

Е. В. Бабаев, Н. С. Зефирюв

МОЛЕКУЛЯРНЫЙ ДИЗАЙН ГЕТЕРОЦИКЛОВ I. РЕЦИКЛИЗАЦИОННЫЕ ГРАФЫ И СТРУКТУРНАЯ ИЕРАРХИЯ ТРАНСФОРМАЦИЙ ГЕТЕРОЦИКЛОВ*

(ОБЗОР)

Предложен новый способ моделирования процессов трансформации цикла гетероциклических систем с помощью рециклизационных графов, вершины и ребра которых отвечают только атомам и скелетным связям, входящим в исходный или конечный гетероцикл. Анализ структуры таких графов и использование различных типов меток для ребер и вершин позволяют впервые выстроить удобную и гибкую иерархию рециклизаций по принципу структурного подобия их рециклизационных графов. Эффективность подхода иллюстрируется примерами классификации известных взаимопревращений азолов и азинов и прогнозом новых рециклизаций.

ВВЕДЕНИЕ

Настоящий цикл работ ставит своей целью показать эффективность использования нечисленных математических методов (теории графов, комбинаторики, топологии) в дизайне гетероциклических структур, планировании путей их синтеза и реакций. Трактую термин «дизайн» в широком смысле, мы предполагаем сочетать разработку новых теоретических конструкций и компьютерных программ для нужд химии гетероциклов с их экспериментальной реализацией и проверкой.

Данное сообщение посвящено проблеме структурной классификации весьма важного и широкого семейства реакций — процессов трансформации цикла (или рециклизаций) гетероциклических систем. Реакции рециклизации составляют едва ли не самую яркую страницу химии гетероциклов; эти реакции, зачастую весьма элегантные и обнаруживаемые случайно, нередко приводят к циклическим структурам с неожиданным распределением гетероатомов и заместителей либо продуктам, получать которые другими путями трудно или невозможно. Хорошо известными примерами рециклизаций, могут служить именные реакции типа Юрьева, Цинке—Кенига, Гафнера или перегруппировки Димрота, Корнфорта, Боултона—Катрицкого, Коста—Сагитуллина и др. В настоящее время склонность к трансформации цикла обнаружена у подавляющего большинства известных гетероциклов, в особенности у гетероароматических. Механизмы такого рода превращений разнообразны, и нередко оказывается весьма затруднительным предположить структурную схему превращения, рассматривая лишь начальный и конечный гетероцикл. Несмотря на огромный фактический материал в этой области и изобилие обзоров, охватывающих большие классы рециклизаций (например, [1—24]), следует признать, что общая рациональная классификация трансформаций гетероколец по структурному принципу отсутствует. Как следствие, нередко оказывается трудно установить, например, реальную степень новизны рециклизации или перегруппировки, декларируемой в качестве «новой». Требуется, очевидно, первоначально каким-то образом определить степень и меру структурного подобия уже известных реакций.

* Пленарный доклад на V Всесоюзной конференции по химии азотсодержащих гетероциклических соединений. Черноголовка, 22—25.10.1991.

В большинстве существующих обзоров по рециклизациям вопрос о единой структурной классификации практически не возникал; лишь в единичных обзорных работах затрагивались частные аспекты этой проблемы. В классической монографии Ван дер Пласа, например, рециклизации не столько классифицированы, сколько упорядочены по размерам исходного и конечного циклов, числу и природе гетероатомов [1]. Между тем, еще в работах Швайки [2, 3] по гидраинолизу азолов предлагалось использовать структурный признак — размер фрагмента исходного цикла, используемого для построения нового гетерокольца — для выявления структурного подобия рециклизаций. Сходная идея, развиваемая Балабаном, нашла применение для классификации разнообразных типов трансформаций пириллиевых солей [4, 5], а также солей 1,3-тиазиния [6]. В недавнем обзоре Л'Аббе [7] предложил иной (альтернативный) способ удобной классификации типов перегруппировок азолов — по размеру имеющейся в них боковой цепи, вовлекаемой в формирование нового цикла. (Например, в перегруппировках типа Димрота, Корифорта, Боултона-Катрицкого эта величина равна 1, 2, 3). В других обзорах (например, [8—24]) обсуждаются обычно либо частные трансформации тех или иных классов гетероциклических систем, либо обобщенные схемы реакций, объединяющих гетероциклы с разным распределением гетероатомов или аннелированных ядер.

Заметим, что наиболее разработанные из упомянутых классификационных схем используют, в конечном итоге, принцип структурного подобия гетероциклических молекул (или разных типов содержащихся в них фрагментов), но не подобия реакций. Между тем, за последние 10...15 лет был достигнут существенный прогресс в разработке и создании математических моделей, в частности, использование языка теории графов, для описания и структурной классификации химических реакций [25—32].

Ключевая идея подходов, связанных с классификациями реакций на базе теории графов заключается в замене привычного химического уравнения реакции (содержащего, очевидно левую и правую части) некоторым иным объектом — единой структурной диаграммой перераспределения связей в ходе реакции; для получения такой диаграммы необходимо сначала установить соответствие между атомами (и связями) молекул левой и правой части химического уравнения, а затем осуществить "суперпозицию" исходных соединений и продуктов реакции [25, 28, 30]. Результирующая диаграмма — граф перераспределения связей — концентрирует всю информацию о разрывающихся или вновь образующихся в ходе реакции связях (остающиеся неизменными атомы и связи обычно не рассматриваются). Именуемые в разных подходах по-разному, такие «графы реакций» в дальнейшем использовались с различными целями, например, для установления степени структурного подобия реакций, их удобного кодирования, решения задач планирования синтеза соединений или планирования новых реакций [27, 28, 31, 32].

Впервые эффективность использования диаграмм перераспределения связей для классификации рециклизаций гетероциклов была отмечена нами ранее [33]; в дальнейшем, анализируя основные структурные тенденции трансформации цикла азолов [34], авторы пришли к выводу о необходимости создания специального типа графов — рециклизационных графов (или графов перераспределения циклических связей), ориентированных на решение задач химии гетероциклов. В настоящей работе рассмотрено обобщение этого подхода, приводящее к удобной в теоретическом (и методическом) плане иерархической структурной классификации рециклизаций. Такая классификация может быть эффективно использована как для систематизации уже имеющегося обширного экспериментального материала по трансформации циклов, так и для прогноза принципиально новых — беспрецедентных — примеров рециклизаций.

Будем считать рециклизацией (или трансформацией цикла), как это обычно принято, любое превращение гетерокольца, включающее стадии раскрытия исходного цикла и замыкания конечного цикла в любой последовательности. Ограничимся рассмотрением моноциклических гетероциклов, добавив к ним из числа конденсированных систем лишь те, в которых аннелированные циклы имеют не более чем одну общую связь (например, обычный тип бензаннылирования как в индоле или акридине). Назовем трансформацию цикла (ТЦ) простой (ПТЦ) если выполнены следующие три условия:

1. В ходе рециклизации не происходит промежуточное образование каких-либо иных циклов (кроме формирования конечного цикла);

2. В ходе рециклизации не имеет места промежуточная миграция атомов или групп атомов;

3. Исходный цикл превращается не более чем в один цикл конечной гетероциклической структуры.

Подавляющее большинство известных ТЦ являются простыми, в частности, большое семейство ANRORC-реакций [1], «моноциклические» перегруппировки азолов [7, 14], большинство межмолекулярных реакций циклоприсоединения-элиминирования (типа реакции Кондратьевой [19]). Рассматриваемые ниже рециклизации на схемах 1—3 дают некоторые примеры типичных ПТЦ, для которых условия 1—3 выполняются. К числу менее распространенных ТЦ не являющихся ПТЦ (нарушено одно из условий 1—3) относятся, например, ряд фотохимических перегруппировок [1, 12, 21], некоторые процессы внутримолекулярного циклоприсоединения (например, [19, 22, 35]), отдельные типы трансформаций с глубокой перестройкой исходного цикла (типа превращения пиридиниевых солей в индолы [36]) и некоторые другие рециклизации (см. также [34]). В настоящем сообщении мы ограничимся рассмотрением лишь ПТЦ.

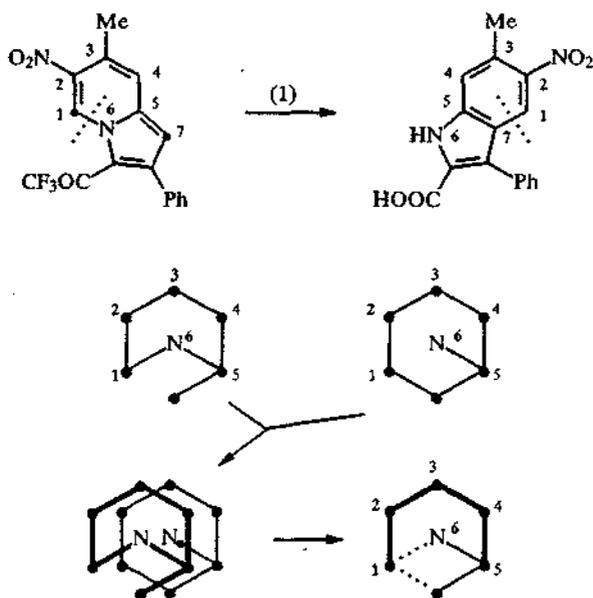
В самой общей форме центральная идея нашего подхода базируется на весьма простых и ясных химическом принципах: коль скоро механизм рециклизации известен, то скелет исходного цикла можно легко обнаружить среди атомов и связей конечных продуктов и наоборот, скелет конечного цикла отвечает вполне конкретным атомам и связям исходных соединений. Сконцентрировав внимание только на таких скелетных подструктурах циклов (и игнорируя все остальные детали строения) можно резко упростить структурное уравнение реакции. Очевидно, что механизм реакции задает строгое (взаимно-однозначное) соответствие между выделенными "циклическими" атомами (и связями) левой и правой части уравнения. Как результат, именно эти атомы и связи можно отобразить для суперпозиции исходных и конечных структур (так, как это делается в рассмотренных выше подходах) с целью получения структурного уравнения (или графа) рециклизации.

Дадим более строгие формулировки. Определим для произвольной ПТЦ (с заранее известным механизмом) два типа графов — молекулярные графы (для исходных и конечных соединений, участвующих именно в этой реакции) и рециклизационный граф (для любой реакции типа ПТЦ).

В качестве *молекулярного графа* M_s исходных реагентов выберем из совокупности атомов и связей левой части химического уравнения лишь те V атомов (вершин M_s -графа) и те R_s скелетных связей (ребер M_s -графа), которые *либо присутствуют в начальном цикле, либо включаются в конечный цикл*. Определим аналогичным образом молекулярный граф конечных продуктов M_f (с V вершинами и R_f ребрами) как те же самые V атомов и лишь те R_f скелетных связей, которые *имеются в конечном цикле, либо присутствовали в начальном цикле*. При этом оставим в графах M_s и M_f в качестве меток лишь символы гетероатомов и отбросим все остальные символы — водородных атомов, кратных связей и любых заместителей (включая аннелированные ядра). Хотя для конструирования графов M_s и M_f

выбираются одни и те же вершины, ясно, что число и/или распределение ребер в этих графах различно. Пример выделения молекулярных M_5 - и M_7 -графов 1a,б для ПТЦ (1) [37] представлен на схеме 1; в этом примере лишь часть аннелированного пиррольного фрагмента исходного индолизина (или результирующего индола) выбирается для конструирования графов M_5 и M_7 .

Схема 1



Занумеровав некоторым образом вершины M_5 -графа, припишем соответствующие номера вершинам M_7 -графа согласно механизму реакции. Рассмотрим мысленную суперпозицию структур графов M_5 и M_7 таким образом, чтобы пары вершин с одинаковыми номерами отождествлялись в одну вершину, а соответствующие пары ребер — в некоторое новое ребро. Указанную суперпозицию M -графов определим как новый *рециклизационный граф* (РГ) содержащий V вершин; при этом ребра, присутствовавшие в M -графах и отображаемые в новые ребра рециклизационного графа, будем обозначать в РГ следующим образом:

пунктирной линией (ребром) — если ребро, присутствует лишь в одном из M -графов, но отсутствует в другом,

жирной линией — если ребро входит в состав и начального и конечного циклов M -графов,

обычной линией — если ребро входит в состав лишь одного из циклов M -графов.

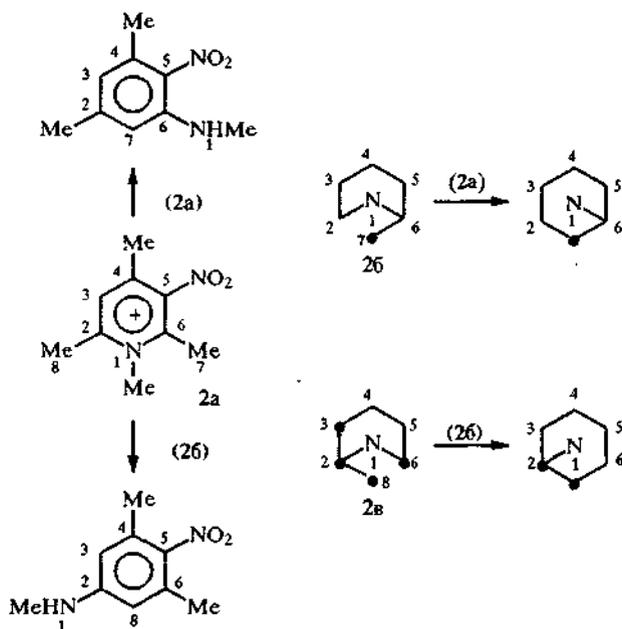
Пример суперпозиции M -графов 1a,б для ПТЦ (1) и результирующий рециклизационный граф 1в приведены на схеме 1*. С формальной точки зрения РГ есть некий помеченный граф, содержащий метки на вершинах (символы гетероатомов) и ребрах (три указанных типа или «цвета» ребер — пунктирный, жирный и обычный). Включение всех трех типов помечений для ребер в одну и ту же структуру РГ позволяет в рамках единой диаграммы реакции проследить за судьбой каждой из циклических скелетных связей гетероцикла в ходе рециклизации.

* Для получения РГ можно поступить еще проще: сразу пометив в M_5 -графе жирными ребрами фрагмент цикла, вовлекаемый в формирование нового цикла и пунктирными — разрывающиеся связи цикла, добавив новые пунктирные ребра, необходимые для построения цикла M_7 -графа.

Из алгоритма конструирования РГ следует, что структура рециклизационного графа не зависит от того, какой именно молекулярный граф (M_S или M_I) взят для суперпозиции первым, т. е. рециклизационные графы прямой и обратной реакций совпадают.

Подчеркнем, что выбор атомов в качестве вершин для построения как РГ, так и М-графов определяется их принадлежностью как исходному, так и конечному циклам. Например, в реакциях (2а) и (2б) [15] на схеме 2 выбор фрагмента в качестве молекулярного M_S -графа диктуется направлением реакции, т. е. тем, какая из метильных групп вовлекается в построение нового цикла. В результате, одна и та же исходная структура 2а — в зависимости от типа трансформации — представлена двумя различными M_S -графами 2б,в.

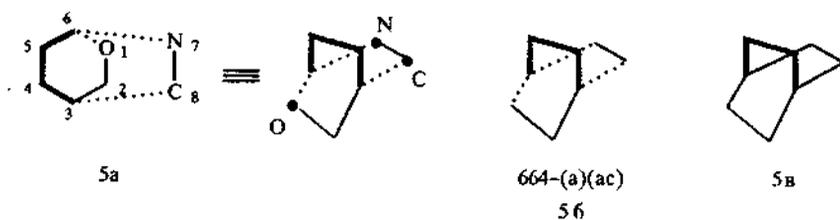
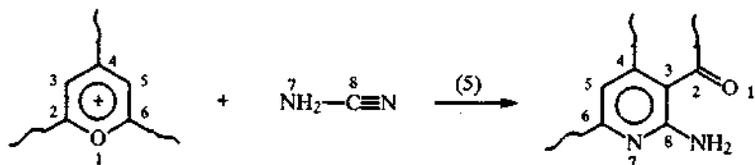
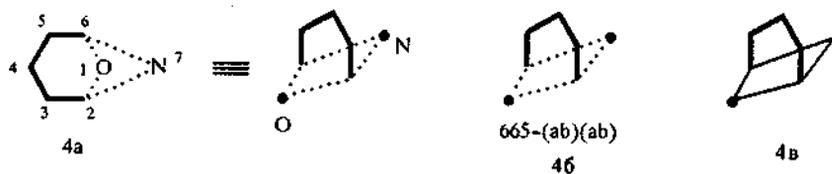
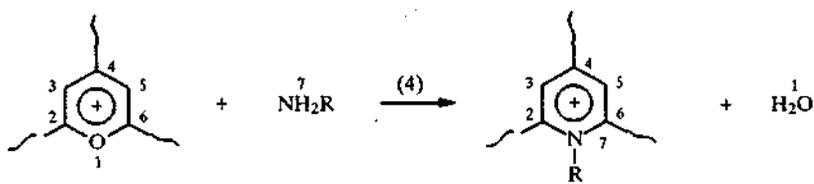
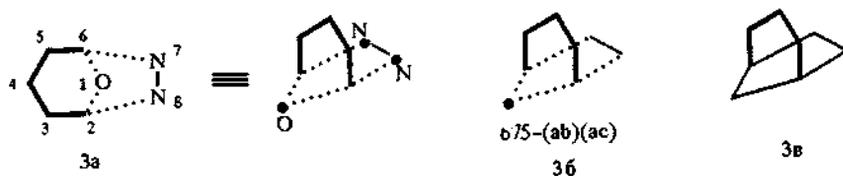
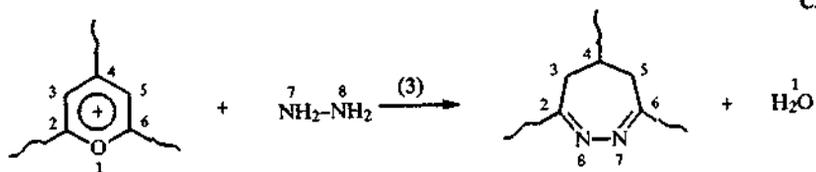
Схема 2

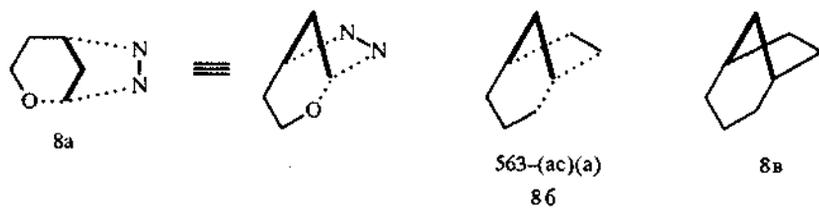
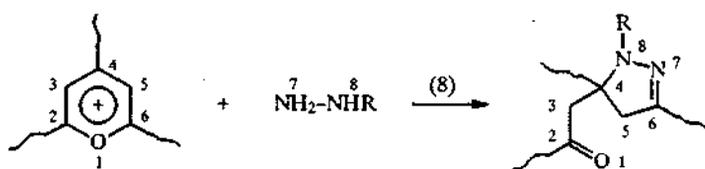
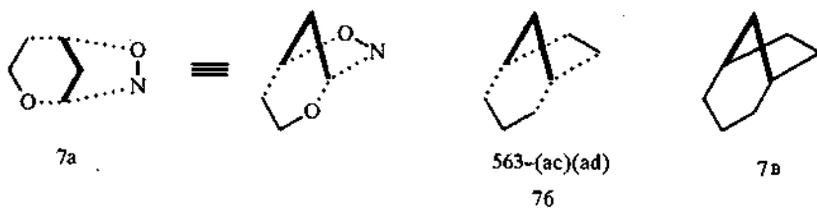
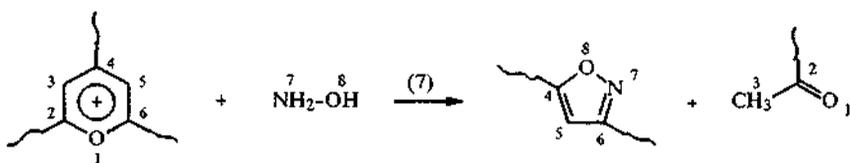
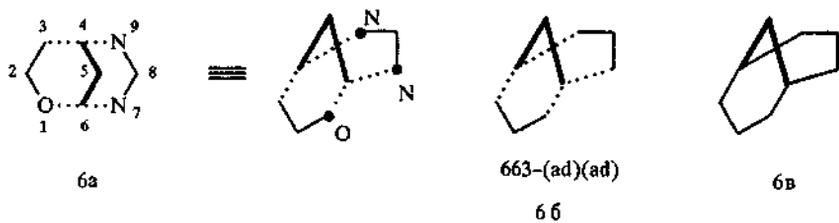
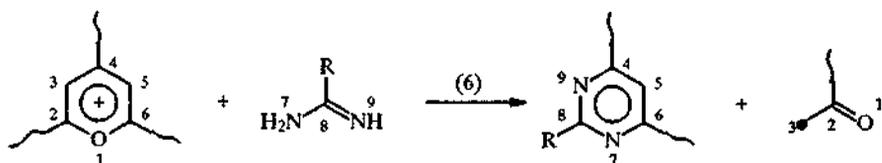


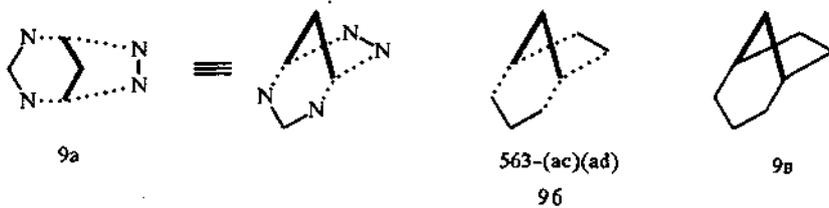
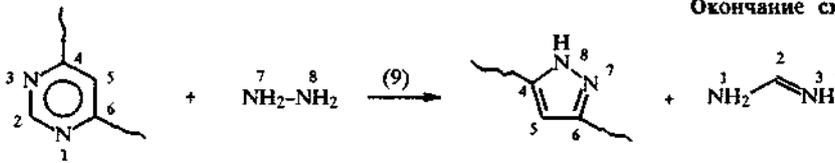
РЕЦИКЛИЗАЦИОННЫЕ ГРАФЫ И СТРУКТУРНОЕ ПОДОБИЕ РЕАКЦИЙ

Следуя приведенному выше алгоритму выбора М-графов и конструирования рециклизационных графов легко построить РГ для любых известных рециклизаций, являющихся простыми, например, для ставших уже хрестоматийными трансформаций солей пириллия (3)—(8) или пиримидина (9), схема 3; результирующие РГ представлены на схеме диаграммами 3а—9а. Эти реакции одновременно иллюстрируют процессы трансформации с расширением (3), сокращением (7)—(9) или сохранением (4)—(6) размера исходного цикла.

Виды рециклизаций. Заметим, что M_S - и M_I -графы связаны лишь для перегруппировок (например, М-графы 1в, 2б,в), тогда как во всех остальных случаях (т.е. для процессов включения внешнего реагента и/или элиминирования фрагмента исходного цикла, иллюстрируемых реакциями (3)—(9)) хотя бы один из М-графов обязан состоять более чем из одного компонента. Выберем число компонент в М-графах в качестве естественного признака для классификации ПТЦ по видам. Обозначая исходный и конечный гетероцикл символами А и В, а внешние фрагменты, внедряемые в скелет цикла (или элиминируемые из него) символами X, Y..., можно сразу выделить три важнейших вида рециклизаций, символически представленные схемами







A → B (т. е. перегруппировки, например (1) или (2a))

A + X → B или A → B + X (реакции с внедрением или элиминированием фрагментов цикла, например (5), (8))

A + X → B + Y (реакции с внедрением и элиминированием фрагментов цикла, например (3), (4), (6), (7), (9))

Очевидно, что ПТЦ одного и того же вида содержат равное число пунктирных ребер в структуре РГ. Заметим также, что принадлежность к одному виду определяется лишь связностью М-графов и не имеет ничего общего с кинетическим порядком реакции, (связный М₅-граф, например, может отвечать какой-либо бимолекулярной реакции, скажем ANRORC-типа).

G₀-, G₁- и G₂-графы. Сопоставление РГ на схеме 3 показывает, что рециклизационные графы отдельных ПТЦ могут отличаться друг от друга либо структурой самих графов, либо числом и распределением в них меток вершин и/или пунктирных пометок ребер. Именно этим фактом можно воспользоваться в качестве главного критерия для выявления степени структурного подобия разных рециклизаций. С этой целью рассмотрим три различных типа рециклизационных графов:

Рециклизационный граф, определенный выше как суперпозиция М-графов и содержащий символы гетероатомов и три сорта ребер, назовем графом G₂-типа.

Определим G₁-граф как G₂-граф, в котором метки вершин (символы гетероатомов) опущены.

Определим G₀-граф как G₁-граф, в котором опущены пунктирные метки ребер.

Примеры G₁- и G₀-графов (соответственно диаграммы 3б—9б и 3в—9в) приведены для каждой из рециклизаций (3)—(9) на схеме 3. Смысл введения дополнительных конструкций РГ очевиден: сопоставляя между собой различные типы РГ можно выявить различные аспекты структурного подобия реакций имеющих одинаковые графы G₀, G₁ или G₂.

Семейства, классы и типы рециклизаций. Рассмотрим самый близкий тип структурного подобия рециклизаций — совпадение соответствующих им G₂-графов. В этом случае будем говорить, что рециклизации принадлежат к одному и тому же *семейству*. Тип семейства однозначно задается скелетом и положением гетероатомов в исходном и конечном гетероциклах (т. е. фиксированной парой М-графов), поэтому любые вариации в степени ненасыщенности, таутомерном строении, числе и природе заместителей (и/или конденсированных ядер) не выводят за пределы семейства. Примером реакций одного семейства могут служить перегруппировки (1), (2a) и (2б).

Следующий тип структурной близости — идентичность (точнее изоморфизм) G_1 -графов; будем говорить, что ПТЦ с одинаковыми G_1 -графами образуют один и тот же класс. В одинаковый класс попадают структурно сходные ПТЦ разных гетероциклов, например, рециклизации пириллиевых солей в изоксазолы и пиримидинов в пиразолы, реакции (7) и (9) с изоморфными G_1 -графами 7в, 9в.

Наконец к одному и тому же структурному типу отнесем такие ПТЦ, G_0 -графы которых совпадают. Рециклизационный тип — наиболее грубый тип сходства ПТЦ; определяющим фактором такого подобия служит совпадение размера фрагмента (помечаемого в G_0 -графе жирными ребрами) общего для начального и конечного циклов. Понятие типа удобно использовать, сопоставляя ПТЦ разного вида, например, сравнивая рециклизации с включением и/или элиминированием фрагментов цикла с обычными перегруппировками. Как легко видеть, перегруппировки (1) и (2а,б) принадлежат к тому же типу, что и реакция (4); тип рециклизаций (7) — (9) также одинаков.

Качественные рассуждения о степени подобия ПТЦ удастся сформулировать в строгой форме на основе анализа структурных особенностей разных типов рециклизационных графов.

СТРУКТУРА, СИММЕТРИЯ И КОДЫ РЕЦИКЛИЗАЦИОННЫХ ГРАФОВ

Поскольку РГ отражают различную степень структурного родства рециклизаций, зададимся следующей проблемой: как именно связаны между собой структуры графов G_0 , G_1 и G_2 , т.е. типы, классы и семейства рециклизаций. Графы G_0 и G_1 конструировались выше путем удаления определенных типов меток (вершинных или реберных) из структуры графа G_2 . Указанную процедуру можно инвертировать, рассматривая G_1 -граф как реберно помеченный граф G_0 , а граф G_2 как вершинно помеченный граф G_1 .

Таким образом, связь между разными РГ носит чисто комбинаторный характер и представляет собой вариант задач на расстановку меток (реберных, а затем и вершинных) в графах G_0 . Число неэквивалентных реберных помечений в графе G_0 определяет количество теоретически возможных G_1 -графов, а число неэквивалентных размещений символов гетероатомов (меток вершин) в графах G_1 — число G_2 -графов. Задачи такого типа являются хорошо изученными (ср. перечисление изомеров замещения [38]) и решаются, например, с использованием леммы Бернсайда [38—40] для случаев, когда известно как именно нужно размещать заданное число меток и каков тип симметрии исходного графа, т.е. в нашем случае — графа G_0 .

Структура и коды графов G_0 . Нетрудно показать [34, 41], что графы G_0 имеют вполне определенный тип структуры, а именно, являются помеченными бициклами или графами с цикломатическим числом равным двум [42]. Иными словами, любой из них представим в виде некоторой пары аннелированных циклов, где фрагмент (мост) общий обоим циклам помечен жирными линиями (ребрами). Примером такого представления G_0 -графов могут служить диаграммы 4в, 5в и 6в, представляющие три разных типа сочленения 6-членных циклов. В общем случае графы с цикломатическим числом 2 содержат три цикла, между тем лишь два из них линейно независимы [42]. (В этом случае линейно-зависимым может считаться любой из трех циклов, например, 10-членный периметр в «бициклическом» молекулярном графе нафталина.) Условимся считать независимыми в графах G_0 лишь те циклы которые содержат «жирную» перемычку (их всегда только два) и именно их называть *циклами* G_0 -графа. Размеры таких аннелируемых циклов РГ G_0 определяются, очевидно, размерами циклов M -графов, т.е. исходного и конечного гетероциклов.

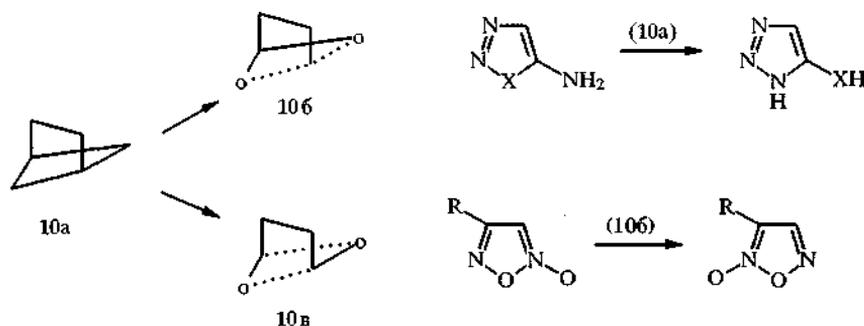
Сравнительно простой тип структуры графов G_0 допускает введение удобного кодирования рециклизационных типов [34]. Пусть K и L — размеры исходного и конечного циклов ПТЦ, а N — число атомов во фрагменте, общем для обоих циклов (т. е. число вершин в мосте G_0 -графа, помеченного жирными линиями). Тогда упорядоченная тройка чисел (или вектор) вида KLN однозначно определяет структуру бицикла G_0 и отвечающего такому графу рециклизационного типа. (Например код 564 означает взаимную трансформацию 5- и 6-членных гетероциклов, так что фрагмент из четырех атомов исходного цикла появляется в новом цикле.)

Перечисление и коды G_1 -графов. Сравнивая структуры G_0 - и G_1 -графов, заметим, что граф G_1 получается путем некоторого помечения части ребер графа G_0 пунктирными линиями. Выше пунктирное ребро было отождествлено с разрывающейся или образующейся скелетной связью цикла. Например, в случае перегруппировок графы G_1 содержали пару пунктирных ребер, с очевидностью находившихся в разных *циклах*. Поэтому если задаться целью перечислить все возможные графы G_1 (классы ПТЦ), отвечающие перегруппировкам с заданной структурой графа G_0 (заданным типом ПТЦ), то следует разместить в G_0 -графе пару пунктирных меток всеми возможными способами, лишь бы эти метки попадали в разные *циклы**.

В результате можно легко вывести (даже вручную) число теоретически возможных классов перегруппировок по заданному типу. Например, если задан G_0 -граф с кодом 554 (диаграмма 10 на схеме 4), то то могут существовать лишь два неэквивалентных графа G_1 — 10а,б, содержащие по пунктирному ребру в каждом из циклов, т. е. лишь два класса перегруппировок такого типа. Оба этих класса хорошо известны в литературе как перегруппировки типа Димрота и перегруппировки фуроксанов [21] и иллюстрируются реакциями (10а) и (10б) на схеме 4.

Классы рециклизаций удобно характеризовать простыми *кодами* соответствующих им G_1 -графов [34, 41]. Для этого достаточно указать точное положение пунктирных связей в предшествующем графе G_0 . В случае перегруппировок таких пунктирных меток, расположенных в разных циклах G_0 -графа всего две. Занумеруем последовательность обычных ребер в меньшем из циклов графа G_0 : ближайшее к жирному мосту ребро обозначим символом a , следующее за ним ребро — символом b и т. д. Аналогичным образом занумеруем теми же символами $a, b, c...$ обычные ребра в другом (большем) из циклов, начиная с ребра, смежного с a -ребром первого цикла. Результирующий код вида $KLN-ij$ однозначно определяет класс перегруппировки**. Рассмотренным выше примерам реакций (10а) и (10б) соответствуют коды 554- aa и 554- ab .

Схема 4



* Естественно, что помечать пунктиром можно лишь «обычные» ребра графа G_0 , поскольку помеченные жирными линиями ребра, по определению, суть неизменные скелетные связи.

** Для случая $K=L$ символы i, j из набора $a, b, c...$ должны быть лексикографически упорядочены.

Ранее были выведены комбинаторные формулы для числа G_1 -графов в связи со структурой предшествующих им G_0 -графов и затабулированы все возможные рекомендуемые коды вида $KLN-ij$ для перегруппировок [34]. Попытаемся обобщить принцип кодирования классов ПТЦ на рециклизации с включением или элиминированием фрагментов. В общем случае код рециклизационного класса будет определяться распределением более чем двух пунктирных меток между циклами исходного G_0 -графа и может быть записан в виде $KLN-(ikjkk\dots)(ijlkl\dots)$, где символы в скобках представляют собой определенную по тем же правилам (и лексикографически упорядоченную) последовательность пунктирных ребер в K - и L -членных циклах графа G_0 . В химической практике общее число букв в коде не превышает четырех (две связи рвутся в исходном гетероцикле и две — образуются в конечном), т.е. максимально длинные коды для рециклизаций имеют вид $KLN-(ikjk)(ijl)$. Коды графов G_1 для рассмотренных рециклизаций (3)–(9) указаны на схеме 3.

G_2 -графы. Следующий тип РГ — графы G_2 — представляют собой вершинно помеченные графы G_1 . Для того чтобы вычислить число всевозможных G_2 -графов (семейств ПТЦ) по заданному G_1 -графу (классу ПТЦ) необходимо знать симметрию графа G_1 (определяемую его группой автоморфизмов) и число меток (гетероатомов) каждого типа, включая углерод в качестве равноправной метки. В этом случае для подсчета неэквивалентных помечений, т.е. G_2 -графов, можно эффективно использовать лемму Бернсайда [38–40]. Не останавливаясь подробно на вычислениях заметим, например, что число теоретически возможных семейств перегруппировок типа Димрота в ряду азолов (если ограничиться лишь тремя метками, к примеру C , N и S) составляет 486; пример лишь одного из таких семейств дает реакция (10а).

Выше отмечалось, что G_2 -графы (полученные путем суперпозиции M -графов) для прямой и обратной реакции совпадают. Справедливо и обратное: если задан рециклизационный граф G_2 , то ему соответствует вполне определенная пара M -графов, т.е. пара реакций — прямая и обратная, но не направление рециклизации*. Для получения пары M -графов из данного РГ G_2 можно предложить следующий простой алгоритм. Выбирается любой из циклов графа G_2 и пунктирные ребра только этого цикла меняются на обычные, одновременно с этим пунктирные ребра во втором из циклов удаляются, что дает первый из M -графов. Повторная процедура начиная со второго цикла порождает второй из M -графов, определяющих пару реакций.

ИЕРАРХИЧЕСКАЯ КЛАССИФИКАЦИЯ ТРАНСФОРМАЦИЙ ГЕТЕРОЦИКЛОВ

Рассмотренная выше в самых общих чертах структурная взаимосвязь различных рециклизационных графов позволяет впервые предложить классификацию трансформаций гетероциклов по иерархическому принципу.

Рассмотрим последовательно уровни этой иерархии:

1. На вершине иерархического дерева задан лишь общий «архетип»: размеры начального и конечного гетероциклов в виде пары (K, L) .

2. На следующем уровне иерархии учитывается что гетероциклы могут иметь различную величину (N) фрагмента, отвечающего неизменной части начального цикла, вовлекаемого в формирование конечного цикла. Позволив величине N варьировать от единицы до $K \leq L$, можно получить всевозможные способы аннелирования пары K - и L -членных циклов, определяющие помеченные G_0 -графы с кодами KLN или рециклизационные типы.

* В ряде подходов [28, 32] реакционный граф содержит особые типы меток (например, стрелки или перечеркнутые линии), указывающих какие именно связи рвутся, а какие — вновь образуются; хотя такие метки можно естественно включить в структуру G_1 - и G_2 -графов, это приведет к потере наглядности и в настоящем сообщении использоваться не будет.

3. На новом уровне учитывается число компонент, на которые распадается исходный цикл или из которых формируется конечный цикл, т.е. определяется вид рециклизации. Ясно, что тем самым устанавливается число разрываемых и вновь образующихся связей, т.е. суммарное число пунктирных меток, которые можно разместить в структуре РГ G_0 .

4. Для фиксированного типа и вида ПТЦ известен размер фрагмента исходного цикла, остающегося неизменным и общее число исчезающих/возникающих скелетных связей циклов. Это позволяет на основе чисто комбинаторных соображений перебрать все мыслимые способы разрыва и образования связей, т.е. классы рециклизаций (G_1 -графов) заданного типа (и вида) путем расстановки определенного числа пунктирных меток на ребрах в каждом из циклов графа G_0 .

5. Коль скоро известно, в каких именно положениях начального цикла рвутся связи, и образование каких именно скелетных связей отвечает за формирование нового цикла, нетрудно на основе соображений симметрии перебрать теоретически возможные начальные и конечные гетероциклические системы (семейства рециклизаций), отвечающих заданному классу. Это достигается расстановкой гетероатомных меток на вершинах заданного графа G_1 приводящее к набору неизоморфных графов G_2 .

6. Заданное семейство рециклизаций (с точностью до направления реакции) отвечает строго определенным парам гетероциклических систем, содержащих лишь те заместители, которые необходимы для построения конечного гетерокольца или ранее входили в циклический остов исходного гетероцикла (пары M -графов). На этом уровне можно выбрать конкретное направление рециклизации.

7. На самом нижнем уровне заданная парой M -графов символическая запись трансформации гетероцикла отвечает набору конкретных химических превращений, компоненты которых различаются степенью ненасыщенности, таутомерией, наличием заместителей, аннелированных циклов и т.д.

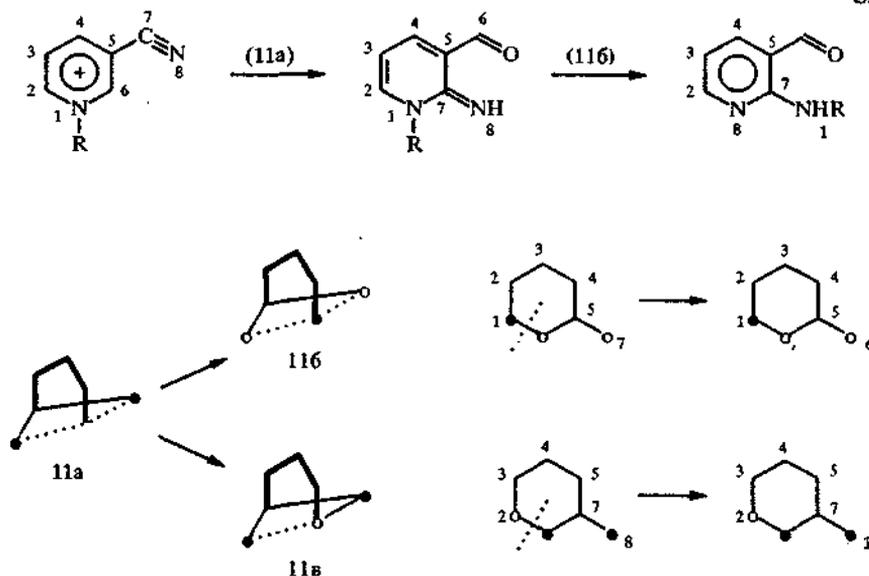
Таким образом, иерархичность предлагаемой классификации заключается в том, что число и структура рециклизационных графов низших уровней жестко детерминированы структурой и симметрией РГ высших уровней. Неразличимость рециклизаций на каком-либо уровне снимается при переходе на более низкий уровень.

НАПРАВЛЕНИЕ ГЕТЕРОЛИЗА СВЯЗЕЙ

Рассмотренная классификация ПТЦ носит структурный характер; тем не менее, уровни ее иерархии можно эффективно дополнять эмпирическими данными о деталях механизмов ПТЦ, в частности, химически значимой информацией о распределении донорно-акцепторных центров в исходной молекуле и о направлении разрыва циклических связей.

Рассмотрим перегруппировку пиридинового ядра определяемую последовательностью реакций (11a) и (11б), схема 5, см. [18]. В обоих случаях происходит гетеролиз связи C—N и формирование нового 6-членного цикла. Нетрудно видеть, что это реакции одного и того же класса (с идентичным G_1 -графом 11a с кодом 665-aa), принадлежащие разным семействам. Между тем, несомненно важное и глубокое химическое различие между ними состоит в том, что в первой реакции (11a) электрофильный атом углерода цикла заменяется электрофильным экзо-атомом углерода циано-группы, тогда как во второй реакции (11б) нуклеофильный атом азота цикла меняется местами с нуклеофильным экзо-атомом азота. Иными словами, число электрофильных и нуклеофильных центров в реакциях (11a) и (11б) не совпадает.

Такой результат легко отразить в одной и той же структуре рециклизационного графа G_1 11a: пусть электрофильные атомы (E -центры) изображаются в вершинах этого графа светлым кружком, а нуклеофильные



(*N*-центры) — черным, тогда реакции (11a) и (116) представимы различными окрашенными диаграммами 11б и 11в (см. схему 5). Полученные G_1 -графы с раскрашенными вершинами будем далее называть $G_1(EN)$ -графами. В таких $G_1(EN)$ -графы учитывается уже не только структурный аспект перераспределения связей, но и химический фактор — донорно-акцепторная природа атомов, между которыми рвутся или вновь образуются связи. Будем говорить, что ПТЦ относится к одному и тому же сорту, если соответствующие им графы $G_1(EN)$ совпадают; таким образом, реакции (11a) и (116) относятся к различным сортам с противоположным расположением *E,N*-центров.

Нетрудно видеть, что с точки зрения комбинаторики графы $G_1(EN)$ получаются из исходного G_1 -графа путем неэквивалентной раскраски последнего, т. е. расстановки меток (черных и белых кружков) на концах пунктирных ребер графа G_1 . Если пунктирные ребра в G_1 -графе несмежны, их концы необходимо раскрашивать в противоположные цвета. Проблема перебора неэквивалентных способов такой раскраски в общем виде решена нами ранее [34]; заметим, что поскольку число $G_1(EN)$ -графов меняется (в зависимости от группы симметрии исходного графа G_1) от одного до четырех, нетрудно перечислять такие семейства вручную.

Поскольку графы $G_1(EN)$ (как и G_2 -графы), являются вершинно помеченными G_1 -графами, их положение в рассмотренной выше иерархии рециклизаций несколько особое: их можно рассматривать как независимый подуровень рециклизационных классов. В свою очередь с уровня $G_1(EN)$ -графов можно перейти к несколько иным (чем были определены ранее) M -графам, которые мы будем называть $M(EN)$ -графами, содержащим лишь черные и белые метки вершин; соответствующие пары $M(EN)$ графов реакций (11a), (116) представлены на схеме 5. Такие $M(EN)$ -графы с заранее заданным распределением *E,N*-центров содержат важную эвристическую информацию, подсказывая, какие именно химические функции должны присутствовать в структурах гетероциклов любых семейств, чтобы отвечающие им ПТЦ принадлежали одному сорту. Например, появление метки черного цвета в графе $G_1(EN)$ требует наличия в $M(EN)$ -графах (и соответствующих гетероциклах, независимо от распределения гетероатомов) функции со строго фиксированным свойством электронодонорности (нуклеофильности), скажем, amino- или *CN*-кислотной группы, тогда как

«белая» метка вершины требует соответственно акцепторной (электрофильной) группировки типа карбонильной, нитрильной или нитрозо-группы. Нетрудно убедиться что трансформации пириллиевых солей под действием гидроксилamina (7) и гидразинолиз пириимидинов (9) — реакции одного сорта.

ПРОБЛЕМА ПРОГНОЗА ПРИНЦИПИАЛЬНО НОВЫХ ПРИМЕРОВ РЕЦИКЛИЗАЦИЙ

Рассмотренная классификация естественным образом приводит к проблеме «пустых клеток» — неизвестных примеров рециклизаций на разных ветвях иерархического дерева. Прогнозирование беспрецедентных примеров перегруппировок различных типов, видов, классов и т.д. требует, очевидно, надежной информации о рециклизационных графах уже существующих ПТЦ и об их распределении по уровням иерархии.

Для решения такого рода проблем нами разрабатывается на базе изложенного подхода компьютерная программа GREH (Graphs of REcyclizations of Heterocycles) [44], представляющая, с одной стороны уникальную, постоянно пополняемую базу данных и учебную базу знаний по рециклизациям гетероциклов, а с другой — инструмент прогнозирования принципиально новых гетероциклических трансформаций. Разумеется, использование этой программы для целей прогноза предполагает использование на последнем этапе как экспертной оценки со стороны химиков, так и расчетных методов из арсенала современной квантовой химии. Заметим, что последовательная классификация известных литературных примеров ПТЦ и статистика встречаемости различных типов РГ также является независимым и полезным критерием экспертной оценки прогнозов.

Методология прогнозирования неизвестных рециклизаций включает в себя общий обзор литературы, построение для рециклизаций соответствующих им РГ, поиск различных типов РГ для которых не имеется прототипа в виде реакции, построение пары М-графов и экспертную оценку исходного соединения, которое могло бы подвергаться неизвестному ранее структурному превращению.

Проведенный нами ранее обзор многообразия типов перегруппировок азолов [34] был первым примером подобного прогнозирования. Как оказалось, наиболее распространенными являются те из перегруппировок, G_1 -граф которых представим кодом типа $55N-aa$ (т. е. разрываемая связь исходного цикла смежна с образующейся связью формируемого цикла). В частности, хорошо известные классы перегруппировок азолов типа Димрота, Корнфорта, Боултона-Катрицкого имеют соответствующие коды $554-aa$, $553-aa$ и $552-aa$, см. реакции I, III, V, а также трансформация VII, с кодом $551-aa$ на схеме 6. Перегруппировки других классов (с кодами, отличными от $55N-aa$) хотя и встречаются в литературе, но довольно редко, что оставляет экспериментаторам широкое поле деятельности по обнаружению действительно новых классов азол-азольных трансформаций.

Для каждого из классов перегруппировок с кодом G_1 -графа $55N-aa$ теоретически возможны лишь два рециклизационных сорта (два $G_1(EN)$ -графа) с противоположным распределением донорных и акцепторных центров. Примеры пар перегруппировок, различных по сорту, представляют собой пары реакции I и II, V и VI, тогда как реакция Корнфорта III и ПТЦ VII [44] подобных антиподов, не имеют. Таким образом приведенные пары М-графов превращений IV и VIII представляют собой перспективные прототипы для поиска азольных перегруппировок ранее неизвестных сортов.

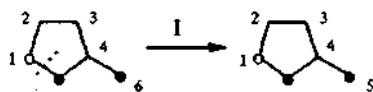
Попытаемся провести хотя бы качественное рассмотрение проблемы более глобального характера — распределение по уровням иерархии взаимных *перегруппировок* 5- и 6-членных гетероароматических систем,

(5 → 5) — Трансформация гетаренов

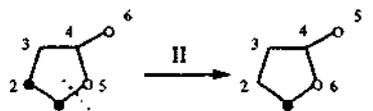
G₁ (EN)-граф
(сорт реакции)

M (EN)-графы
и нумерация вершин

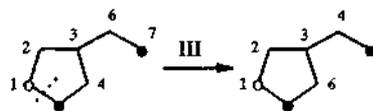
Пример расположения
гетарилов в схеме



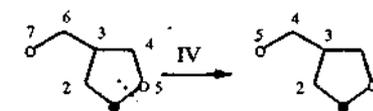
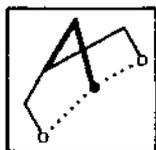
S(5)N(6)



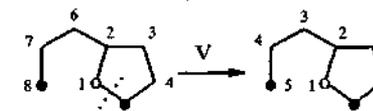
S(1)



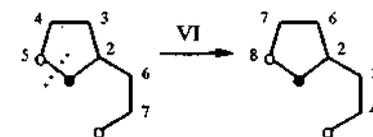
N(2)O(5)O(7)



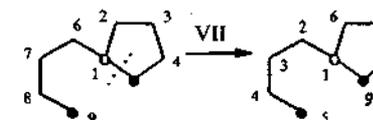
—



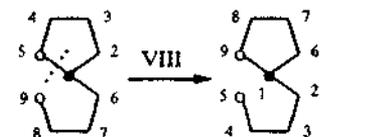
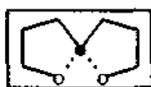
O(5)N(7)N(8)



S(1)N(5)

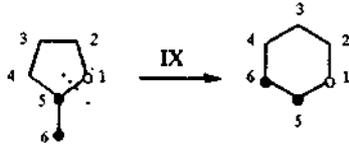


O(5)N(8)N(9)

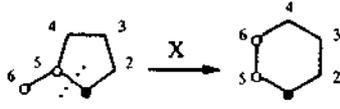


—

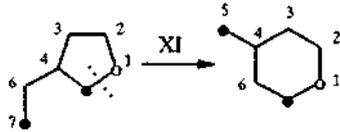
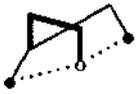
(5 → 6) — Трансформация гетаренов



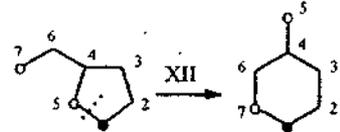
N(5)N(6)



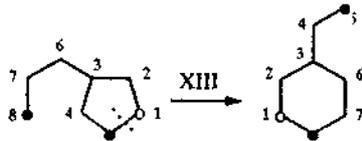
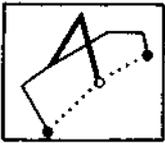
N(1)



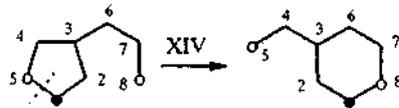
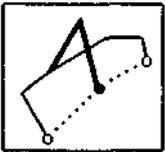
O(5)N(7)



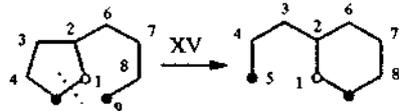
N(3)N(5)



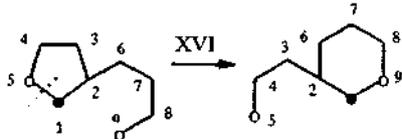
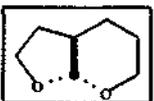
—



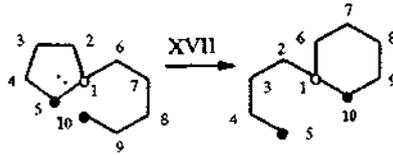
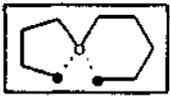
—



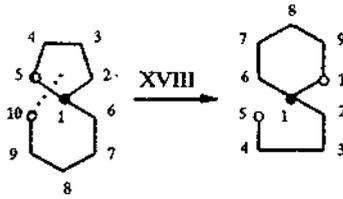
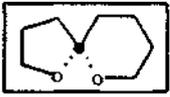
O(5)O(9)



—

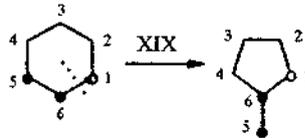


—

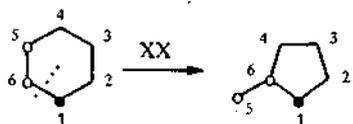


—

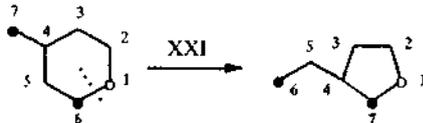
(6 → 5) — Трансформация гетаренов



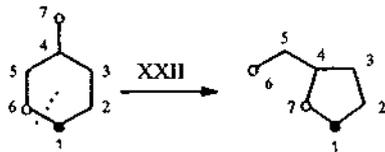
N(5)



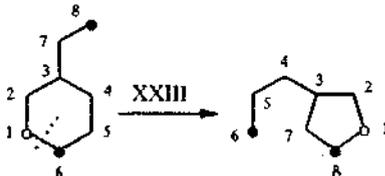
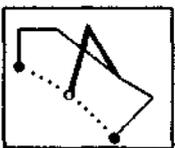
O(1)



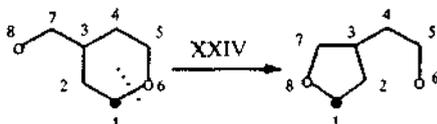
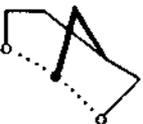
N(5)N(6)N(7)S(3)



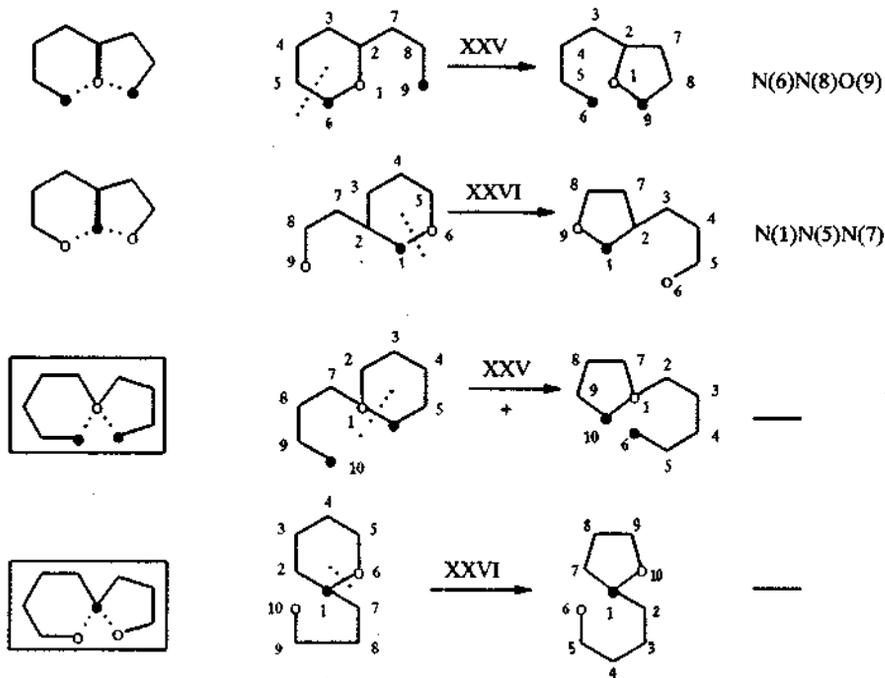
N(1)



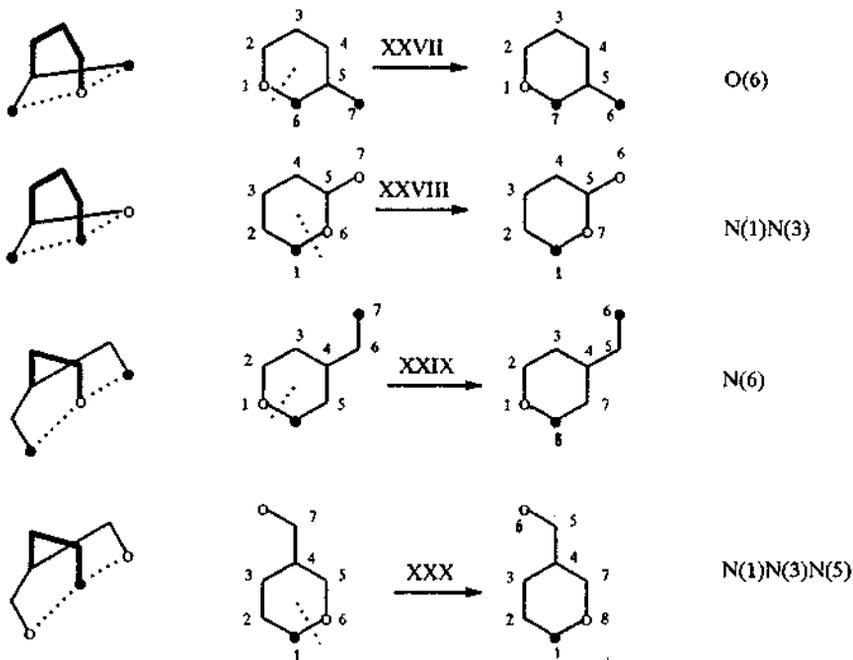
—

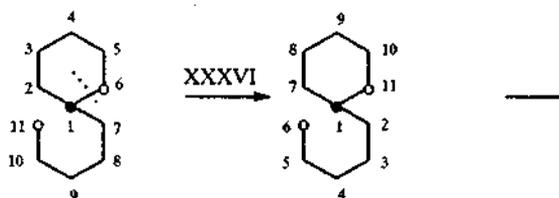
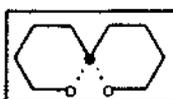
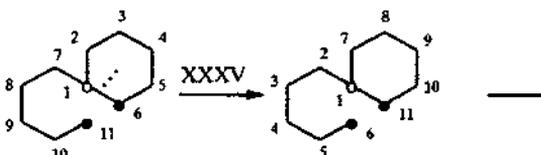
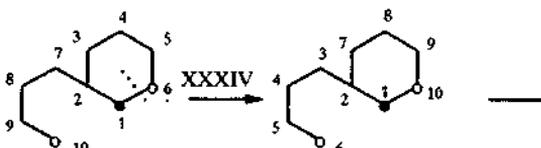
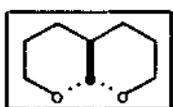
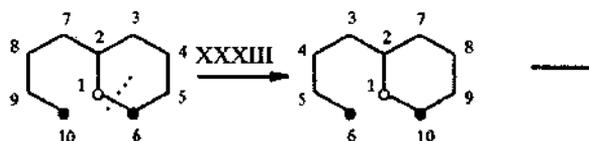
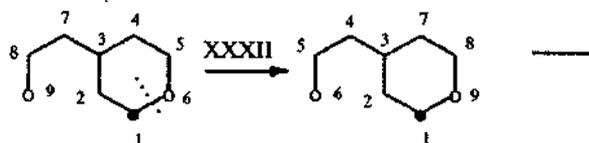
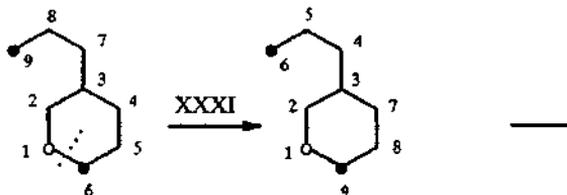


N(1)N(7)N(8)



(6 → 6) — Трансформация гетаренов





архетипы взаимопреращений которых задаются (K,L) -парами вида $(5,6)$, $(6,5)$ и $(6,6)$. Заметим, что как и прежде, мы ограничиваемся во-первых, простыми трансформациями циклов и во-вторых, *полностью отбрасываем* случаи включения и/или элиминирования каких-либо фрагментов исходного либо конечного циклов.

Анализ доступных литературных данных по таким перегруппировкам показывает, что (как и в случае азол-азольных ПТЦ) наибольшее распространение имеют те трансформации циклов, коды G_1 -графов которых представимы в виде $KLN-aa$. Структуры всех теоретически возможных G_1 -графов с таким кодом для взаимных ПТЦ 5- и 6-членных циклов представлены на схеме 6. Заметим, что код aa для G_1 -графов уникален: только для этого кода общее число реакционных центров (донорных и акцепторных), между которыми перераспределяются связи равно трем. (При любом другом распределении букв в коде G_1 -графов число таких центров, с

необходимостью должно возрасти до четырех.) Таким образом работает своеобразное магическое правило «минимального числа реакционных центров», нарушение которого может, по-видимому, приводить к конкурентным процессам, например, перекрестным побочным конденсациям. Таким образом, новым стратегическим направлением в области рециклизаций как азолов, так и азинов мог бы явиться целенаправленный поиск классов ПТЦ с кодами G_1 -графов, отличными от $KLN-aa$.

В качестве примера прогнозирования новых сортов взаимных трансформаций пяти- и шестичленных циклов для всех G_1 -графов aa -типа на схеме 6 приведены пары возможных $G_1(EN)$ -графов отвечающих превращениям I—XXXVI. В случае, если соответствующая реакция обнаружена в литературе, указан ее частный пример (не отражающий статистику встречаемости данного сорта). В случае же, если соответствующий сорт перегруппировки нам пока неизвестен («пустая клетка»), дана структурная схема реакции в виде пары $M(EN)$ -графов, а $G_1(EN)$ заключен в рамку.

Схема 6 носит предварительный характер и публикуя ее, мы ставим несколько целей:

- привлечь внимание химиков-экспериментаторов к формальным моделям математической химии;
- расширить существующую точку зрения на процессы рециклизации в целом и обсудить достоинства и недостатки идеи структурной иерархии и кодирования рециклизаций;
- начать разработку уникальной компьютерной базы данных по уже существующим рециклизациям;
- стимулировать экспериментальные исследования в области поиска действительно беспрецедентных примеров трансформаций гетероциклов.

По мере пополнения компьютерной базы данных, мы надеемся улучшать результаты схемы 6, а также публиковать атласы аналогичных схем для других типов, видов, классов, сортов и семейств трансформаций гетероциклов. Любая критика, замечания, дополнения и контрпримеры к схеме 6 будут приняты с благодарностью, в особенности если они будут носить характер экспериментальных опровержений.

Намереваясь в дальнейшем попытаться ограничить количество пустых клеток схемы 6 (и будущих аналогичных схем) в рамках собственных поисковых экспериментальных исследований, авторы открыты к сотрудничеству с любыми другими коллективами.

СПИСОК ЛИТЕРАТУРА

1. *Plas Van der H. C. Ring Transformation of Heterocycles.* — L., NY: Academic Press, 1973. — Vol. 1, 2.
2. *Швайка О. П., Артемов В. Н. // Усп. химии.* — 1972. — Т. 41. — С. 1788.
3. *Швайка О. П. Реакции рециклизации гетероциклических соединений под действием гидразина и некоторые исследования в ряду азолов. Автореферат дисс.... докт. хим. наук.* — Киев, 1976.
4. *Pyrrilium Salts. Syntheses, Reactions and Physical properties.* / Ed. A. T. Balaban e. a. — N. Y.: Academic Press, 1982.
5. *Балабан А. Т. // Доклады VI Междунар. Конгресса IUPAC по Органическому Синтезу.* — 1986. — Москва. Пленарный доклад.
6. *Schroth W. // Chem. of Heteroc. Compounds.* / Eds. J. Kovac, P. Zalupsky — Amsterdam e. a.: Elsevier, 1988. — P. 115.
7. *L'abbe G. // J. Heterocycl. Chem.* — 1984. — Vol. 21. — P. 627.
8. *Boulton A. J. // Lectures in Heterocyclic Chemistry.* / Eds. R. N. Castle, L. B. Townsend. — Orem, Utah: Heterocorporation, 1974. — Vol. 2. — P. S-45.
9. *Плас ван дер Х. // ХГС.* — 1978. — № 7. — С. 867.
10. *Страдынь Я. П. // ХГС.* — 1979. — № 11. — С. 1570.
11. *Becher J. // Synthesis.* — 1980. N 8. — P. 589.
12. *Padwa A. // Rearrangements in Ground and Exited States.* / Ed. P. de Mayo. — N. Y.: Academic Press, 1980. — Vol. 3. — P. 501.

13. *Ramsden C. A.* // *Adv. Heterocycl. Chem.* — 1980. — Vol. 26. — P. 1.
14. *Ruccia M., Vivona N., Spinelli D.* // *Adv. Heterocycl. Chem.* — 1981. — Vol. 29. — P. 141.
15. *Kost A. N., Gromov S. P., Sagitullin R. S.* // *Tetrahedron.* — 1981. — Vol. 37. — P. 3423.
16. *Робев С. К.* // *ХГС.* — 1981. — № 000. — С. 1587.
17. *Чарушин В. Н., Чупахин О. Н.* // *Усп. Химии.* — 1984. — Т. 53. — С. 1648.
18. *Сажитуллин Р. С.* // *ХГС.* — 1984. — № 4. — С. 563.
19. *Boger D. L.* // *Chem. Rev.* — 1986. — Vol. 86. — P. 781.
20. *Kato A., Nishio T., Kashina C.* // *Heterocycles.* — 1987. — Vol. 26. — P. 2223.
21. *Comprehensive heterocyclic chemistry.* / Eds. Katritzky A. R., Rees C. W. — Oxford: Pergamon Press, 1984. — Vol. 1-8.
22. *Katritzky A. R., Dennis N.* // *Chem. Rev.* — 1989. — Vol. 89. — P. 827.
23. *Алексеева Н. В., Яхонтов Л. Н.* // *Усп. Химии.* — 1990. — Т. 59. — С. 888.
24. *Андрианов В. Г., Еремеев А. В.* // *ХГС.* — 1990. — № 11. — С. 1443.
25. *Зефиоров Н. С., Трач С. С.* // *ЖОрХ.* — 1975. — Т. 11. — С. 225.
26. *Zefirov N. S., Tratch S. S.* // *Chem. Scripta.* — 1980. — N 15. — P. 4.
27. *Zefirov N. S.* // *Acc. Chem. Res.* — 1987. — Vol. 20. — P. 237.
28. *Fujita S.* // *J. Chem. Inf. Comput. Sci.* — 1986. — Vol. 26. — P. 205.
29. *Баскин И. И., Трач С. С., Зефиоров Н. С.* // *Тез. докл. 7 Всес. конф. "Исп. ЭВМ в хим. исслед. и спектроск. молекул"*. — Рига, 1986. — С. 88.
30. *Vladutz G.* / *Modern Approaches to Chemical Reaction Searching.* / Ed. Willet P. — Gover: Andershot, 1987. — P. 202.
31. *Herges R.* // *Tetrahedron Computer Methodology.* — 1988. — Vol. 1. — N 1. — P. 15.
32. *Hopkinson G. A., Cook T. P., Buchan I. P.* // *Chemical Information Systems. Beyond the Structure Diagram.* / Ed. D. Bawden, E. Mitchell — N. Y.: Ellis Horwood, 1990. — P. 83.
33. *Бабаев Е. В., Зефиоров Н. С.* / *Молекулярные графы в химических исследованиях (Тезисы докл. межвуз. конференции)*. — 1990. — Калинин. — С. 3.
34. *Babaev E. V., Zefirov N. S.* // *Bull. Soc. Chim. Belg.* — 1992. — Vol. 101. — P. 67.
35. *Салахов М. С., Исмаилов С. А.* // *Усп. Химии.* — 1986. — Т. 55. — С. 2008.
36. *Теренин В. И., Бабаев Е. В., Юровская М. А., Бундель Ю. Г.* // *ХГС.* — 1992. — № 6 — В печати (этот номер).
37. *Бабаев Е. В., Вобровский С. И., Бундель Ю. Г.* // *ХГС.* — 1988. — № 11. — С. 1570.
38. *Зефиоров Н. С., Трач С. С., Чижов О. С.* Каркасные и полициклические соединения. Молекулярный дизайн на основе принципа изоморфного замещения. (Итоги науки и техники.). — М., 1979. — 87 с.
39. *Папулов Ю. Г.* // *Симметрия молекул.* — Калинин, 1979. — 84 с.
40. *Klin M. Ch., Pochel R., Rosenbaum K.* *Angewandte Algebra.* — Bin.: VEB Deutsche Verlag der Wiss., 1988.
41. *Babaev E. V., Zefirov N. S.* / *IV-th Intern. Conf. on Mathem. and Comput. Chem.* 3—7.06.1991, Bled, Yugoslavia. (*J. Mathematical Chem.*, 1992, в печати.)
42. *Харари Ф.* // *Теория графов.* — М.: Мир, 1973. — 300 с.
43. *Бабаев Е. В., Лушиников Д. Е., Зефиоров Н. С.* / *Химическая информатика. (Тез. докл. IX Всес. конфер.)*. — Черногоровка, 1992. — С. 93.
44. *Holla B. S., Ambekar S. I.* // *J. Chem. Soc. Chem. Comm.* — 1979. — N 5. — P. 221.

Московский государственный университет
им. М. В. Ломоносова
Москва 119899

Поступило в редакцию 30.03.1992