

Д. Е. Лушников, Е. В. Бабаев

## МОЛЕКУЛЯРНЫЙ ДИЗАЙН ГЕТЕРОЦИКЛОВ

## 4\*. ПРИМЕНЕНИЕ КОМПЬЮТЕРОВ В ХИМИИ ГЕТЕРОЦИКЛОВ

(ОБЗОР)

В обзоре рассмотрены и систематизированы основные принципы применения компьютеров в дизайне гетероциклических структур и реакций. Кратко рассмотрена оригинальная программа авторов GREN, перечисляющая рециклизации гетероциклов, и программа Heterocycland, основанная на предложенной ранее [2] концепции «структура-синтез» и использованная для ретросинтеза шестичленных гетероциклов.

Настоящий обзор преследует несколько целей: во-первых, познакомить широкую аудиторию химиков-гетероциклистов с методологией и самыми последними достижениями в области применения компьютеров в органической химии вообще и химии гетероциклов в частности. Во-вторых, акцентировать внимание самих специалистов в области компьютерной химии на специфике проблем, имеющих в химии гетероциклов, и на известных способах их решения. Наконец, обзор может быть адресован просто думающим химикам-органикам, которым не безразличны исследования на стыке наук. Для специалистов, имеющих о предмете обзора отдаленное представление, мы старались сделать изложение идей компьютерной химии максимально доступным.

Рассматривая применение компьютеров в химии, нельзя не удивляться пестроте и широкой гамме подходов. Только с помощью компьютеров можно осуществлять квантовохимические расчеты многоатомных молекул [3] или оптимизировать методами молекулярной механики конформации полициклов [4]. Ключевой роли компьютеров в обработке химической информации посвящены самостоятельные обзоры [5, 6], созданы (и уже широко используются) компьютерные программы для установления структуры нового соединения на основе имеющейся о нем спектральной и иной информации [7, 8]. Наконец, сформировалось целое направление компьютерной химии, ориентированное на поиск закономерностей типа «структура-свойство» и «структура-активность» (в английской аббревиатуре — QSPR и QSAR) [9—11].

Разумеется, в цитируемых подходах, где основной целью является анализ статических (например, физико-химических или структурных) свойств веществ, гетероциклические объекты играют далеко не последнюю роль (просто в силу практической значимости гетероциклов, скажем, как лекарственных препаратов, красителей и т. д.). Вместе с тем, понимая главную задачу и предмет химии как изучение изменения веществ в химических процессах, мы постараемся сконцентрировать основное внимание на использовании компьютеров при анализе именно химических превращений, т.е. рассмотрим компьютерные аспекты синтеза и реакций органических веществ, а именно гетероциклических соединений. Несмотря на существование обзоров в этой области [12—16], специфика компьютерной химии гетероциклов, по-видимому, до сих пор не была предметом самостоятельного изучения.

\* Сообщение 3 см [1]

**КОМПЬЮТЕРНЫЙ ДИЗАЙН СТРУКТУР И РЕАКЦИЙ:  
КЛАССИФИКАЦИЯ ЦЕЛЕЙ И КОНЦЕПЦИЙ**

При обсуждении использования компьютеров для решения синтетических проблем можно говорить о трех главных «компонентах» синтеза — об *исходных реагентах, реакциях и их продуктах*. Обычно один или несколько из этих компонентов не известны и являются предметом исследования при использовании компьютера, а другие известны. Будем считать эти компоненты независимыми, а следовательно, их можно комбинировать между собой со знаком «плюс» (если исходная информация имеется) или «минус» (если информации нет или она составляет задачу исследования). Простой перебор дает нам 8 возможных комбинаций, которые представлены в таблице «Цели и концепции компьютерной химии», которая, по-видимому, предлагается впервые. Следует подчеркнуть принципиальную разницу между информацией о веществах и информацией о реакциях; первая поддается довольно простой обработке, а вторая нет\*. Поэтому мы условно разделили таблицу на 4 возможных типа задач, которые отражают наличие или отсутствие *данных* о реагентах и продуктах и разные оттенки *знания* о протекании реакций.

Отметим, что ориентированное на математику и логику абстрактное представление реакций в литературе нередко характеризуется словами «формальное», «логическое» или их сочетанием [12]. Противоположный подход обозначается как «эмпирический» или «информационно-ориентированный».

Случаи I—III (знак плюс в графе «Продукты») отвечают ретросинтетическому планированию синтеза, т. е. нахождению способов синтеза *целевой структуры* (далее — ЦС), неважно, каким методом и из чего. В реальной практике приходится вводить ограничения на способы (стратегию) синтеза. В случае I, кроме того, имеется информация о доступных реагентах.

Цели и концепции компьютерной химии

Тип	Реагенты	Реакции	Продукты
I	+	+	+
II	-	+	+
III	-	-	+
IV	+	-	+
V	+	+	-
VI	+	-	-
VII	-	+	-
VIII	-	-	-

+ — используется эмпирическое знание о предмете. - — не используется эмпирическое знание о предмете.

Случай IV (есть ЦС и реагенты) — отвечает «несвободе» в выборе реагентов для ЦС (подходящие реагенты должны иметься в наличии в лаборатории или в каталоге). Кроме того, неявно подразумевается некоторое знание о реакциях, что сближает случаи IV и I.

Случаи V, VI (задан набор реагентов). Задача прямого (или синтетического) планирования синтеза заключается в предсказании продуктов реакции между реагентами в заданных условиях. Возможны элементарные сведения (случай VI) о правилах протекания реакций.

\* Информация о веществах — реагентах и продуктах — обычно представляет собой *данные* (либо структуры, либо список структур, например, каталог коммерчески доступных реагентов): они либо есть, либо нет. Напротив, вводимая в компьютер информация о реакциях представляет собой некоторое *знание* о том, как происходят превращения соединений.

Случай VII (есть знание о реакциях) отвечает классификации и поиску новых реакций и изучению их механизмов. Сюда же относится подбор оптимальных объектов для реализации абстрактных идей.

Два предельных случая I и VIII имеют особенности. Первый из них (цели поставлены, реакции известны, реагенты даны) смыкается с экспериментальной практикой. Второй случай (отсутствие реагентов, знаний о реакциях и целей исследования) подразумевает отсутствие также и компьютеров.

Ниже мы обсудим все случаи, используя в качестве реагентов и/или продуктов гетероциклические структуры и акцентируя внимание на специфике знания о гетероциклических реакциях (например, о рециклизациях или «магических» правилах для циклизаций гетероциклов). Особую роль будут играть примеры, приведенные на схемах: все они предсказаны с использованием компьютера (вместо реагентов над стрелками мы будем ставить имя программы, в которой предложена реакция). Экспериментальное соответствие будет оговариваться для каждого случая отдельно.

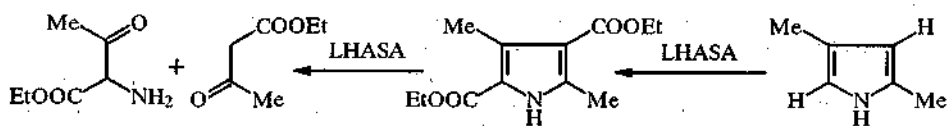
### РЕТРОСИНТЕТИЧЕСКИЙ АНАЛИЗ (ПОДХОДЫ I, II)

Давая своеобразную «ретроспективу идей ретросинтеза», заметим, что впервые идею использовать электронно-вычислительные машины для поиска путей синтеза органических соединений предложил еще в 1963 г. Г. Влэдуч [17], бывший сотрудник ВИНТИ (г. Москва). В 1967 г. Кори, с именем которого обычно ассоциируется рождение компьютерного ретросинтеза, предложил концепцию *ретросинтетического анализа* целевой молекулы и разработал общие принципы планирования синтеза [18].

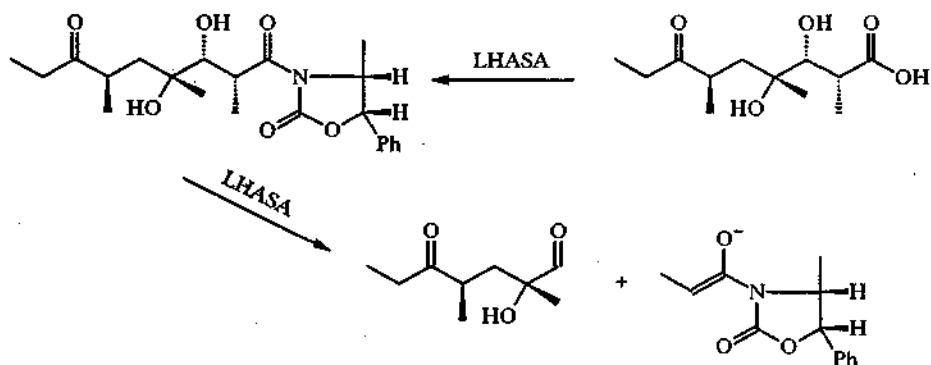
В концепции Кори результатом анализа является набор синтетических предшественников, а применение к ним процедуры ретросинтеза приводит к образованию *дерева синтеза*. Таким образом, при разработке плана синтеза принципиальны два момента: 1) как генерировать синтетические предшественники исходя из заданной ЦС (тактика); 2) как управлять построением дерева синтеза, чтобы избежать «комбинаторного взрыва» и найти оптимальные пути в этом дереве (стратегия).

Эмпирическое компьютерное планирование синтеза, или ретросинтетический анализ, в целом имитирует способ мышления химика. Он основан на выделении в ЦС *ретронов* — неких структурных особенностей, дающих химику подсказку о конкретной реакции (или последовательности реакций), приводящей к созданию ЦС. Трансформации структуры, обратные реакциям, получили название *ретрореакций*. Когда структурный фрагмент выявлен, компьютер оценивает, является ли окружение фрагмента благоприятным для осуществления задуманного превращения. Если да, то компьютер заменяет фрагмент его предшественником. Решение о том, какие ретрореакции могут быть использованы, принимается на основе эмпирических правил, составляющих стратегию синтеза. Классический набор правил Кори (стратегий), реализованный в программе LHASA, общеизвестен [19].

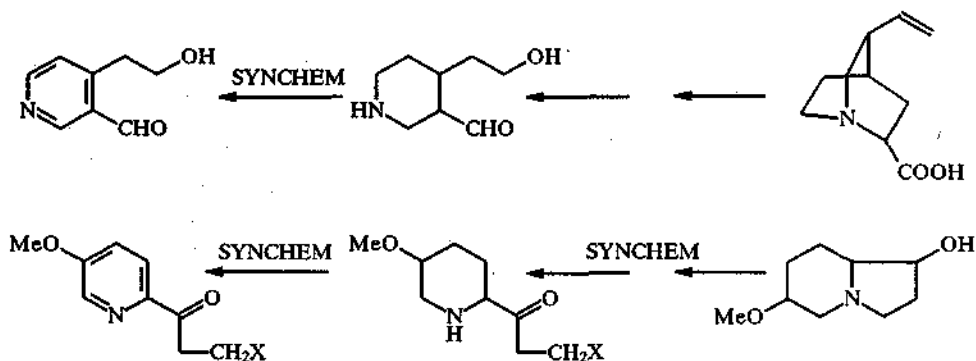
Следующий пример показывает, как LHASA использует стратегию изменения функциональных групп для «разборки» пиррола по Кнорру [20] (к сожалению, в оригинале статьи опечатка — вместо этоксикарбонильной группы приведена этильная):



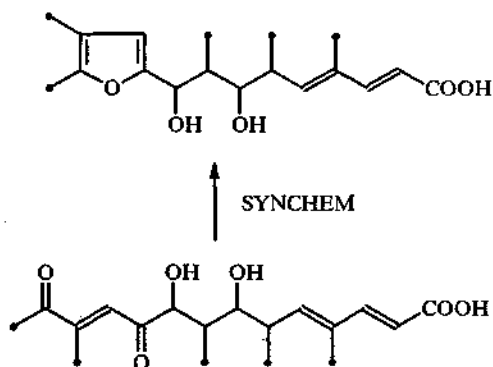
Заметим, что наибольшая эффективность подхода Кори была продемонстрирована при анализе сложных алициклических или насыщенных гетерополициклических систем [19]. Что касается гетероциклов с сопряженными связями, то они использовались как защитные [21] или вспомогательные [22] группы, например:



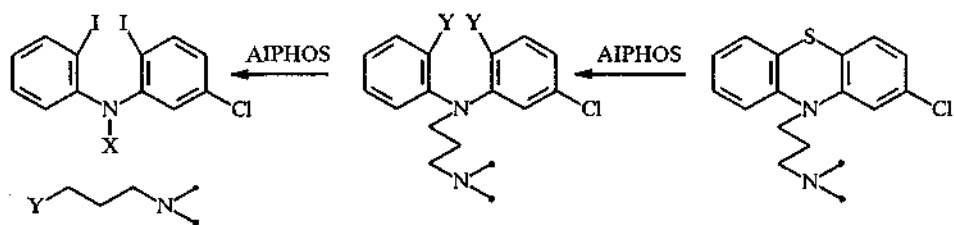
Лишь в одной из работ пиридиновый цикл с подходящей функциональностью рассматривался как предшественник насыщенных бициклов:



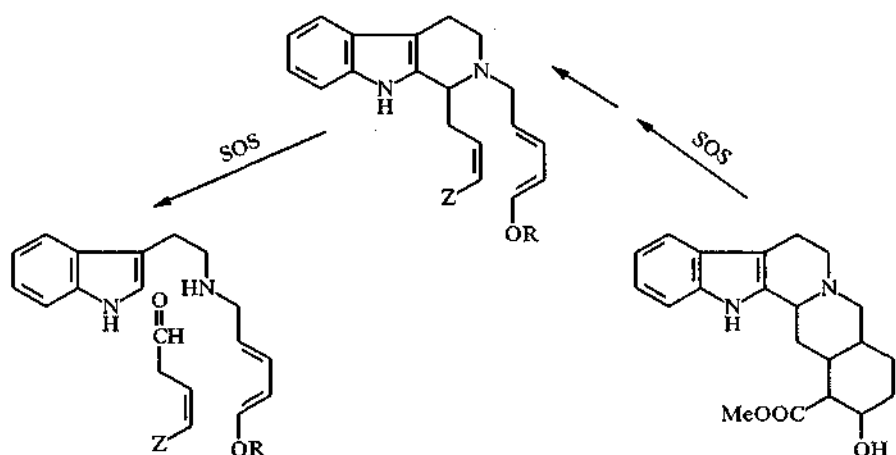
а фурановый цикл — как предшественник цепи с требуемым расположением функций [23]:



В работах и программах, посвященных дальнейшему развитию ретросинтетического подхода, гетероциклические структуры стали появляться и в качестве ЦС. Между тем ретрореакции, предлагавшиеся для сборки гетероциклов, не выходили за рамки простейших двухкомпонентных синтезов, главным образом типа 5+1 циклизаций, как, например, в ретросинтезе хлорпромазина [24]:

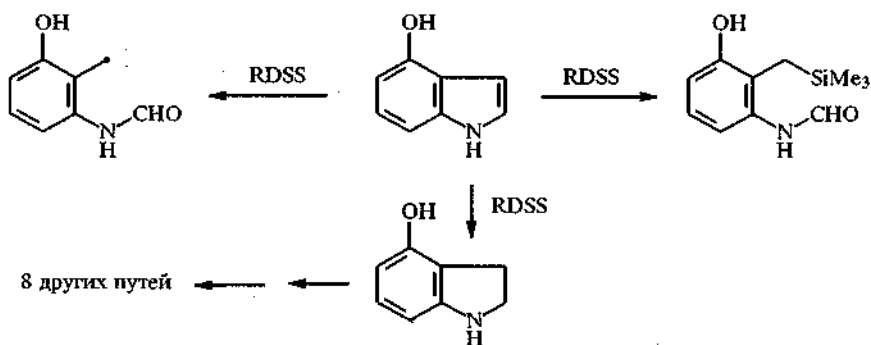


Аналогичная 5+1 циклизация (с заменой 1,1-бинуклеофила на 1,1-бизлектрофил) применена в ретросинтезе алкалоида иохимбина [25]:



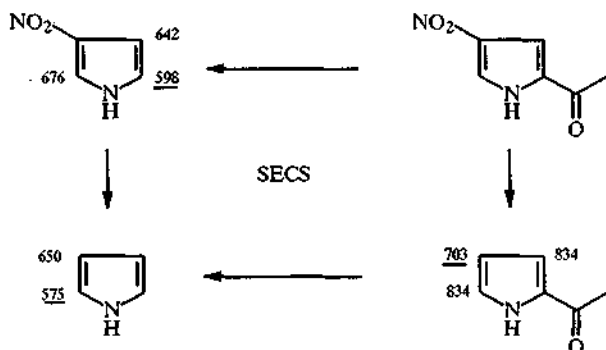
Следует подчеркнуть, что методология программ типа LHASA не может выйти за рамки библиотеки ретрореакций. В результате при ретросинтезе, например, гетероциклического ядра, способ получения которого не введен в библиотеку реакций, программа не сможет предложить адекватного пути синтеза.

По-видимому, наибольшее число гетероциклизаций (более 300) содержится в библиотеке трансформаций программы RDSS [26]. В качестве примера ее работы представлена часть дерева синтеза 4-гидроксииндола. Из 11 лучших предшественников только два пути приводят к целевому индолу:

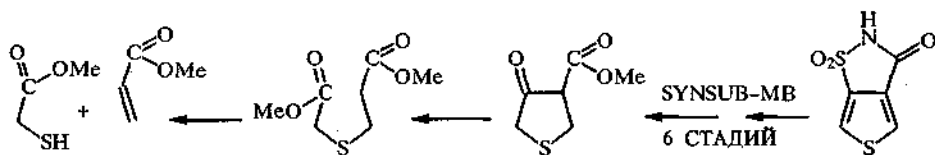


Остальные пути базируются на появлении в качестве предшественника гидрированной структуры 4-окси-2,3-дигидроиндола и построены на обычных превращениях алифатических фрагментов и реакциях ароматического замещения, приводящих к нему.

В ряде подходов удается построить модель для оценки селективности введения функций в ароматические и гетероароматические системы. Например, в программе SECS [27] предпочтительное направление реакций этого типа определяется на основе сравнения электронных энергий локализации для разных положений, рассчитанных по методу Хюккеля. Приведенный пример иллюстрирует идею такого планирования синтеза дизамещенного пиррола:



Разработка программ поиска путей синтеза, которые работают неинтерактивно (т. е. без участия пользователя), накладывает целый ряд дополнительных ограничений на используемые ретрореакции и дерево синтеза (например, число стадий, возможный выход, цена реагентов, возможность усложнения предшественников в ходе синтеза и т. д.). Сложность проблемы может быть показана на примере работы программы химической компании Sumimoto [28], ориентированной на предельную дешевизну используемых реактивов. При максимально разрешенном числе стадий, равном семи, программа не нашла ни одного «дешевого» пути синтеза «тиофен-сахарина», а для восьмистадийного синтеза воспроизвела метод, описанный в литературе:



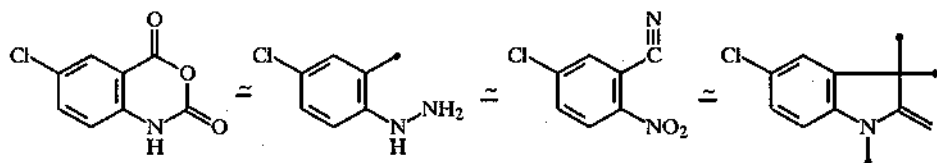
Приведенные примеры показывают, что, хотя гетероароматические системы встречались в числе объектов ретросинтетического анализа, никаких решающих обобщений о способах ретросинтеза гетероциклов сделано не было.

#### ПОИСК ИСХОДНЫХ РЕАГЕНТОВ ДЛЯ СИНТЕЗА ЦС (ПОДХОД IV)

Одним из принципов, используемых химиками в практической работе, является взаимозаменяемость или синтетическая эквивалентность реагентов. Между тем ретросинтез приводит к каким-то конкретным структурам, и, таким образом, если обучить компьютер распознаванию химически эквивалентных или подобных структур, то каждый предшественник, генерируемый компьютером, можно будет сравнить с его аналогами, например, содержащимися в коммерческих каталогах.

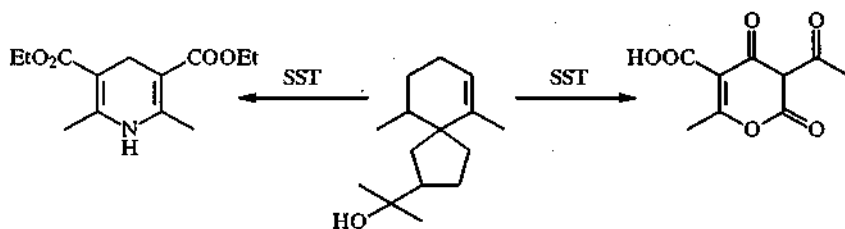
Ключевым моментом является определение того, что именно считать мерой подобия. О сложности этой проблемы говорит тот факт, что в одной из

последних программ [29] для поиска исходных реагентов было использовано 48 (!) определений структурного подобия молекул. Один из таких критериев, введенный в компьютер, отождествил следующие ароматические и гетероароматические системы из каталога Jannsen:



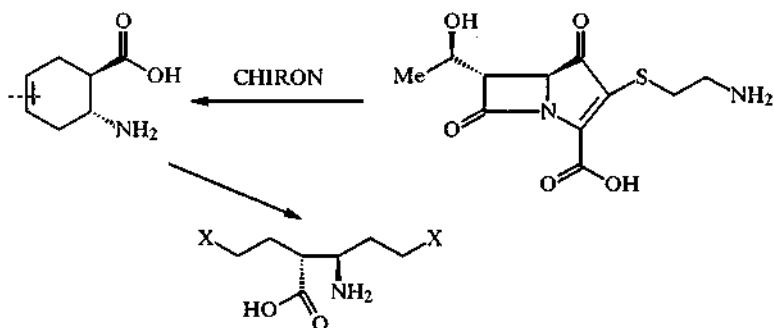
(Приведенные структуры объявлены подобными, так как все они имеют бензольный фрагмент и одинаковую природу атомов, примыкающих к ядру.)

Второй вариант оптимального подбора реагентов для синтеза ЦС можно сопоставить с «интуитивным скачком» [30], который позволяет химик сразу увидеть в исходных реагентах прообраз конечной структуры без планирования стадий синтеза. В программе SST [30] авторы попытались имитировать такой подход с помощью компьютера. В частности, реализация идеи структурного сходства в программе позволила найти в каталоге гетероциклические структуры, подходящие в качестве предшественников для синтеза карбоциклического соединения агароспирала:



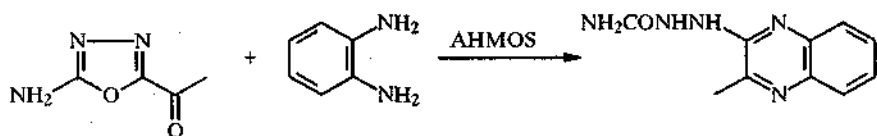
Тем не менее нам не известны другие примеры использования гетероциклов как специфических предшественников раскрытых форм в компьютерном синтезе.

Специально для поиска хиральных синтетических предшественников была разработана программа CHIRON. Она содержит библиотеку из более тысячи доступных оптически чистых и около 300 рацемических соединений и позволяет находить хиральные шаблоны, в том числе скрытые, для синтеза оптически чистых природных соединений [31]. Например, для синтеза хирального гетероцикла тиенамицина компьютером был найден следующий шаблон:



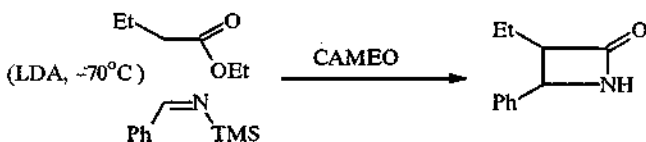
**АНАЛИЗ МЕХАНИЗМОВ РЕАКЦИИ  
(ПОДХОД V)**

Компьютерный анализ сложных механизмов требует введения «элементарных сведений об элементарных стадиях». Например, программа AHMOS рассматривает 8 типов элементарных реакций (присоединение, замещение, диссоциация, протонирование, секстетная перегруппировка, поляризация, электрофильное замещение и элиминирование) и 6 типов реакционных центров (жесткие и мягкие электрофилы и нуклеофилы, нуклеофуги и электрофуги) [32]. Небольшой банк данных содержит индексы реакционной способности, количественно отражающие нуклеофильные и электрофильные свойства различных функциональных групп. На его основе программа определяет наиболее приоритетные способы взаимодействия электрофильных и нуклеофильных центров. Такой сравнительно простой подход позволил правильно предсказать результат протекания несложных ионных реакций, например [33]:

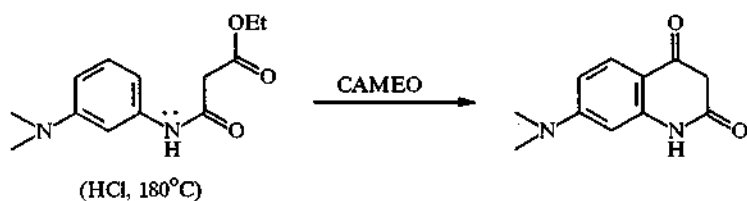


Наиболее последовательно моделирование механизмов реакций реализовано в работах группы Йоргенсена [34], который попытался тщательно проанализировать фундаментальные типы механизмов реакций и конкуренцию между ними. Созданная в результате программа SAMEO может прогнозировать нуклеофильные, электрофильные, перициклические, окислительно-восстановительные, свободнорадикальные и карбеновые реакции.

Таким образом, программа SAMEO позволяет решать задачи достаточно высокого уровня. Если ввести структуры исходных реагентов и задать условия реакции, то программа сама определит, какие классы механизмов могут быть использованы. Впечатляют примеры предсказанных гетероциклизаций, совпадающие с экспериментом [35]. Например, предсказано следующее катализируемое основанием нуклеофильное присоединение-замещение:

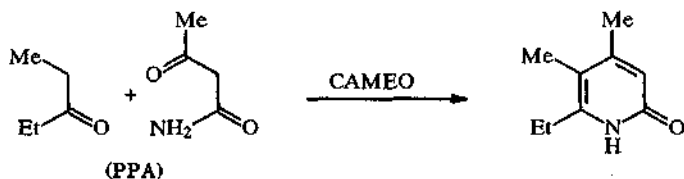


Другой пример — гетероциклизация путем внутримолекулярного электрофильного ароматического замещения:

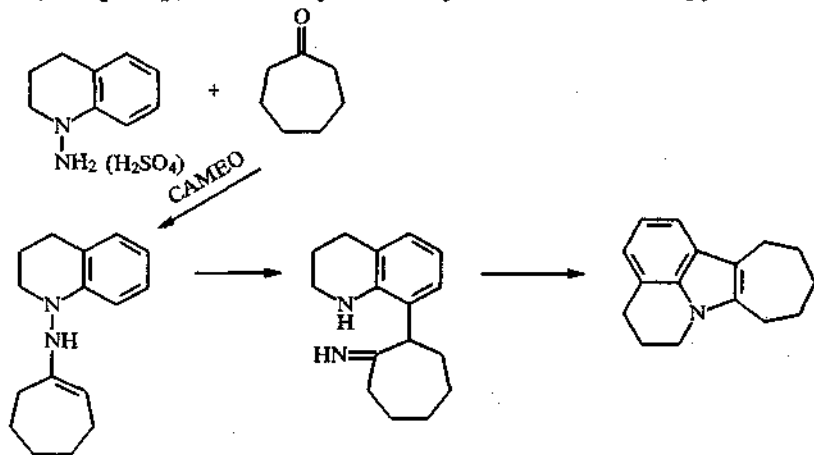




Еще один тип прогнозов представлен реакциями конденсации:

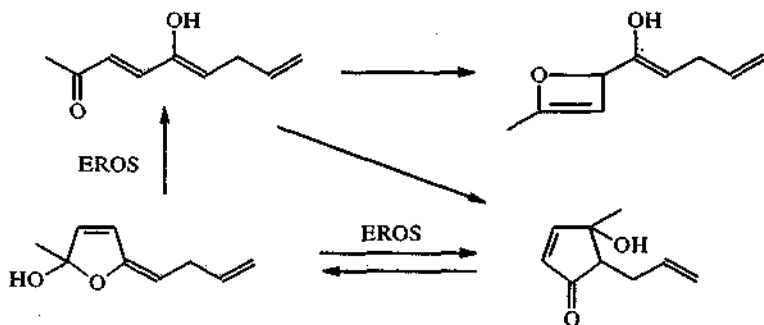


Отличительной особенностью программы CAMEO является возможность исследовать сложные синтезы, в которых реализуется несколько типов механизмов. Так, программа правильно предсказывает последовательность реакций, например, отвечающую синтезу индолов по Фишеру:



#### МОДЕЛИРОВАНИЕ РЕАКЦИОННОЙ СПОСОБНОСТИ СОЕДИНЕНИЙ (ПОДХОД VI)

Другой подход к предсказанию результата протекания органических реакций разрабатывается группой Гастайгера в программе EROS [36, 37]. Реакции в программе EROS генерируются формальным образом путем смещения электронов, удаления и расстановки связей. Другими словами, предметом исследования является не механизм, а суммарный результат реакции. Основная задача программы состоит в автоматическом выделении химически осуществимых реакций из множества формально возможных. С этой целью были разработаны численные модели расчета энтальпии реакций и способности связей к гомолитическому и гетеролитическому разрыву. Учитывая при оценке формальных интермедиатов как энтальпию, так и реакционную способность связи, программа правильно предсказывает результат весьма сложных перегруппировок [36]:



Рассмотренные подходы охватывают достаточно широкие разделы химии и в целом обладают весьма высокой предсказательной способностью. Оценка альтернативных путей протекания реакции позволяет предсказывать образование побочных продуктов.

#### МЕТОДЫ КЛАССИФИКАЦИИ ОРГАНИЧЕСКИХ РЕАКЦИЙ (ПОДХОД VII)

До настоящего времени в органической химии не разработана универсальная систематическая классификация органических реакций, которая была бы применима во всех компьютерных приложениях — от систематического поиска реакций в базах данных до рационального планирования синтеза и поиска новых реакций [5]. Первые математические подходы к описанию структурных изменений в ходе химической реакции были разработаны в 70-х годах Балабаном [38, 39], Ути и Дугунджи [40, 41], Хендриксоном [42], Трачом и Зефириным [43, 44]. Более поздние публикации в этом направлении представлены работами Фуджиты [45, 46], Робертса [47], Аренса [48], Кваснички [49, 50], Уилкокса [51] и других.

Центральная идея практически всех подходов заключается в установлении соответствия между атомами реагентов и продуктов путем их «наложения» друг на друга. После такого слияния или суперпозиции можно однозначно описать изменения связей в ходе химической реакции. Далее возникает проблема выразить такое перераспределение связей в какой-либо доступной компьютеру форме на языке математики.

Заметим, что способы представления информации о реакциях были разными у разных исследовательских групп, однако доминировали две стороны одной идеи: либо матричное представление реагентов и реакций, либо использование графических диаграмм, более близкое к языку структурной химии. Не вдаваясь в обсуждение особенностей каждого подхода, мы представили на схеме 1 воображаемую реакцию гетероциклизации и попытались показать, каким образом на языке различных подходов выглядят диаграммы одной и той же реакции. В нижней части приведен более громоздкий способ матричного описания той же реакции. Более детально различные аспекты представления реакций рассмотрены в обзорах [13, 52].

Отметим, что все представленные подходы описывают суммарный результат реакций и никак не учитывают их механизм.

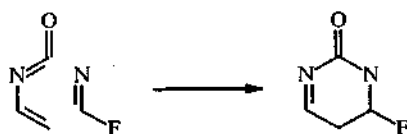
#### НЕЭМПИРИЧЕСКИЙ ДИЗАЙН НОВЫХ РЕАКЦИЙ (ПОДХОД VII)

Представленные на схеме 1 разные диаграммы описания реакций сходны в том, что они существенно упрощают обычные химические уравнения. Остается обычно графический образ («скелет») реакции, содержащий, как правило, атомы, меняющие степень окисления (или какое-либо окружение), и связи, меняющие кратность в ходе реакции. Таким образом, химические реакции можно свести к конечному числу диаграмм перераспределения связей. Сопоставление таких диаграмм для разных реакций позволяет, прежде всего, выявить сходство или степень новизны самих реакций [40, 44].

Другой способ использования диаграмм перераспределения связей — прогнозирование новых реакций. Для этого требуется нарастить на эти диаграммы реальные заместители, цепи, циклы и т. д. В ряде случаев такое прогнозирование являлось эвристичным.

Указанный принцип был положен в основу программ SYMBEQ [53, 54] и IGOR [55]. С помощью программы SYMBEQ осуществлено генерирование следующего символического уравнения и экспериментально реализован соответствующий ему синтез фуранового ядра [53]:

Различные языки формального описания реакций

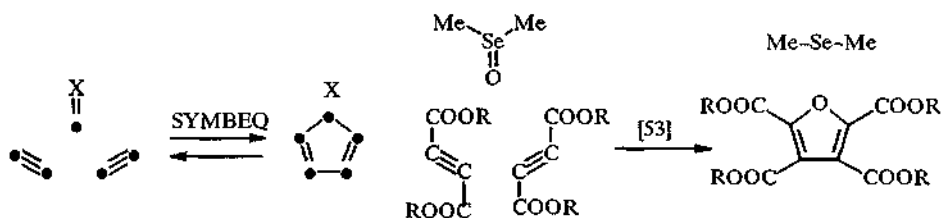


<p>Хергес - Уги (1985) [83]</p> <p>гетерореакция</p> <p>базовая реакция</p> <p>категория реакции</p>	<p>Трач - Зефилов (1975) [43]</p> <p>реакционное уравнение</p> <p>символическое уравнение</p> <p>топология</p> <p><math>CA_4 CA_2 \rightarrow RA_4 RA_2</math></p> <p>классификационное уравнение</p>	<p>Фуджита (1986) [45]</p> <p>Уилкоккс (1985) [51]</p> <p>Хендриксон (1974) [42]</p>
--	---	--

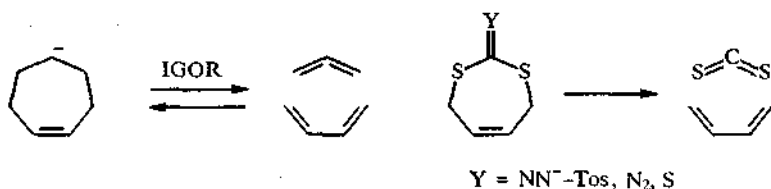
Уги - Дугунджи (1971) [41]

<p>ССНСНСФОНННН</p> <table border="1"> <tr><td>C</td><td>/</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td></tr> <tr><td>C</td><td>2</td><td>/</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>N</td><td>0</td><td>1</td><td>/</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>C</td><td>0</td><td>0</td><td>2</td><td>/</td><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>N</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>C</td><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td><td>/</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td></tr> <tr><td>F</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>O</td><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>H</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>H</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>H</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>H</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td></tr> <tr><td>H</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td></tr> <tr><td>H</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td></tr> </table> <p>В-матрица исходных реагентов</p>	C	/	2	0	0	0	0	0	0	0	0	1	1	0	0	C	2	/	1	0	0	0	0	1	0	0	0	0	0	0	N	0	1	/	2	0	0	0	0	0	0	0	0	0	0	C	0	0	2	/	0	0	0	2	0	0	0	0	0	0	N	0	0	0	0	/	2	0	0	0	0	1	0	0	0	C	0	0	0	2	0	/	1	0	0	0	0	0	1	0	F	0	0	0	0	1	0	/	0	0	0	0	0	0	0	O	0	0	0	2	0	0	0	/	0	0	0	0	0	0	H	0	1	0	0	0	0	0	0	/	0	0	0	0	0	H	1	0	0	0	0	0	0	0	0	/	0	0	0	0	H	0	0	0	1	0	0	0	0	0	0	/	0	0	0	H	0	0	0	0	1	0	0	0	0	0	0	/	0	0	H	0	0	0	0	0	1	0	0	0	0	0	0	/	0	H	0	0	0	0	0	0	1	0	0	0	0	0	0	/	<p>ССНСНСФОНННН</p> <table border="1"> <tr><td>/</td><td>-1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>-1</td><td>/</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>1</td><td>/</td><td>-1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>-1</td><td>/</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>1</td><td>/</td><td>-1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>0</td><td>-1</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td></tr> </table> <p>Р-матрица реакции</p>	/	-1	0	0	0	1	0	0	0	0	0	0	0	0	0	-1	/	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	/	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	/	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	/	-1	0	0	0	0	0	0	0	0	0	1	0	0	0	-1	/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	/	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	/	0	<p>ССНСНСФОНННН</p> <table border="1"> <tr><td>/</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>C</td></tr> <tr><td>1</td><td>/</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>C</td></tr> <tr><td>0</td><td>2</td><td>/</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>N</td></tr> <tr><td>0</td><td>0</td><td>1</td><td>/</td><td>1</td><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>C</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>1</td><td>/</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>N</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td><td>/</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>C</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>F</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>2</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>O</td></tr> <tr><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>H</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>0</td><td>H</td></tr> <tr><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>0</td><td>H</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>0</td><td>H</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>0</td><td>H</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>/</td><td>H</td></tr> <tr><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>1</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>0</td><td>H</td></tr> </table> <p>Е-матрица продуктов</p>	/	1	0	0	0	1	0	0	0	1	1	0	0	0	C	1	/	2	0	0	0	0	0	1	0	0	0	0	0	C	0	2	/	1	0	0	0	0	0	0	0	0	0	0	N	0	0	1	/	1	0	0	2	0	0	0	0	0	0	C	0	0	0	1	/	1	0	0	0	0	1	0	0	0	N	1	0	0	0	1	/	1	0	0	0	0	1	0	0	C	0	0	0	0	1	0	/	0	0	0	0	0	0	0	F	0	0	0	2	0	0	0	/	0	0	0	0	0	0	O	0	1	0	0	0	0	0	0	/	0	0	0	0	0	H	1	0	0	0	0	0	0	0	0	/	0	0	0	0	H	1	0	0	0	0	0	0	0	0	0	/	0	0	0	H	0	0	0	1	0	0	0	0	0	0	0	/	0	0	H	0	0	0	0	1	0	0	0	0	0	0	0	/	0	H	0	0	0	0	0	1	0	0	0	0	0	0	0	/	H	0	0	0	0	0	0	1	0	0	0	0	0	0	0	H
C	/	2	0	0	0	0	0	0	0	0	1	1	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
C	2	/	1	0	0	0	0	1	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
N	0	1	/	2	0	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
C	0	0	2	/	0	0	0	2	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
N	0	0	0	0	/	2	0	0	0	0	1	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
C	0	0	0	2	0	/	1	0	0	0	0	0	1	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
F	0	0	0	0	1	0	/	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
O	0	0	0	2	0	0	0	/	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
H	0	1	0	0	0	0	0	0	/	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
H	1	0	0	0	0	0	0	0	0	/	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
H	0	0	0	1	0	0	0	0	0	0	/	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
H	0	0	0	0	1	0	0	0	0	0	0	/	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
H	0	0	0	0	0	1	0	0	0	0	0	0	/	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
H	0	0	0	0	0	0	1	0	0	0	0	0	0	/																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
/	-1	0	0	0	1	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
-1	/	1	0	0	0	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	1	/	-1	0	0	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	-1	/	1	0	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	1	/	-1	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
1	0	0	0	-1	/	0	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	0	0	/	0	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	0	0	0	/	0	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	0	0	0	0	/	0	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	0	0	0	0	0	/	0	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	0	0	0	0	0	0	/	0	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	0	0	0	0	0	0	0	/	0	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	0	0	0	0	0	0	0	0	/	0	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	0	0	0	0	0	0	0	0	0	/	0																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
/	1	0	0	0	1	0	0	0	1	1	0	0	0	C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
1	/	2	0	0	0	0	0	1	0	0	0	0	0	C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	2	/	1	0	0	0	0	0	0	0	0	0	0	N																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	1	/	1	0	0	2	0	0	0	0	0	0	C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	1	/	1	0	0	0	0	1	0	0	0	N																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
1	0	0	0	1	/	1	0	0	0	0	1	0	0	C																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	1	0	/	0	0	0	0	0	0	0	F																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	2	0	0	0	/	0	0	0	0	0	0	O																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	1	0	0	0	0	0	0	/	0	0	0	0	0	H																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
1	0	0	0	0	0	0	0	0	/	0	0	0	0	H																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
1	0	0	0	0	0	0	0	0	0	/	0	0	0	H																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	1	0	0	0	0	0	0	0	/	0	0	H																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	1	0	0	0	0	0	0	0	/	0	H																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	0	1	0	0	0	0	0	0	0	/	H																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
0	0	0	0	0	0	1	0	0	0	0	0	0	0	H																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									

Приведено описание гипотетической реакции Дильса—Альдера (вверху) при помощи разных формальных подходов математической химии (см. текст), как если бы ее описывали сами авторы подходов.



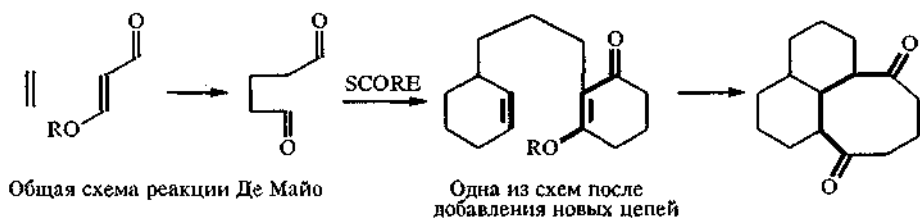
При помощи программы IGOR был, например, найден и осуществлен экспериментально новый способ генерации 1,4-диенов на основе гетероциклической системы [56]:



Следует отметить, что и программа SYMBEQ, и программа IGOR лишь генерируют исчерпывающим образом примеры реакций, а оценка их достоверности целиком зависит от пользователя, его химического опыта и интуиции.

#### ИССЛЕДОВАНИЕ ДИАПАЗОНА ПРИМЕНИМОСТИ РЕАКЦИЙ (ПОДХОД VII)

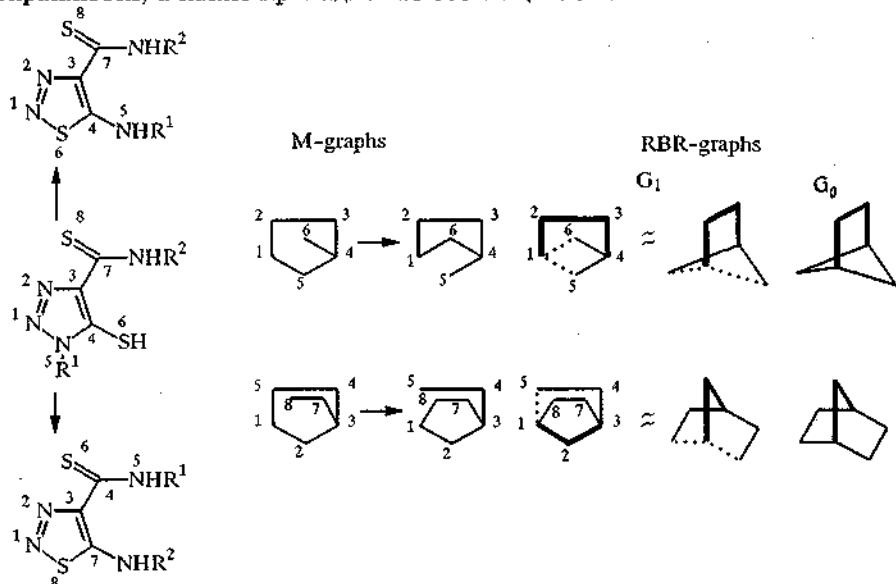
Программа SCORE [57] пытается ответить на вопрос: какие виды циклических структур можно получить при помощи заданной реакции? Для этого она добавляет цепочки из 2—4 атомов углерода последовательно ко всем парам атомов на заданной абстрактной диаграмме реакции. Ниже представлено одно из решений, предложенных этой программой для реакции Де Майо:



На схеме слева показана схема реакции, которая задана пользователем; остальные связи были последовательно добавлены программой. Замена атомов углерода в полученных скелетах на гетероатомы и расстановка кратных связей проводятся химиком самостоятельно. Таким образом, программа может быть использована для дизайна новых гетероциклических систем. (Заметим, что программы SCORE и IGOR написаны для IBM-совместимого компьютера и приложены на дискете к журналам [55, 57].)

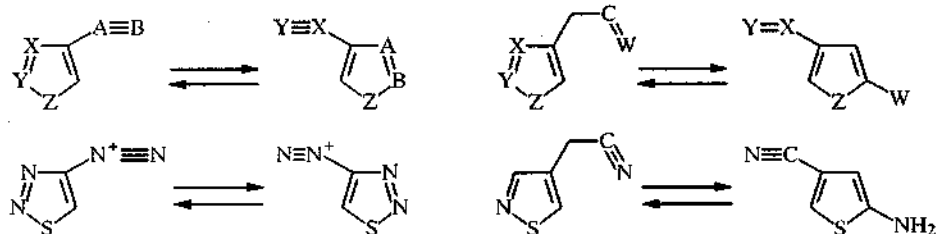
## КЛАССИФИКАЦИЯ И ДИЗАЙН РЕЦИКЛИЗАЦИЙ ГЕТЕРОЦИКЛОВ (ПОДХОД VII)

Реакции рециклизации известны для многих гетероциклов с любым размером цикла и почти любым типом, числом и распределением гетероатомов [58]. В сообщении 1 настоящего сериала [59] была предложена новая методология классификации гетероциклических перегруппировок, далее развитая в работе [60]. Ее основная идея состоит в обсуждавшейся выше суперпозиции исходной и конечной гетероциклических структур. Специфика подхода (и его главное отличие от абстрактных диаграмм на схеме 1) заключается в учете лишь тех связей, которые входят как в исходный, так и в формируемый цикл. Специальные типы меток показывают, какие именно циклические связи рвутся, образуются или сохраняются, а какие принадлежат обоим циклам:



Полученная диаграмма реакции имеет естественный химический смысл: на единой схеме химик видит одновременно бывший и будущий гетероциклы, а также положение рвущихся и вновь формируемых связей.

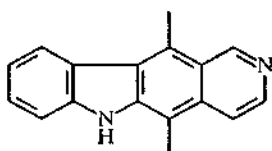
Результирующие диаграммы реакций всегда имеют простую бициклическую структуру, что существенно упрощает анализ сложных перегруппировок. На их основе была разработана последовательная иерархическая классификация трансформаций гетероциклов, которая позволяет учитывать и чисто химический фактор — распределение электрофильных и нуклеофильных центров в разрываемых и образуемых связях. В компьютерной программе GREH [60] были реализованы алгоритмы исчерпывающего перечисления всех перегруппировок на каждом уровне иерархии. Простая возможность вводить в программу схемы уже известных реакций позволила использовать такой подход для прогнозирования неизвестных структурных типов перегруппировок. При помощи этой программы были предложены некоторые новые типы рециклизаций:



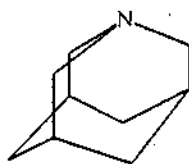
Z: NR, O, S; X, Y: CH, N; A≡B: C≡CH, N<sup>+</sup>≡CR, N<sup>+</sup>≡N; C≡W: COR, CN

ЛОГИЧЕСКИЕ МЕТОДЫ ПОИСКА ПУТЕЙ СИНТЕЗА  
(ПОДХОД П)

Простейший подход к ретросинтезу, не требующий от программы никаких химических знаний, — разорвать одну или несколько связей в ЦС и выдать полученные структуры химику для дальнейшего анализа. Именно так работала первая версия микрокомпьютерной программы SOS [61]. (Этот подход можно смело назвать «методом *ab initio* компьютерного синтеза», хотя сами авторы оправданно ставят его на один уровень с методом Хюккеля [13].) Такой тип подходов, позволяющий фиксировать внимание химика на неочевидных возможностях сборки ЦС, оказался полезным при обсуждении путей синтеза простых гетероциклических молекул. Например, в работе [62] компьютер генерировал 253 способа рассечения двух связей в скелете эллиптицина (А), из которых авторы выбрали для обсуждения 41 комбинацию. Для азаадамантана (В) было генерировано 2510 способов рассечения, из которых (вручную) выбрано для анализа 25 схем [61].



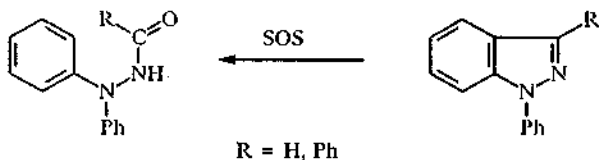
А



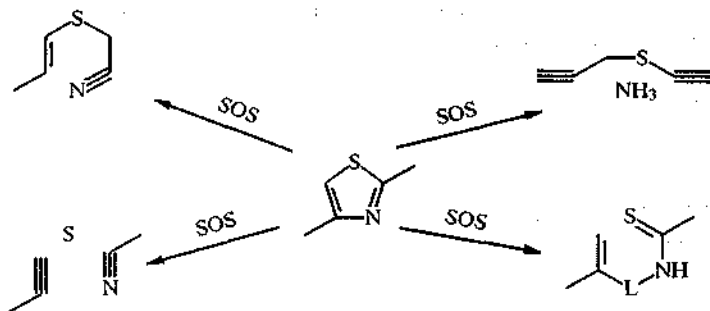
В

Несмотря на то, что такой неэмпирический подход к компьютерному синтезу, разумеется, способен предлагать новые пути синтеза, трудоемкость оценки большого числа предшественников, сложность узнавания известных и распознавания новых путей синтеза сильно ограничивают сферу его применения.

Между тем, введение даже минимальной информации об элементарных реакциях значительно облегчает восприятие и оценку результатов работы программы при сохранении потенциальной новизны решений. Например, в той же программе SOS [63] был использован небольшой набор схем, описывающих основные типы реакций: нуклеофильное замещение и присоединение, элиминирование, ароматическое электрофильное и нуклеофильное замещение, таутомерный сдвиг, а также окисление и восстановление. При помощи данной программы проведен анализ путей синтеза индазолов. Спрогнозированный в работе [64] новый путь синтеза ядра индазола был реализован экспериментально той же исследовательской группой:

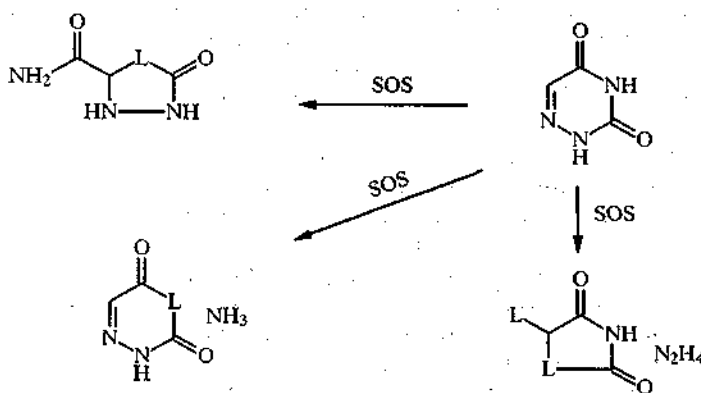


При разработке новых путей синтеза триазола [65] компьютер предложил 800 интермедиатов, из которых 140 было отобрано для анализа. Четыре приведенные ниже реакции оценены как наиболее перспективные:

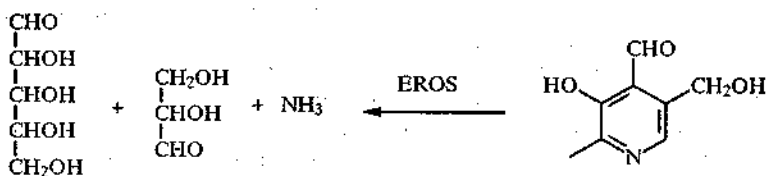


Один из предложенных в работе [66] способов был позднее подтвержден экспериментально [67].

В работе тех же авторов для синтеза 6-азаурацила компьютер предложил 250 интермедиатов, из которых авторы приводят и обсуждают только 12 путей синтеза [68]. Отметим, что среди предложенных компьютером путей синтеза были и рециклизации, например:

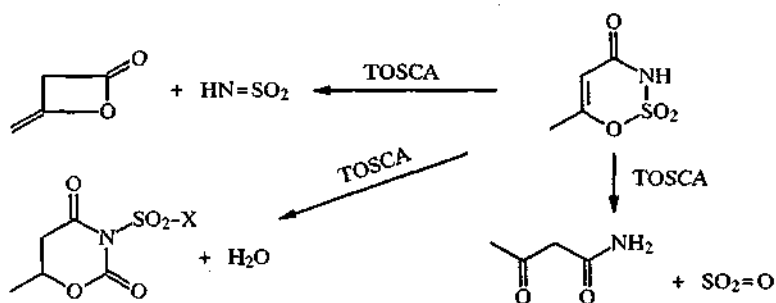


Программы EROS и TOSCA используют наборы «генераторов реакций», которые описывают в самом общем виде перераспределение связей в ходе химических реакций [36, 69, 70]. Выбор синтетических предшественников в программе EROS в основном проводится на основании оценки энтальпии реакции (см. раздел «Оценка реакционной способности»). Данный подход может быть полезен, в принципе, для установления путей биосинтеза простых природных соединений. Например, для пиридоксала в числе возможных предшественников были найдены аммиак, глицериновый альдегид и пентоза [69]:



Программа TOSCA концерна Hoechst использует идею о так называемых «консонантных» и «диссонантных» структурах и реакциях для планирования синтеза ЦС [70]. В структуре определяется донорный или акцепторный характер входящих в нее атомов, на основе чего молекула классифицируется как диссонантная (возможно соседство двух донорных или двух акцепторных

центров) или консонантная, в противном случае. Априорно постулируется использование «консонантных синтезов для консонантных структур» [70]. Приведенные ниже пути синтеза, найденные программой TOSCA, представляют пример использования такой стратегии:



Наиболее серьезная проблема, возникающая при практическом использовании программы, заключается в появлении большого числа шумовых структур на стадии генерации реагентов для синтеза гетероцикла. Так, приведенные выше структуры получены при отборе из более чем 700 предложенных структур. Заметим, что идея консонантных молекул и реакций имела чисто умозрительный характер и, тем не менее, представляется наиболее близкой к проблематике гетероциклического синтеза.

В раннем сообщении этого сериала на основании обобщения экспериментальных данных по синтезу шестичленных гетероциклов [2] был предложен набор магических правил «структура-синтез», имеющий определенную аналогию с идеей консонантности. Центральное различие заключается в том, что правила «структура-синтез» предлагают класс синтонов, тогда как программа TOSCA генерирует конкретные (в том числе неправдоподобные) реагенты, приводя к «комбинаторному взрыву».

Нами разработана компьютерная программа Heterocycland [71], реализующая предложенные ранее правила «структура-синтез» [2]. Программа исчерпывающим образом генерирует все возможные способы гетеролитического формирования связей в ЦС.

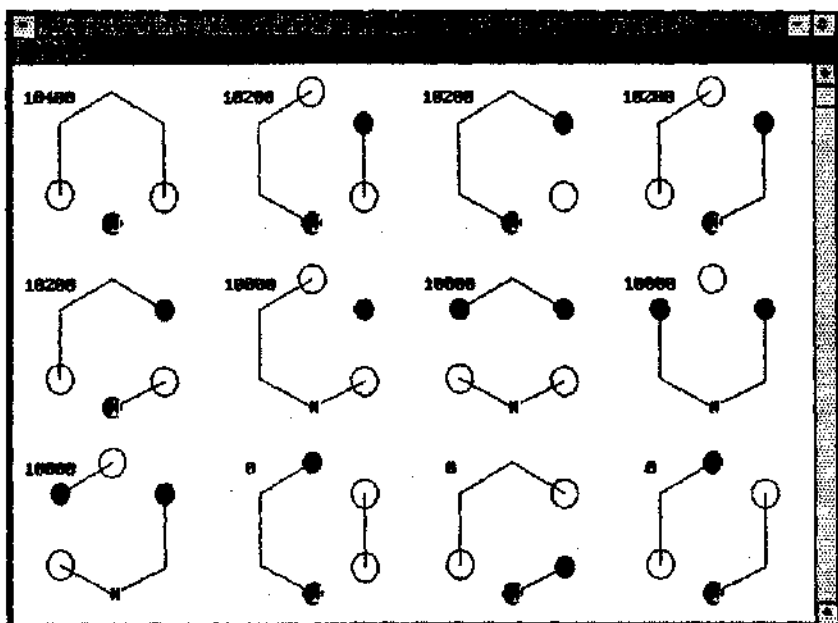
В качестве примера рассмотрим генерацию и отбор наиболее перспективных типов гетеролитических реакций, приводящих к пиридиновому ядру. Первым шагом работы программы является исчерпывающая неповторная генерация гетеролитических диссекций связей ЦС. Так, возможным двухкомпонентным синтезам ядра пиридина отвечает 33 таких разбиения, которые включают, в частности, и неправдоподобные сочетания полярных центров в реагентах.

Пользователю предоставляется две возможности: 1) отсечь сочетания синтонов, запрещенные правилом гетероальтернирования; 2) ранжировать синтезы на основе предпочтительного образования какого-либо типа связи (например, углерод—гетероатом). Пример осуществления такого отбора приведен на схеме 2. Как видно, количество типов синтезов сократилось до 9 (высокий рейтинг), 8 из которых отвечают известным в литературе синтезам пиридинового ядра, а девятый обнаружен экспериментально в предыдущем сообщении [1]. Остальные схемы синтеза с низким рейтингом соответствуют неизвестным или весьма редким синтезам пиридинового ядра [2].

Нельзя не отметить крупный вклад в развитие идей планирования синтеза, внесенный работами Хендриксона [42, 72—77]. В его подходе чисто химическая идея гетеролитического формирования  $\text{C}-\text{C}$  связи удачно сопряжена с численным выражением электро- и нуклеофильности углеродного атома [42, 74]. В разработанной им методологии использования полуреакций [73, 75], охватывающих большинство известных схем реакций



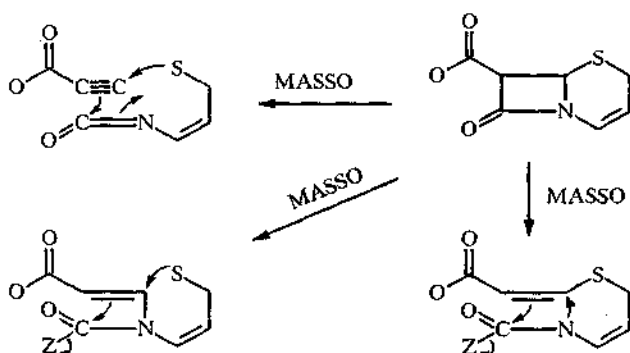
Пример генерации синтезов программой Heterocycland  
(печать с экрана компьютера)



Видны предложенные программой полярные типы реагентов (черная метка — нуклеофильный центр, белая — электрофильный) для «разрешенных» двухкомпонентных синтезов ядра пиридина. Цифры возле схем синтеза отвечают предпочтительности того или иного метода («ноль» соответствует редким или экспериментально неизвестным схемам синтеза).

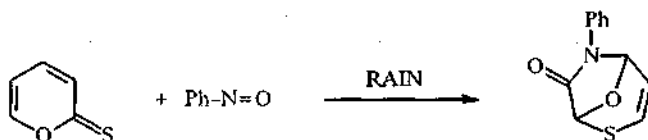
[76], ключевую роль играет ближайшее окружение углеродного атома (например, число протонов, кратных связей и уходящих групп). К сожалению, в этом подходе до сих пор гетероциклические структуры не являлись объектами анализа.

Применение идей Хендриксона к образованию связей углерод—гетероатом реализовал Моро в своей программе MASSO [78]. Это перспективное направление [79], однако рассматривается только один опубликованный пример:



## ДВУСТОРОННИЙ КОМПЬЮТЕРНЫЙ СИНТЕЗ (ПОДХОД IV)

Группой Уги разрабатывается оригинальный двусторонний метод планирования синтеза [80, 81]. Его идея заключается в одновременном построении дерева синтеза с двух сторон — со стороны ЦС и со стороны реагентов, что весьма эффективно сокращает общее число интермедиатов по сравнению с однонаправленным поиском. Двусторонний синтез был реализован в программе RAIN [81], которая использовалась для поиска неочевидных механизмов реакций и объяснения путей образования побочных продуктов. Например, для такого неочевидного превращения, как приведенная на схеме реакция Стрейта, программа RAIN предложила несколько вполне правдоподобных механизмов [82]:



Таким образом, по своей сути большинство рассмотренных подходов компьютерной химии в применении к гетероциклам не выходят за рамки обычных реакций образования С—С или С—Х связей. Между тем, как мы попытались доказать, отдельные подходы к дизайну гетероциклических структур и реакций, использующие специфику гетероциклического синтеза, лишь начинают формироваться.

### СПИСОК ЛИТЕРАТУРЫ

1. Бабаев Е. В. // ХГС. — 1993. — № 7. — С. 962.
2. Бабаев Е. В. // ХГС. — 1993. — № 7. — С. 937.
3. Clark T. A Handbook of Computational Chemistry. — N. Y.: Wiley, 1985. — 332 p. (Рус. перевод М.: Мир, 1990. — 384 с.)
4. Burkert U., Allinger N. L. Molecular Mechanics. — Washington: Am. Chem. Soc., 1982. — 339 p. (Рус. перевод М.: Мир, 1986. — 364 с.)
5. Chemical Information Systems: Beyond the Structure Diagram / Eds. D. Bawden, E. Mitchell. — Chichester: E. Horwood, 1990. — 178 p.
6. Computer Aids to Chemistry / Eds. G. Vernin, M. Chanon. — Chichester: E. Horwood, 1986. — 411 p. (Рус. перевод Л.: Химия, 1990. — 383 с.)
7. Artificial Intelligence Applications in Chemistry / Eds. T. H. Pierce, B. A. Hohne. — Washington: Am. Chem. Soc. — 1986. — 395 p. (Рус. перевод М.: Мир, 1988. — 428 с.)
8. Lindsay R. K., Buchanan B. G., Feigenbaum E. A., Lederberg J. Applications of the Artificial Intelligence for Organic Chemistry. The DENDRAL Project. — N. Y.: McGraw Hill, 1980. — 332 p.
9. Розенблит А. В., Голендер В. Е. Логико-комбинаторные методы в конструировании лекарственных. — Рига: Зинатне, 1983. — 351 с.
10. Станкевич М. И., Станкевич И. В., Зефиоров Н. С. // Усп. химии. — 1988. — Т. 57. — С. 337.
11. Раевский О. А., Сапегин А. М. // Усп. химии. — 1988. — Т. — 57. — С. 1565.
12. Зефиоров Н. С., Гордеева Е. В. // Усп. химии. — 1987. — Т. 56. — С. 1753.
13. Varone R., Chanon M. Computer Aids to Chemistry / Eds. G. Vernin, M. Chanon. — Chichester: E. Horwood, 1986. — P. 9. (Рус. перевод Л.: Химия, 1990. — С. 11.)
14. Winter J. H. Chemische Syntheseplanung. — Berlin: Springer-Verlag, 1982. — 215 S.
15. Репинская И. Б. Ретросинтетический подход к планированию синтеза органических соединений. — Новосибирск, 1989. — 92 с.
16. Serrafosa F. Organic Chemistry in Action. The Design of Organic Synthesis. — Amsterdam: Elsevier, 1990. — 395 p.
17. Vledits G. // Inf. Storage Retr. — 1963. — Vol. 1. — P. 117.
18. Corey E. J. // Pure Appl. Chem. — 1967. — Vol. 14. — P. 19.
19. Corey E. J. // Chem. Soc. Rev. — 1988. — Vol. 17. — P. 111.
20. Johnson A. P. // Chem. Brit. — 1985. — Vol. 21. — P. 59.
21. Corey E. J., Long A. K., Greene T. W., Miller J. W. // J. Org. Chem. — 1985. — Vol. 50. — P. 1920.

22. Corey E. J., Long A. K., Rubenstein S. D. // *Science*. — 1985. — Vol. 228. — P. 408.
23. Gelernter H. L., Sanders A. F., Larsen D. L., Agarwal K. K., Boivie R. H., Spritzer G. A., Searleman J. E. // *Science*. — 1977. — Vol. 197. — P. 1041.
24. Funatsu K., Sasaki S. I. // *Tetrahedron Comput. Methodol.* — 1988. — Vol. 1. — P. 27.
25. Azario P., Barone R., Chanon M. // *J. Org. Chem.* — 1988. — Vol. 53. — P. 720.
26. Haase F., Biedka K. // *Tetrahedron Comput. Methodol.* — 1990. — Vol. 3, N 6B. — P. 461.
27. Wipke W. T., Braun H., Smith G., Choplin H., Steber W. // *Computer-Assisted Organic Synthesis*. — ACS Symp. Ser. N 61. — 1977. — P. 97.
28. Takahashi M., Dogane I., Yoshida M., Yamachika H., Takabatake T., Bersohn M. // *J. Chem. Inf. Comput. Sci.* — 1990. — Vol. 30. — P. 436.
29. Gasteiger J., Ihlenfeldt W. D., Rose P. // *Rec. Trav. Chim.* — 1992. — Vol. 111. — P. 270.
30. Wipke W. T., Rogers D. // *J. Chem. Inf. Comput. Sci.* — 1984. — Vol. 24. — P. 71.
31. Hanessian S., Franco J., Larouche B. // *Pure Appl. Chem.* — 1990. — Vol. 62. — P. 1887.
32. Weise A. // *Z. Chem.* — 1973. — Bd 13. — S. 155.
33. Weise A., Westphal G., Rabe H. // *Z. Chem.* — 1981. — Bd 21. — S. 218.
34. Jorgensen W. L., Laird E. R., Gushurst A. J., Fleischer J. M., Gothe S. A., Helson H. E., Paderes G. D., Sinclair S. // *Pure Appl. Chem.* — 1990. — Vol. 62. — P. 1921.
35. Bures M. G., Jorgensen W. L. // *J. Org. Chem.* — 1988. — Vol. 53. — P. 2504.
36. Gasteiger J., Hutchings M. G., Christoph B., Gann L., Hiller C., Low P., Marsili M., Saller H., Yuki K. // *Topics Curr. Chem.* — 1987. — Vol. 137. — P. 19.
37. Gasteiger J., Marsili M., Hutchings M. G., Saller H., Low P., Rose P., Rafeiner K. // *J. Chem. Inf. Comput. Sci.* — 1990. — Vol. 30. — P. 467.
38. Balaban A. T. // *Rev. Chim.* — 1967. — Vol. 12. — P. 875.
39. Barabas A., Balaban A. T. // *Rev. Chim.* — 1974. — Vol. 19. — P. 1927.
40. Dugundji J., Ugi I. // *Topics Curr. Chem.* — 1973. — Vol. 39. — P. 19.
41. Ugi I., Gillespie P. // *Angew. Chem., Int. Ed. Engl.* — 1971. — Vol. 10. — P. 914.
42. Hendrickson J. B. // *Angew. Chem., Int. Ed. Engl.* — 1974. — Vol. 13. — P. 47.
43. Зедуров H. C., Трав C. C. // *ЖОрХ*. — 1975. — Т. 11. — С. 1785.
44. Zefirov N. S. // *Acc. Chem. Res.* — 1987. — Vol. 20. — P. 237.
45. Fujita S. // *J. Chem. Inf. Comput. Sci.* — 1986. — Vol. 26. — P. 205.
46. Fujita S. // *J. Chem. Soc. Perkin Trans. II*. — 1988. — P. 597.
47. Roberts D. C. // *J. Org. Chem.* — 1978. — Vol. 43. — P. 1473.
48. Arens J. F. // *Rec. Trav. Chim.* — 1979. — Vol. 98. — P. 155.
49. Kvasnicka V. // *Coll.* — 1984. — Vol. 49. — P. 1090.
50. Kvasnicka V., Kratochvil M., Koca J. // *Coll.* — 1983. — Vol. 48. — P. 2284.
51. Wilcox C. S., Levinson R. A. *Artificial Intelligence Applications in Chemistry* / Eds. T. H. Pierce, B. A. Hohne. — Washington: Am. Chem. Soc. — 1986. — P. 209. (Рус. перевод М.: Мир, 1988. — С. 238.)
52. Hendrickson J. B. // *Rec. Trav. Chim.* — 1992. — Vol. 111. — P. 323.
53. Zefirov N. S., Tratch S. S. // *Anal. Chim. Acta.* — 1990. — Vol. 235. — P. 115.
54. Baskin I. I., Palyulin V. A., Zefirov N. S. // *J. Chem. Inf. Comput. Sci.*, в печати.
55. Bauer J. // *Tetrahedron Comput. Methodol.* — 1989. — Vol. 2. — P. 269.
56. Herges R., Hoock C. // *Science*. — 1992. — Vol. 255. — P. 711.
57. Barone R., Arbelot M., Chanon M. // *Tetrahedron Comput. Methodol.* — 1988. — Vol. 1. — P. 3.
58. *Comprehensive Heterocyclic Chemistry* / Eds. A. R. Katritzky, C. W. Rees. — Oxford: Pergamon, 1984. — Vols. 1...8.
59. Бабаев Е. В., Зедуров H. C. // *XTC*. — 1992. — № 6. — С. 808.
60. Babaev E. V., Lushnikov D. E., Zefirov N. S. // *J. Amer. Chem. Soc.* — 1993. — Vol. 115. — P. 2416.
61. Barone R., Boch A., Chanon M., Metzger J. // *Comput. & Chem.* — 1979. — Vol. 3. — P. 83.
62. Barone R., Chanon M. // *Heterocycles*. — 1981. — Vol. 16. — P. 1357.
63. Barone R., Chanon M., Cadiot M., Cense J. M. // *Bull. soc. chim. Belge*. — 1982. — Vol. 91. — P. 333.
64. Barone R., Camps P., Elguero L. // *Ann. Quim.* — 1979. — Vol. 75. — P. 736.
65. Barone R., Chanon M., Metzger J. // *Chimia*. — 1978. — Vol. 32, N 6. — P. 216.
66. Barone R., Chanon M., Metzger J. // *Tetrah. Lett.* — 1974. — P. 2761.
67. Dubs P., Stuessi R. // *Synthesis*. — 1976. — N 10. — P. 696.
68. Barone R., Chanon M. // *Nouv. J. Chim.* — 1978. — Vol. 2. — P. 659.
69. Gasteiger J., Jochum C. // *Topics Curr. Chem.* — 1978. — Vol. 74. — P. 93.
70. Doenges R., Groebel B. T., Nickelsen H., Sander J. // *J. Chem. Inf. Comput. Sci.* — 1985. — Vol. 25. — P. 425.
71. Лушиников Д. Е., Бабаев Е. В., Цитовский С. В., неопубл. данные.
72. Hendrickson J. B. // *J. Amer. Chem. Soc.* — 1977. — Vol. 99. — P. 5439.
73. Hendrickson J. B., Grier D. L., Toczko A. G. // *J. Amer. Chem. Soc.* — 1985. — Vol. 107. — P. 5228.
74. Hendrickson J. B. // *Acc. Chem. Res.* — 1986. — Vol. 19. — P. 274.
75. Hendrickson J. B., Toczko A. G. // *J. Chem. Inf. Comput. Sci.* — 1989. — Vol. 29. — P. 137.

76. *Hendrickson J. B., Miller T. M. // J. Amer. Chem. Soc. — 1991. — Vol. 113. — P. 902.*
77. *Hendrickson J. B., Parks C. A. // J. Chem. Inf. Comput. Sci. — 1992. — Vol. 32. — P. 209.*
78. *Moreau G. // Nouv. J. Chim. — 1978. — Vol. 2. — P. 187.*
79. *Pollet P. // New J. Chem. — 1990. — Vol. 14. — P. 957.*
80. *Ugi I., Bauer J., Bley K., Dengler A., Fontain E., Knauer M., Lohberger S. // J.Mol.Struct. (THEOCHEM). — 1991. — Vol. 230. — P. 73.*
81. *Dengler A., Fontain E., Knauer M., Stein N., Ugi I. // Rec. Trav. Chim. — 1992. — Vol. 111. — P. 262.*
82. *Fontain E., Bauer J., Ugi I. // Z. Naturforsch. — 1987. — Vol. 42b. — P. 889.*
83. *Bauer J., Herges R., Fontain E., Ugi I. // Chimia. — 1985. — Vol. 39. — P. 43.*

*Московский государственный университет  
им. М. В. Ломоносова, Химический факультет,  
Москва 119899*

*Поступило в редакцию 16.06.93*