci. 1985. V. 25. P. 425.
63. Bersohn M.//Bull. Chem. Soc. Japan. 1972. V. 45. P. 1897.
64. Bersohn M., Esack E., Luchini J.//Comput. and Chem. 1978. V. 2. P. 105.
65. Bersohn M.//ACS Symp. Ser. 1979. V. 112. P. 341.
66. Blower P. E., Whitlock H. W.//J. Amer. Chem. Soc. 1976. V. 98. P. 1499.
67. Пиотух-Пелецкий В. И., Смирнов В. И., Коптюг В. А.//Тезисы докл. VII Всесоюз. конф. «Использование вычислительных машин в химических исследованиях и спектроскопии молекул». Рига, 1986. С. 3.
68. Corey E. J., Wipke W. T., Cramer R. D. III, Howe W. J.//J. Amer. Chem. Soc. 1972. V. 94. P. 421.
69. Masinter L. M., Lederberg J., Sridharan N. S., Smith D. H.//Ibid. 1974. V. 96. P. 7702.
70. Bersohn M., Esack A.//Chem. Scripta. 1974. V. 6. P. 122.
71. Esack A., Bersohn M.//J. Chem. Soc., Perkin Trans. I. 1975. № 12. P. 1124.
72. Bersohn M.//Comput. and Chem. 1978. V. 2. P. 113.
73. Bersohn M.//J. Chem. Soc., Perkin Trans. I. 1979. № 8. P. 1975.
74. Wipke W. T., Dyott T. M.//J. Amer. Chem. Soc. 1974. V. 96. P. 4825.
75. Wipke W. T., Dyott T. M.//Ibid. 1974. V. 96. P. 4834.
76. Hendrickson J. B., Toczko A. G.//J. Chem. Inf. Comp. Sci. 1983. V. 23. P. 171.
77. Corey E. J., Petersson G. A.//J. Amer. Chem. Soc. 1972. V. 94. P. 450.
78. Corey E. J., Wipke W. T., Cramer R. D. III, Howe W. J.//Ibid. 1972. V. 94. P. 431.
79. Bersohn M.//J. Chem. Soc., Perkin Trans. I. 1973. № 12. P. 1239.
80. Esack A., Bersohn M.//Ibid. 1974. № 21. P. 2463.
81. Esack A.//Ibid. 1975. № 12. P. 1120.
82. Bersohn M., Esack A.//Chem. Scripta, 1976. V. 9. P. 211.
83. Bersohn M.//J. Chem. Soc. Perkin Trans. I. 1982. № 2. P. 631.
84. Jochum C., Gasteiger J.//J. Chem. Inf. Comp. Sci. 1979. V. 19. P. 43.
85. Jochum C., Gasteiger J.//Ibid. 1979. V. 19. P. 49.
86. Roos-Kozel B. L., Jorgensen W. L.//Ibid. 1981. V. 21. P. 101.
87. Hendrickson J. B., Grier D. L., Toczko A.//Ibid. 1984. V. 24. P. 195.
88. Bersohn M., Esack A.//Comput. and Chem. 1977. V. 1. P. 103.
89. Харари Ф., Палмер Э. Перечисление графов. М.: Мир, 1977. С. 324.
90. Hendrickson J. B.//J. Amer. Chem. Soc. 1971. V. 93. P. 6847.
91. Hendrickson J. B.//Ibid. 1975. V. 97. P. 5763.
92. Hendrickson J. B.//Ibid. 1975. V. 97. P. 5784.
93. Hendrickson J. B. Top. Curr. Chem. 1976. V. 62. P. 49.
94. Hendrickson J. B.//J. Chem. Inf. Comp. Sci. 1979. V. 19. P. 129.
95. Hendrickson J. B.//Acc. Chem. Res. 1986. V. 19. P. 274.
96. Corey E. J., Jorgensen W. L.//J. Amer. Chem. Soc. 1976. V. 98. P. 189.
97. Corey E. J., Orf H. W., Pensak D. A.//Ibid. 1976. V. 98. P. 210.
98. Corey E. J., Jorgensen W. L.//Ibid. 1976. V. 98. P. 203.
99. Corey E. J., Long A. K., Greene T. W., Miller J. W.//J. Org. Chem. 1985. V. 50. P. 1920.
100. Gasteiger J.//Comput. and Chem. 1978. V. 2. P. 85.
101. Gasteiger J., Dammer O.//Tetrahedron. 1978. V. 34. P. 2939.
102. Gasteiger J., Jacob P., Strauss U.//Ibid. 1979. V. 35. P. 139.
103. Gasteiger J.//Ibid. 1979. V. 35. P. 1419.
104. Gasteiger J., Marsili M.//Tetrahedron Lett. 1978. N 34. P. 3181.
105. Gasteiger J., Marsili M.//Tetrahedron. 1980. V. 36. P. 3219.
106. Gasteiger J., Hutchings M. G.//Tetrahedron Lett. 1983. V. 24. P. 2541.
107. Gasteiger J., Saller H.//Angew. Chem. 1985. Jg. 97. S. 699.
108. Gasteiger J., Hutchings M. G.//Tetrahedron Lett. 1983. V. 24. P. 2537.
109. Guillen M. D., Gasteiger J.//Tetrahedron. 1983. V. 39. P. 1331.
110. Gasteiger J., Hutchings M. G.//J. Chem. Soc., Perkin Trans. II. 1984. № 3. P. 559.
111. Gasteiger J., Hutchings M. G.//J. Amer. Chem. Soc. 1984. V. 106. P. 6489.
112. Mortier W. J., van Genechten K., Gasteiger J.//Ibid. 1985. V. 107. P. 829.
113. Зедуров Н. С., Трач С. С.//Журн. орган. химии. 1981. Т. 17. С. 2465.
114. Зедуров Н. С., Трач С. С.//Там же. 1975. Т. 11. С. 225.
115. Зедуров Н. С., Трач С. С.//Там же. 1975. Т. 11. С. 1785.

116. Зефилов Н. С., Трач С. С.//Там же. 1976. Т. 12. С. 7.
117. Зефилов Н. С., Трач С. С.//Там же. 1976. Т. 12. С. 697.
118. Zefirov N. S., Tratch S. S.//Chem. Scripta. 1980. V. 15. P. 4.
119. Трач С. С., Зефилов Н. С.//Журн. орган. химии. 1982. Т. 18. С. 1561.
120. Жидков Н. П., Зефилов Н. С., Попов А. И. и др.//Математические вопросы структурного анализа и алгоритмы машинных экспериментов в органической химии/Под ред. Жидкова Н. П. и др. М.: Изд. МГУ, 1979. С. 59.
121. Зефилов Н. С., Гордеева Е. В., Кузнецова Т. С., Кожушков С. И.//Журн. орган. химии (в печати).
122. Zefirov N. S., Gordeeva E. V.//J. Org. Chem. (в печати).
123. Venkataratnam R. V., Thyagarajan G., Husain A.//CEW, Chem. Eng. World. 1985. V. 20. № 6. P. 61.

Институт органической химии им. Н. Д. Зелинского
АН СССР, Москва