

УДК 547.12:541.14(083.73)

## ЛИНЕЙНО-ЦЕПНОЕ КОДИРОВАНИЕ ФОРМУЛ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ. V. КОДИРОВАНИЕ НЕУГЛЕРОДНЫХ АТОМОВ

© Я.Э. Брюске

Y.E.Brüske. *The line-chain notation of organic-compound formulae. V. The notation of non-carbon atoms.* Non-carbon atoms (heteroatoms) are labeled by suitable symbols from the periodic system of elements for the line-chain notation of hydrocarbon derivatives. Chain bonds of a carbon atom, denoted by a number, with a heteroatom and heteroatoms with each other are denoted by simply writing their symbols together. The heteroatom symbol should be next to the number, and the two heteroatoms are written in the order determined by their bond in the encoded molecule. Both letters in the two-letter atom symbol are capital, therefore in order to discern them, periods are used. The atom, which symbol is written in linear sequence after the comma, is considered to be combined by the non-chain bond with the first (left) atom of the nearest chain bond on the left. Another method of non-chain bond notation consists in including the group of heteroatoms being in the lateral chain in brackets.

В [1 - 4] рассмотрено линейно-цепное кодирование углеводородов. Поскольку все атомы углерода в молекуле нумеруют, код состоит исключительно из чисел и разделяющих их знаков «препинания». Нумерация атомов по неразветвленным углеродным цепям позволяет не давать в коде информации о связях между атомами углерода с последовательными номерами (цепные связи). Информацию дают о нецепных связях и об отсутствии некоторых цепных связей. Нецепную связь обозначают номерами ее атомов, разделяя их запятой, при этом больший номер помещают впереди. При отсутствии цепной связи пишут ноль перед большим из ее возможных последовательных номеров. Сообщения о различных связях в одной молекуле разделяют тире или дефисами.

Атомы водорода в коде не обозначают: ввиду постоянной валентности углерода их число у каждого углеродного атома определяется путем вычитания из четырех числа обозначенных в коде связей этого атома с другими атомами углерода.

Сложные углеводороды нумеруют и кодируют по фрагментам, каждый из которых нумеруют сначала (автономно) и заключают в круглые скобки [3].

Информация об атомах химических элементов настолько хорошо «закодирована» в периодической системе элементов Д.И. Менделеева, что в линейно-цепном коде органической молекулы, содержащей, наряду с атомами углерода, атомы других элементов, ее можно использовать без всяких принципиальных изменений в виде символов этих элементов. Обычно неуглеродных атомов (гетероатомы) в органической молекуле меньше, чем атомов углерода и водорода, а их разнообразие позволяет в большинстве случаев обходиться без нумерации. В отличие от символов самой периодической системы, обе буквы в двухбуквенном символе элемента прописные, чтобы не увеличивать разнообразия знаков в коде.

Из 26 букв латинского алфавита в однобуквенных символах элементов «занято» 15 (без символа углерода, но с символами изотопов водорода). В

других значениях в линейно-цепном коде их использовать нельзя. Также в другом значении нельзя использовать ни одну пару прописных букв, написанных рядом, без пробелов и без разделительных знаков, если она соответствует двухбуквенному символу химического элемента. Остальные 11 «одиноких» букв можно применять в коде для его сокращения в качестве специальных символов. Таким образом, в [4] использованы буквы R и E для обозначения ароматической и сопряженной циклической систем связей в углеводородах.

**Цепные связи неуглеродных атомов.** Так же, как и между углеродными атомами, различают цепные и нецепные связи неуглеродных атомов. Цепную связь обозначают простым написанием символа элемента после (справа от) номера, если это атом углерода или же рядом с символом другого связанного с ним неуглеродного атома. Например,

CH <sub>3</sub> NH <sub>2</sub>	1N
CH <sub>3</sub> NHON	1NO
CH <sub>3</sub> ONH <sub>2</sub>	1ON
n-C <sub>3</sub> H <sub>7</sub> OH	1O-3

(здесь нельзя написать более экономный код 3O, т.к. это противоречит одному из основных принципов номенклатуры: заместитель должен быть у углерода с меньшим номером).

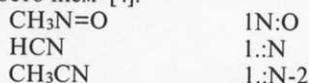
Номера считаются старше гетероатомов и поэтому для кодирования связи углерода с неуглеродным атомом нахождение символа последнего справа от (после) номера обязательно [1], а порядок написания символов двух неуглеродных атомов зависит от порядка связей между ними в самой кодируемой молекуле, что видно из сравнения кодов формул CH<sub>3</sub>NHON и CH<sub>3</sub>ONH<sub>2</sub>.

Если написанный несколько раз один и тот же номер обозначает (в пределах одного и того же фрагмента) один и тот же атом углерода, то несколько раз повторенный символ элемента обозначает, сколько атомов этого вида в молекуле, сколько раз он повторен. Например,

CH <sub>3</sub> NHNNH <sub>2</sub>	1NNN
------------------------------------	------

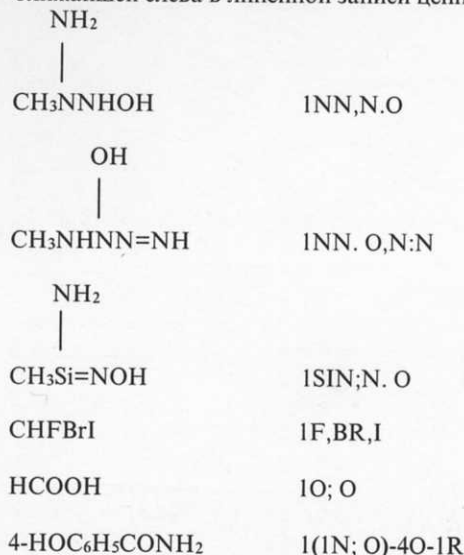
Углерод и все атомы азота находятся в одной неразветвленной цепи, поэтому и связи, так обозначенные, названы цепными.

Кратные цепные связи также, как и между атомами углерода, обозначают двоеточием и точкой с двоеточием [4]:



Ввиду того, что все буквы в символах элементов большие, возникает проблема различения двухбуквенного символа, состоящего из двух однобуквенных. Если нет специальных указаний, символы определяют в порядке следования букв. Например, SINB – кремний, ниобий. Если же нужно в этой же последовательности выделить другие элементы, ставят необходимое число точек: S.INB – сера, индий, бор, SIN.B – кремний, азот, бор, S.I.NB – сера, иод, ниобий, S.I.N.B. – сера, иод, азот, бор. Точка, наряду с обычным написанием символов рядом, может обозначать простую цепную связь, но ставят ее только в случаях, подобных этим. Среди рассмотренных выше соединений в коде CH<sub>3</sub>NHOH между символами азота и кислорода необходимо поставить точку, т.к. иначе эту пару букв следует принять за символ нобелия.

**Нецепные связи неуглеродных атомов.** Нецепной связью неуглеродного атома с углеродом (номером) или с другим неуглеродным атомом считается такая связь между ними, когда их символы в линейной записи нельзя поставить рядом, или же когда надо обозначить связь с боковой цепью, присоединенной к одному из атомов основной цепи. В первом случае символ атома отделяют от предшествующего символа запятой (точкой с запятой для двойной связи) и считают, что этот атом соединен с левым атомом ближайшей слева в линейной записи цепной связи:



По второму варианту группу, выделяемую в боковую цепь, помещают справа от атома (или номера), к которому она присоединена, и заключают в скобки. Рассмотреть этот способ целесообразно на примерах кремнийорганических соединений, в которых число непосредственно соединенных атомов кремния может достигать до девяти [5].



Здесь атом кремния, стоящий после левой скобки, присоединен к атому кремния, находящемуся перед этой скобкой, нецепной связью, а атом кремния, стоящий после правой скобки, присоединен к атому кремния, находящемуся перед левой скобкой, цепной связью. Выводить таким способом в боковую цепь один атом нецелесообразно, т.к. в этом случае можно составить код с меньшим числом знаков или применив кодирование при помощи запятой, как описано выше, или же поместив внутри скобок еще один атом.

В отличие от углеводородного фрагмента, внутри этих скобок других скобок быть не должно, и поэтому такая группа атомов не имеет входящей связи, а выходящая связь идет только от первого (крайнего левого) атома внутри скобок. Такая пара скобок называется бесфрагментной.

Кратную выходящую связь помещают внутри скобок перед первым атомом:



Кратную связь имеющего боковую цепь атома основной цепи обозначают после правой скобки знаком цепной связи:



Здесь следует отметить, что к производным углеводородов относят соединения, в молекулах которых, наряду с углеродом и водородом (органических соединений, не содержащих водорода, относительно мало) имеются так называемые атомы-органогены [5]: O, N, S, P, галогены. Однако фосфорорганические соединения, вместе с органическими соединениями других не перечисленных выше элементов, относят к элементарноорганическим соединениям, очевидно, потому, что «истинные» фосфорорганические соединения, в которых фосфор непосредственно соединен с углеродом, почти не встречаются в природе [5].

**Кодирование гетероциклов.** Гетероциклические соединения целесообразнее всего рассматривать и называть как производные соответствующих циклических углеводородов. По заместительной, «а»-номенклатуре ИЮПАК [6] их так и называют. По этому же принципу их кодируют, используя не применяющийся в кодах углеводородов символ углерода для обозначения номера гетероатома (С-замена). Для такой замены применяют следующие правила:

- номер гетероатома помещают между символом углерода и символом гетероатома (рис. 1);
- номера нескольких одинаковых гетероатомов помещают между символами углерода и этого гете-

роатома в возрастающем порядке и разделяют запятыми (рис. 2);

- если в гетероцикле имеются разные гетероатомы, их кодируют, располагая в порядке увеличения отдельных номеров множеств каждого гетероатома (см. рис. 4 и пояснение к нему). При этом, в отличие от предшествующих правил, символ углерода отделяют от следующего за ним номера запятой, а номер следующего гетероатома от символа предшествующего запятой не отделяют (рис. 3);

- атомы гетероцикла нумеруют по правилам соответствующего циклического углеводорода, и только если его симметрия позволяет, выбирают то множество номеров всех гетероатомов, в котором раньше встретится меньший номер (правило наименьшего локанта [7]).

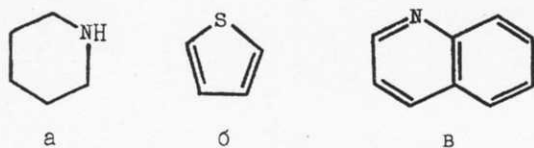


Рис. 1. а) C1N-6,1 б) C1S-1E5 в) C1N-1,5R10

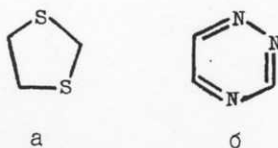


Рис. 2. а) C1,3S-5,1 б) C1,2,4N-1R

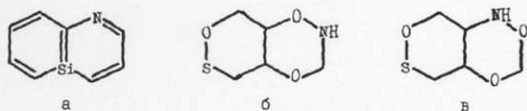


Рис. 3. а) C,1N5SI-1,5R10 б) C,1,4,8O2N7S-10,1,5 в) C,1N2,4,8O7S-10,1,5

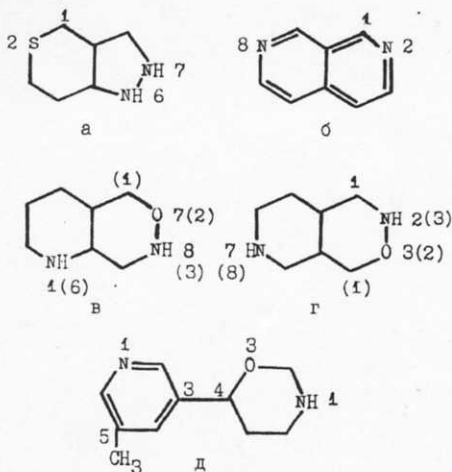
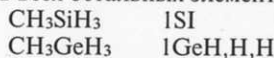


Рис. 4. а) C,2S6,7N-9,5 б) C2,8N-1,5R10 в) C,1,8N7O, но не C,2O3,6N-10,1,5, т.к. на первом месте 1 меньше 2 (сравнивают множества номеров 1, 7, 8 и 2, 3, 6); г) C,2,7N3O-10,1,5, но не C,2O3,8N-10,1,5 (2, 3, 7, лучше, чем 2, 3, 8) д) 4(,3-5(1)-C1N-1R)-C,1N3O-6,1

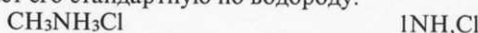
Если и после этого остается выбор, меньшие номера дают старшим гетероатомам. (Старшинство гетероатомов, наряду с вопросами старшинства фрагментов [3], будет рассмотрено в отдельном сообщении). При дальнейшей возможности выбора меньшие номера дают выходящей, затем старшим входящим внешним [3] связям (рис. 4);

- сообщение о гетероатомах так же, как и о внутренних связях, описывает внутреннюю структуру гетероцикла, поэтому его помещают в коде перед сообщениями о внутренних связях (отдельные начальные атомы и нецепные связи).

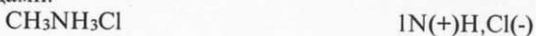
**Водород, присоединенный к неуглеродному атому.** Ввиду большого разнообразия водородных соединений неуглеродных атомов, проблему обозначения присоединенных к гетероатому атомов водорода нельзя решить так же просто, как это сделано для углеводородов. Все элементы "главных" подгрупп (*s*- и *p*-элементы) периодической системы Д.И. Менделеева имеют, аналогично углероду, постоянную валентность по водороду, определяемую номером группы [8], однако многие из них, с большой атомной массой, имеют переменную валентность. *d*-элементы проявляют большее разнообразие как в валентностях по водороду, так и в изменениях самих величин валентностей. Только у атомов второго и третьего периодов строго соблюдается постоянство их валентностей по водороду, определяемое номером его группы. Поэтому вводится правило: не обозначают водороды у атомов элементов второго и третьего периода и обозначают их у атомов всех остальных элементов, например,



Из этого правила есть исключение: водород у элемента второго или третьего периода следует обозначать, если общая валентность элемента превышает его стандартную по водороду:



Стандартная валентность азота по водороду равна трем, а здесь она равна пяти. Обозначают только «лишнее» число атомов водорода (один из трех). Электровалентную связь от ковалентной можно не отличать, но при необходимости можно обозначить соответствующими электрическими зарядами:



Другой пример:



Стандартная валентность фосфора по водороду равна трем, а здесь она равна пяти. Поэтому единственный атом водорода у фосфора тоже «лишний» и должен быть обозначен.

Во всех рассмотренных выше примерах неуглеродные атомы принадлежат ко 2 и 3 периодам и водороды в кодах не обозначены.

**Обозначение водорода в коенах.** В циклических неароматических углеводородах с максимальным числом сопряженных связей (коены [4]) и нечетным числом атомов углерода в цикле наименьший номер дают атому, имеющему два атома водорода

(обозначенный водород [6]). Однако, как показано на рис. 5б, в, это не всегда возможно: на рис. 5б дополнительный атом водорода стоит у 9-го атома пятичленного коенового цикла, а его наименьший номер восьмой, а на рис. 5в вторые водородные атомы стоят не у наименьших номеров в обоих коеновых циклах. В коде рис. 5в обозначения обоих атомов водорода объединены в одно сообщение так же, как в [4] объединены в одно сообщение коды нескольких двойных связей. На рис. 5г приведен пример подобного объединения кодов трех гидроксильных групп.

**Хиноны и шестичленные коены.** Уже в молекуле нафталина взаимное влияние сросшихся циклов нарушает ароматическую выравненность сопряженных двойных связей.

Формально это можно отразить, написав все формулы, отличающиеся друг от друга расположением двойных связей (концепция мезомерии - «резонанса» - слово взято в кавычки, т.к. резонанса, как физического явления, здесь нет, а различные классические формулы, изображающие одну и ту же молекулу, называются каноническими структурами (формулами). То, что эти формулы описывают одно и то же вещество, отображено двойными стрелками между ними. [5]). Данные рис. 6 показывают, что при фиксированном расположении двойных связей в одном кольце, в другом их уже нельзя расположить произвольно. О нарушении выравненности связей можно судить по тому, что двойная связь между атомами 1 и 2 и парами атомов, им симметричных, встречается в формулах рис. 6а, б, в вдвое чаще, чем между всеми другими парами атомов. У формул б и в связи в циклах расположены несимметрично, но сами они симметричны друг другу, и одну из них можно не писать, удваивая (утраивая, если таких структур три и т.д.) влияние такой структуры на свойства вещества.

Для антрацена можно написать две канонические формулы (рис. 7), у каждой из которых имеется по одной симметричной, так что канонических структур у антрацена четыре.

У фенантрена имеются четыре несимметричные друг другу канонические структуры, но только формула рис. 8в не имеет внутренней симметрии и поэтому существует симметричная ей пятая каноническая структура.

На всех трех рисунках буквой о помечены орто-хиноидные структуры одного-двух бензольных циклов (в формуле рис. 8г в среднем кольце - двойная орто-пара-хиноидная - оп - структура). Поскольку у этих молекул в любом кольце бензольная система из трех двойных связей имеется не менее чем в двух канонических структурах, все они в значительной степени сохраняют ароматические свойства и кодируются [4], как соединения с бензольными циклами.

Этот анализ показывает, что ароматические свойства бензольного ядра сохраняются, если одна его внутренняя двойная связь заменена двумя внешними. На рис. 9 приведены оба бензохинона и их линейно-цепные коды. Тенденция к образованию бирадикала (рис. 9в) находит некоторое подтверждение в образовании семихинона, с одним неспаренным электроном, при восстановлении хинона в гидрохинон [5].

По официальной номенклатуре (ИЮПАК [6]) «ароматический» корень в названии хинонов сохра-

няется. Особенно наглядно это видно в номенклатуре Реферативного журнала Chemical Abstracts, где антрахинон называют как антрацендион, т.е. как дикетон с сохранением «ароматического» характера в корне слова-названия [7].

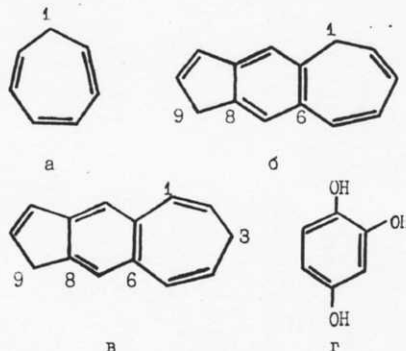


Рис. 5. а) 1E7 б) 9H-1E7-6R-8E5 в) 3,9H-1E7-6R-8E5 г) 1,2,4O-1R

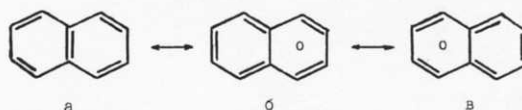


Рис. 6. Канонические формулы нафталина.

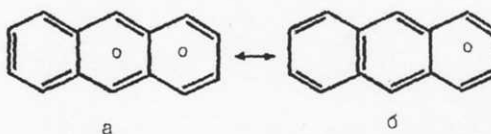


Рис. 7. Несимметричные канонические формулы антрацена.

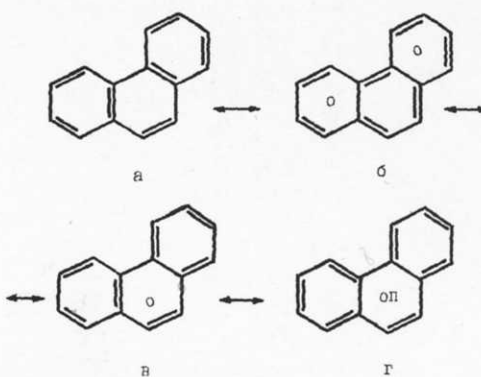


Рис. 8. Канонические формулы фенантрена.

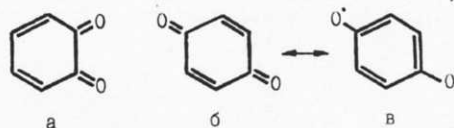


Рис. 9. а) 1,2:O-1R б) 1,4:O-1R в) бирадикальная каноническая структура п-хинона.

Если гексадиеновый цикл имеет только одну выходящую двойную связь, такое соединение не является ароматическим, но способно к легкому химическому превращению в него - изомеризации - с переходом атома водорода (в виде протона) из положения 1 циклов рис. 10а, в к кислороду. Так же, как и в коенах с нечетным числом атомов углерода в цикле, у атомов рис. 10а и в, помеченных единицей, имеются два атома водорода. Поэтому их целесообразно кодировать, как коены, присваивая наименьший номер углеродному атому с дополнительным атомом водорода [6]. Эти коды приведены в подписи к рис. 10.

Оба изомера циклогексадиена целесообразно кодировать, как диен: 1,3:-6,1 и 1,4:-6,1 соответственно. Если в этом случае, вместо углерода с дополнительным водородом, в цикле находится атом с неподеленной электронной парой (кислород), то такое соединение (пиран), существующее только

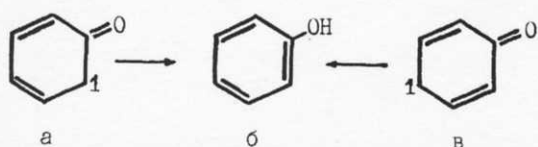


Рис. 10. Изомеризация циклогексадиенов в фенол:  
а) 2:O-1E6 б) 1O-1R в) 4: O-1E6

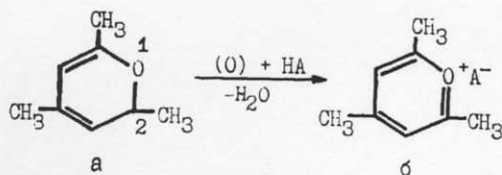


Рис. 11. а) 2,4,6(1)-2H-C1O-1E6 б) 2,4,6(1)-C1O(+)-A(-)-1R.

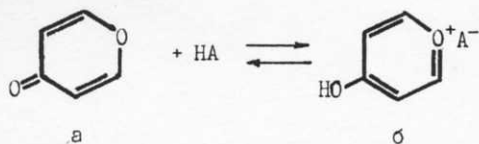


Рис. 12. а) 4: O-C1O-1R б) 4O-C1O(+)-A(-)-1R

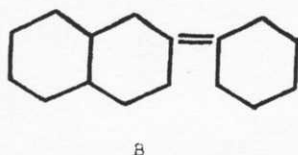
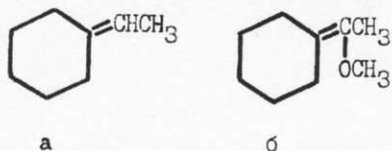


Рис. 13. а) 1(-)-2)-6,1 б) 1(1O(1)-2)-6,1 в) 2(6,1;)-10,1,5

вместе с присоединенными метильными группами [9], легко окисляется в присутствии сильной кислоты ( $\text{HClO}_4$ ) с образованием пирилеиновой соли, имеющей бензольную систему связей (рис. 11б). Таким образом пираны, аналогично соединениям с одной внешней двойной связью, целесообразно кодировать как гексакоены (рис. 11а). Сообщения о присоединении трех метильных групп объединены в одно так, как это сделано для двух присоединенных атомов водорода (рис. 5в), кислорода (рис. 9а, б) и трех гидроксильных групп (рис. 5г). Наличие у пиранового цикла дополнительной внешней двойной связи с кислородом приводит к пиранам, у которых тенденция к ароматичности еще больше увеличивается [5]: они являются основаниями, образующими с кислотами соли, которые имеют шестичленный ароматический цикл. Поэтому пираны целесообразно уже кодировать, как ароматические соединения (рис. 12а, б).

**Соединения со связывающими функциями.** Под функцией здесь будем понимать неуглеродный атом или их группу, не разделенную атомами углерода. Термин, приведенный в этом заголовке, автор в литературе не нашел, однако определенные так соединения, в которых по меньшей мере два углеводородных фрагмента соединены друг с другом не непосредственно, а через функцию, выделял еще основатель теории химического строения органических молекул А.М. Бутлеров [10].

Авторы [11] называют такие соединения функциональными производными, но включают в эту группу еще ряд других производных углеводородов, которые не являются соединениями со связывающими функциями, такие как амиды и галогенангидриды органических кислот, оксимы, нитрилы.

В кодах сложных углеводородов [3] номер атома углерода, имеющего входящую связь, непосредственно предшествует открывающей (левой) скобке того фрагмента, у которого эта же связь является выходящей. Для кодирования простой внешней связи запятую ставят внутри скобок перед номером выходящей связи, если он больше единицы. Вопреки предложенному в [3], отдельное сообщение о выходящей связи целесообразнее помещать на первое место, как это сделано в коде к рис. 4д.

Символ любого гетероатома считается старше любого номера внутри скобок, содержащих код присоединенного фрагмента. Поэтому гетероатом, стоящий непосредственно перед открывающей скобкой фрагмента, так же, как и номер, связан с этим фрагментом входящей связью, соединяя, таким образом, фрагменты с разобщенными углеродными скелетами.

Простейшим соединением со связывающей функцией является диметилловый эфир



линейно-цепной код которого

1O(1)

Здесь главным фрагментом может быть любая из метильных групп. При неодинаковом старшинстве фрагментов более старший из них должен быть главным и находиться вне скобок:

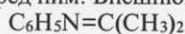


1O(1)-2

2O(2)-04,2

Код бутилизопропилового эфира  
 $n\text{-C}_4\text{H}_9\text{OCH}(\text{CH}_3)_2$  1O(,2-3)-4

т.е. номер выходящей связи, если он больше единицы, ставят на первое место внутри скобок с запятой перед ним. Внешнюю двойную связь в соединении



кодируют так же

1N(;2-3)-1R

Если двойная связь выходит от первого атома, то ее обозначают без единицы на первом месте (рис. 13а), но если единица в коде уже обозначена, то знак двойной связи ставят или перед (рис. 13б), или после нее (рис. 13в). Подобным образом у номеров, больших единицы, обозначают не только двойную, но и простую выходящую связь, если этот номер, аналогично кодам соединений рис. 13б, в, кодирует, наряду с информацией о выходящей связи, другую информацию.

Код соединения  
 $(\text{CH}_3)_2\text{CHN}=\text{CHCH}_3$  2N(;)-2-3

но код изомерного ему соединения  
 $(\text{CH}_3)_2\text{C}=\text{NCH}_2\text{CH}_3$  2:N(2)-3

Код кремнийорганического соединения



Здесь так же, как и во втором варианте кодирования нецепной связи гетероатомов при помощи скобок, код углеводородного фрагмента считается выделенным в боковую цепь, а символ гетероатома после закрывающей скобки обозначает продолжение цепи символа (или номера атома углерода), стоящего перед открывающей скобкой.

Если к одному и тому же углеродному или гетероатому присоединены несколько фрагментов, их помещают в своих скобках один за другим. Так кодируют углеводород, рис. 14. Оба углеводородных фрагмента присоединены к углероду №3, но не связаны друг с другом.

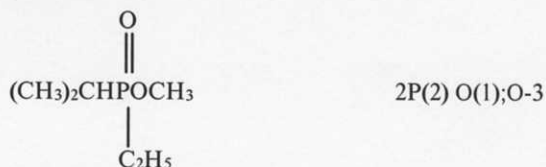
В кремнийорганическом соединении



углеводородные фрагменты также присоединены к атому кремния (его символ перед открывающей левой скобкой), не соединены друг с другом и не прерывают цепи гетероатомов: атом кремния, символ которого стоит после последней правой скобки, присоединен цепной связью к атому кремния, символ которого находится перед первой левой скобкой. Код третичного амина

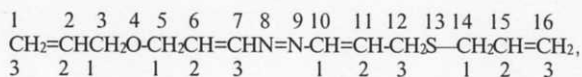


составляется аналогично. В фосфорорганическом соединении



первый фрагмент в скобках присоединен к фосфору, а второй - к кислороду, после символов которых каждый из них находится. Ни один из этих фрагментов не влияет на порядок связей атомов, символы которых стоят вне скобок: так, первый атом кислорода присоединен цепной связью к фосфору, а второй, символ которого стоит после точки с запятой, присоединен двойной нецепной связью к первому атому цепной связи PO, т.е. тоже к фосфору.

**Несколько связывающих функций в одной молекуле.** Среди подобных соединений наиболее часто встречаются такие, у которых связывающая функция состоит из одного гетероатома, а сама молекула имеет форму цепи из чередующихся углеродных и гетероатомов. Такие соединения в номенклатуре ИЮПАК предложено называть по «а»-номенклатуре, аналогично построению названий гетероциклических соединений. В этих соединениях нет циклов, поэтому основу их названий составляют названия соответствующих нециклических углеводородов с наиболее длинной цепью, включающей гетероатомы связывающих функций. Приведенное в [6] в качестве примера соединение



имеющее в неразветвленной цепи 16 атомов, названо 4-окса-13-тиа-8,9-диазагексадекапентаен-1,6,8,10,15. Код этого соединения, составленный по приведенным выше правилам, следующий (нумерация для него под формулой, главный фрагмент подчеркнут)

1O(2)-3N:N(1-3S(2))-2:

В этом коде информация о наличии трех атомов углерода в отрезках цепей дается как двойными связями у вторых номеров, так и третьими номерами, к которым присоединены гетероатомы.

В соответствии с названием по «а»-номенклатуре, линейно-цепной код этого же соединения составляют так же, как для гетероциклических молекул (С-замена, нумерация над формулой)

1, 6, 8, 10, 15:-C,4O8,9N13S-16

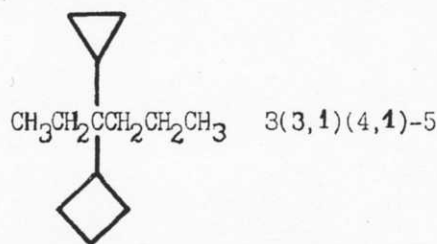


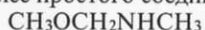
Рис. 14. Код 3-циклопропил-3-циклобутилпентана.

В номенклатуре ИЮПАК названия по «а»-номенклатуре не применяются, если соединение можно назвать по другой системе этой же номенклатуры. Для кодирования целесообразнее применить другой критерий: уменьшение общего числа знаков в коде при переходе к С-замене. Здесь в коде с С-заменой 27 знаков, больше, чем в предыдущем (25), поэтому С-замена нецелесообразна. Коды аналогичного соединения, не имеющего двойных связей,



имеют  
 10(3)-3NN(3S(3))                      16 знаков  
 C,4O8,9N1,3S-16                        14 знаков

т.е. С-замену в этом случае следует произвести. Для более простого соединения



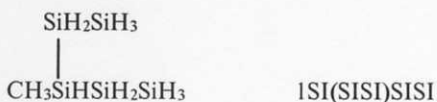
название по номенклатуре ИЮПАК составить легко: 1-метокси-2-метиламиноэтан, поэтому называть его по «а»-номенклатуре нецелесообразно, тогда как его линейно-цепные коды

10(1)-2N(1)    и  
 C,2O5N-6

имеют соответственно 11 и 8 знаков, и С-замена здесь необходима.

**Функции с большим числом одинаковых атомов.**

В кодах рассмотренных выше кремнийорганических соединений многократно повторяется символ одного и того же атома. Код такого соединения иногда можно сократить путем нумерации этих атомов. Для этого применяют тот же прием С-замены, представляя функцию, как углеводородный фрагмент. В соединении



атомы кремния нумеруют по правилам нумерации ациклических углеводородов [2], и если атомов в нем больше, чем в главном углеводородном фрагменте, считают главным фрагментом. На первом месте в коде помещают отдельное сообщение об общей С-замене данного атома в этом фрагменте, после которого следует его код как код соответствующего углеводорода:

C.SI-3(1)-5

(комбинацию CSI следует читать как «цезий-йод»). Критерий для перехода к этому способу кодирования тот же: уменьшение общего числа знаков. Здесь в «обычном» коде их 13, в коде с общей С-заменой их 11.

В соединении



главным является углеводородный фрагмент:

1(C.Si-,3-5)-1R

Коды имеют 16 и 15 знаков соответственно. Общая С-замена действует только в пределах своего фрагмента, не распространяясь ни на младшие, ни на

старшие фрагменты (ни внутрь, ни вне скобок). Сообщение о выходящей связи ставят после сообщения об общей С-замене. Коды соединения



имеют 10 и 11 знаков, и С-замену не производят.

В соединении  
 $\text{CH}_3\text{SiH}_2\text{OSiH}_2\text{NHSiH}_3$

цепь из атомов кремния прерывают атомы кислорода и азота. Этим атомам дают свои номера в общей цепи с атомами кремния (второй и четвертый), а информацию об этом приводят дополнительно в сообщении об общей С-замене:

C,SI2O4N-1(1)-5

Информацией о наличии такой дополнительной С-замены является запятая после символа углерода. Код этого же соединения без С-замены

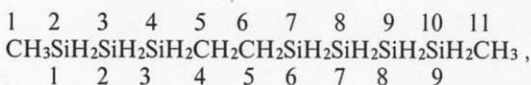
1SIOSINSI

имеет на шесть знаков меньше, но в соединении с большим числом атомов кремния

$\text{CH}_3\text{SiH}_2\text{SiH}_2\text{OSiH}_2\text{SiH}_2\text{NHSiH}_2\text{SiH}_2\text{SiH}_3$   
 1SISI0SISINSISISI  
 C,SI3O6N-1(1)-9

соотношение числа знаков изменяется в пользу кода с общей С-заменой (17 и 15).

Представляет интерес соединение



содержащее две связывающие функции с большим числом одинаковых атомов каждая. Здесь, по сравнению с обычным кодом

1SISISISI(1)-2SISISI(1)

общая С-замена

C.SI-1(2(C.SI-3(1))-4(1)

увеличивает число знаков на один. Короче, код с обычной С-заменой (20 знаков, нумерация над формулой)

C2,3,4,7,8,9,10SI-11

Этот код можно дополнительно сократить, применив общую С-замену и включив в цепь атомов кремния средние атомы углерода аналогично включению в такую цепь других гетероатомов (нумерация под формулой)

C,SI4,5-1,9(1)

Здесь в сообщении об общей С-замене отсутствует символ гетероатома после номеров. Это является информацией о том, что в цепи из атомов кремния атомы углерода с №№ 4 и 5 на другие атомы не заменяются. Последнее сообщение обозначает присоединение двух метильных групп к первому и последнему атомам нумерованной гетероцепи. В этом коде число знаков наименьшее – 14.

Таким образом, относительно простые правила и приемы линейно-цепного кодирования производных углеводородов и гетероциклических соединений и обозначение неуглеродных атомов в коде символами элементов периодической системы Д.И. Менделеева позволяют получить коды органических молекул высокой степени сложности без привлече-

ния дополнительного справочного материала. Специфическое кодирование ароматических и коеновых циклов и соответствующих им гетероциклов дает возможность не только уменьшить число знаков в коде, но и отразить в нем некоторые химические особенности соединений, хотя в кодировании так же, как и в номенклатуре [11] последнее делать не обязательно. Основная цель кодирования: в линейной форме, удобной для ввода в ЭВМ, однозначно описать органическую молекулу и иметь возможность составить по коду структурную формулу соединения.

#### ЛИТЕРАТУРА

1. Брюске Я.Э. Цепная нумерация и кодирование циклических углеводородов // Ж. структурной химии. 1995. Т. 36. № 4. С. 729-734.
2. Брюске Я.Э. Линейно-цепное кодирование и названия алифатических углеводородов // Вестн. ТГУ. Сер. Естеств. и технич. науки. Тамбов, 1996. Т. 1. Вып. 1. С. 34-38.
3. Брюске Я.Э. Линейно-цепное кодирование формул органических соединений. III. Кодирование сложных углеводородов // Вестн. ТГУ. Сер. Естеств. и технич. науки. Тамбов, 1997. Т. 2. Вып. 1. С. 53-56.
4. Брюске Я.Э. Линейно-цепное кодирование формул органических соединений. IV. Кодирование связей в углеводородах // Вестн. ТГУ. Сер. Естеств. и технич. науки. Тамбов, 1997. Т. 2. Вып. 1. С. 57-59.
5. Несмеянов А.Н., Несмеянов Н.А. Начала органической химии: В 2 т. М.: Химия, 1974.
6. Номенклатурные правила ИЮПАК по химии. М.: ВИНТИ, 1979. Т. 2. 896 с.
7. Кан Р., Дермер О. Введение в химическую номенклатуру. М.: Химия, 1983. 224 с.
8. Некрасов Б.В. Основы общей химии. М.: Химия, 1983. Т. 2. 688 с.
9. Ненцеску К.Д. Органическая химия. М.: ИЛ, 1963. Т. 2. 1048 с.
10. Бутлеров А.М. Введение к полному изучению органической химии // Сочинения. М.: Изд-во АН СССР, 1953. Т. 2. С. 257.
11. Терентьев А.П., Кост А.Н., Цукерман А.М., Попанов В.М. Номенклатура органических соединений. М.: Изд-во АН СССР, 1955. 304 с.

Поступила в редакцию 12 января 1998 г.