

УДК 004.81:510.64]:547.63

Н. В. Харчевникова, М. В. Максин, В. Г. Блинова,
Д. А. Добрынин, З. И. Жолдакова

12-17

Прогноз канцерогенности полициклических ароматических углеводородов с использованием квантовохимического модуля генерации метаболитов интеллектуальной ДСМ-системы¹

Для прогноза канцерогенности непрямым канцерогенов — полициклических ароматических углеводородов — применена интеллектуальная компьютерная система, основанная на ДСМ-методе автоматического порождения гипотез. Использован модуль генерации метаболитов с помощью квантовохимических расчетов. При проведении компьютерного эксперимента уровень канцерогенной активности исходного соединения приписывали конечному метаболиту, сгенерированному квантово-химическим модулем. Для описания структур метаболитов использован усовершенствованный вариант языка ФКСП (фрагментарный код суперпозиций подструктур). Обучающая выборка, содержащая 36 соединений, была разбита на три группы по канцерогенности. В результате компьютерного эксперимента сгенерированы гипотезы (структурные фрагменты), наличие которых в молекуле конечного метаболита определяет принадлежность исходного соединения к определенной группе канцерогенности. С использованием этих гипотез предсказана канцерогенность трех изученных соединений, не включенных в обучающую выборку. Результаты согласуются с экспериментальными данными. Предсказан уровень канцерогенной активности двух неизученных соединений.

неф

В последние годы для поиска связи структура — биологическая активность все шире используются экспертные системы. Однако точность прогноза такого важного вида активности как канцерогенность с использованием существующих в мире экспертных систем невысока [1,2]. По-видимому, это определяется сложностью процессов канцерогенеза и, в частности, тем, что канцерогенная активность многих ксенобиотиков определяется их биоактивацией в организме.

Наилучшие результаты прогноза канцерогенности в рамках организованного Национальным Институтом охраны окружающей среды США (NIHNS) открытого проекта Predictive-Toxicology Challenge (PTC) 2000–2001 [3] получены с помощью разработанной в отделе теоретических и прикладных проблем информатики ВИНТИ ДСМ-системы правдоподобных рассуждений [4,5], которая может логико-комбинаторными методами рассуждений выделить факторы, отвечающие за наличие либо отсутствие в соединении каких-либо заранее заданных свойств, например, канцерогенности. При этом обычно в качестве модели для рассуждений используется структурная формула молекул.

Применение ДСМ-системы дает хорошие ре-

зультаты при прогнозе канцерогенной активности прямых канцерогенов, т. е. веществ, активность которых определяется их непосредственным взаимодействием с ДНК. При прогнозе канцерогенной активности непрямым канцерогенов возникает необходимость учета биотрансформации соединений под действием ферментной системы, так как в этом случае активность определяется взаимодействием с ДНК не самого вещества, а его метаболита [6,7]. Логико-комбинаторная система, основанная на чисто структурном описании молекул, в этом случае не может дать правильный ответ на вопрос, будет ли обладать неизученное соединение канцерогенной активностью (недостаточная прогностическая способность). Это сказалось в частности на результатах эксперимента PTC, которые были хотя и лучшими из представленных в 2001 г., но точность прогноза была все же недостаточной.

Полициклические ароматические углеводороды (ПАУ) относятся к непрямым канцерогенам. В результате экспериментальных исследований выявлена определяющая роль метаболической активации в канцерогенной активности этого важного класса чужеродных веществ, воздействующих на

¹ Работа выполнена при поддержке РФФИ (проект № 02-01-00537а)

человека в среде обитания и на производстве. Воздействие ПАУ происходит при вдыхании табачного дыма, загрязненного воздуха, при употреблении загрязненной воды и подвергнутой высокотемпературной обработке пищи [8–11]. Многие процессы сжигания, коксовое производство, очистка нефти, двигатели автомобилей, домашние нагревательные приборы обуславливают попадание ПАУ в окружающую среду. При лесных пожарах ПАУ накапливаются в атмосфере. Вулканическая активность также может быть источником загрязнения окружающей среды ПАУ. Основным источником загрязнения воздушной среды городов ПАУ является автомобильный транспорт [9, 10].

Установлено, что многие представители этого класса соединений являются инициаторами или промоторами опухолей, канцерогенами и мутагенами. Канцерогенная и мутагенная активность различных ПАУ варьируется в очень широких пределах в зависимости от структуры углеводорода [8, 9, 11].

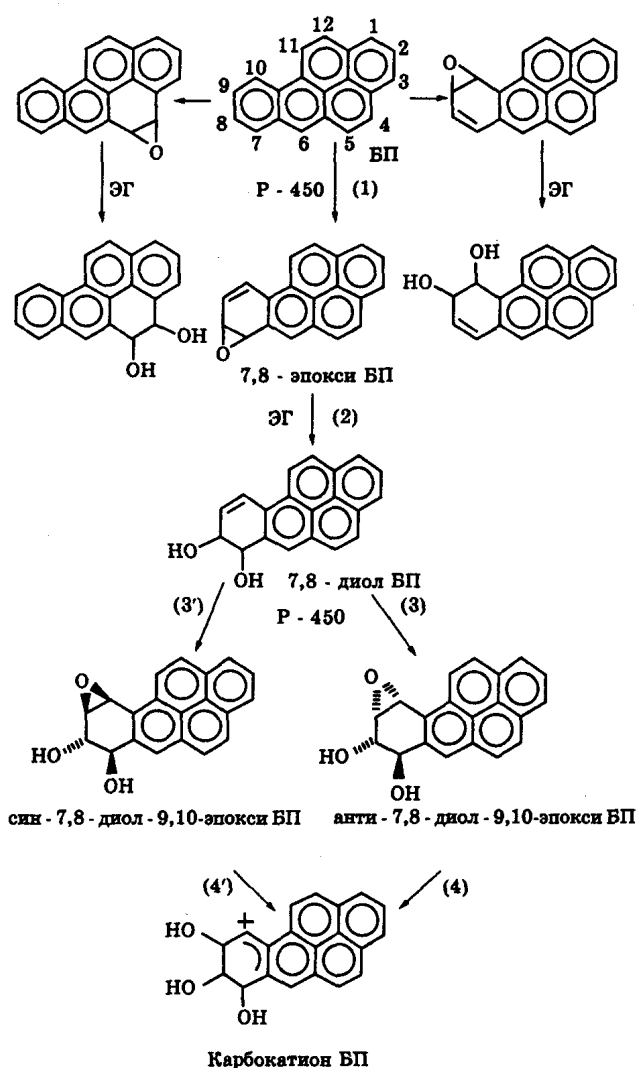


Рис. 1. Схема метаболизма и биоактивации ПАУ на примере бенз(а)пирена

Основные черты биоактивации всех ПАУ аналогичны (рис. 1). На первом этапе в результате окисления ПАУ с помощью цитохрома P-450 образуются эпоксиды, например, из бенз(а)пирена — 7,8-эпоксибенз(а)пирен. Под действием фермента эпоксигидролазы (ЭГ) эпоксиды разлагаются до дигидродиолов. В частности, из 7,8-эпоксибенз(а)пирена получится 7,8-дигидро-

оксибенз(а)пирен. Дигидродиолы подвергаются повторному ферментативному окислению с образованием диолэпоксидов. В случае бенз(а)пирена важнейшим продуктом второго окисления являются 7,8-дигидро-9,10-эпоксибенз(а)пирены, размеры сопряженной π -электронной системы которых меньше, чем в исходном бенз(а)пирене, на одно бензольное кольцо. Наиболее активные соединения в этом ряду метаболитов — эпоксигидродиолы ПАУ. Именно они способны образовывать аддукты с ДНК. Диолэпоксид атакует критические позиции ДНК по механизму нуклеофильного замещения первого порядка (S_N1) с образованием в качестве интермедиатов триолкарбокатионов. Скорость процесса биоактивации ПАУ с образованием аддуктов с ДНК лимитируется образованием карбокатионов из диолэпоксидов. Отсюда следует, что биологическое действие ПАУ — мутагенное или канцерогенное — будет тем сильнее, чем меньше требуется энергии для образования карбокатиона из диолэпоксида [12–13]. Конечным канцерогеном будет карбокатион с наименьшей энергией образования, т. е. наиболее стабильный относительно диолэпоксида.

В ходе окисления, как можно видеть, структурная формула вещества значительно изменяется, а поскольку активность определяется именно свойствами метаболитов или характеристиками реакций их образования, при анализе соотношений структура — канцерогенная активность необходимо использовать данные по структуре метаболитов, а не исходных соединений. Следовательно, для изучения канцерогенности ПАУ требуется иная модель, не сводимая к поиску структурного сходства с исходным веществом, а связывающая конечный метаболит данного соединения с уровнем канцерогенности исходного вещества.

Целью данной работы является апробация метода прогноза канцерогенной активности ПАУ с использованием интеллектуальной ДСМ-системы, при котором компьютерный эксперимент проводится на структурах конечных стабильных метаболитов, а не исходных соединений.

На первом этапе задача состоит в выявлении наиболее устойчивого карбокатиона данного полиароматического углеводорода, для чего был создан программный модуль, основанный на расчете энергий метаболитов методом Хюккеля. Реализация модуля подробно описана в нашей предыдущей работе [14]. Кратко изложим алгоритм нахождения наиболее устойчивого карбокатиона. Необходимо отметить, что на практике энергию образования карбокатиона удобно вычислять относительно тетраола, так как при использовании метода Хюккеля энергия тетраола совпадает с энергией диолэпоксида. Для нахождения наиболее устойчивого метаболита необходимо выполнить следующее.

1. Найти все внешние циклы (т. е. циклы, имеющие с молекулой одно и только одно общее ребро). В рамках данной задачи считается, что только такие циклы могут быть окислены с образованием диолэпоксидов (см. рис. 1).

1.1. Для каждого цикла получить соответствующий ему тетраол. Для получения тетраола связь в цикле разрывается и к каждому внешнему атому углерода присоединяется группа OH; молекула при этом перестаёт быть “плоской” и положения групп H и OH относительно плоскости цикла (“сверху”—“снизу”) чередуются для соседних

атомов С. Вычислить энергию тетраола $E_{тетраол}^i$ (процедура вычисления энергии ПАУ будет описана ниже).

1.2. Из тетраола строим два производных карбокатиона $M_{катион}^{ij}$, $j = 1, 2$, "восстанавливая" ароматическую связь по очереди у каждого из атомов С, принадлежащих общему ребру. Вычисляем энергии карбокатионов $E_{катион}^{ij}$, $j = 1, 2$ и энергии их образования $\Delta E_{катион}^{ij} = E_{катион}^{ij} - E_{тетраол}^i$, $j = 1, 2$. В действительности $\Delta E_{катион}^{ij} < 0$, что говорит об энергетической выгоды такой трансформации.

2. Найти максимальную по модулю энергию образования: $\Delta E_{катион} = \max_{i,j} |\Delta E_{катион}^{ij}|$ и восстановить соответствующий карбокатион $M_{катион}$, который и считается наиболее стабильным.

Пример нахождения наиболее стабильного карбокатиона приведен на рис. 2.

Конечный метаболит триолкарбокатион является хотя и наиболее устойчивым из возможных, но все же короткоживущим интермедиатом реакции электрофильного замещения, реагентами которой

являются диолэпоксид и нуклеиновое основание ДНК. Как видно из рис. 2, каждому карбокатиону соответствует один диолэпоксид. Компьютерный эксперимент проводили, используя в качестве объектов структуры диолэпоксидов — предшественников наиболее стабильных карбокатионов. С помощью языка ФКСП каждый метаболит автоматически представлялся в виде числового множества дескрипторов.

В системе применена новая версия ФКСП кодировщика, позволяющая при кодировании циклических структур (в том числе полиароматических углеводородов) учитывать не только максимальную структуру, но и все не максимальные подструктуры [15].

В ходе апробации метода для 41 соединения класса полициклических ароматических углеводородов с использованием квантово-химического модуля были определены наиболее активные конечные метаболиты — карбокатионы. На массиве метаболитов — диолэпоксидов, предшественников этих наиболее устойчивых карбокатионов (см. рис. 2), был проведен компьютерный ДСМ-эксперимент. Метаболитам соединений был приписан

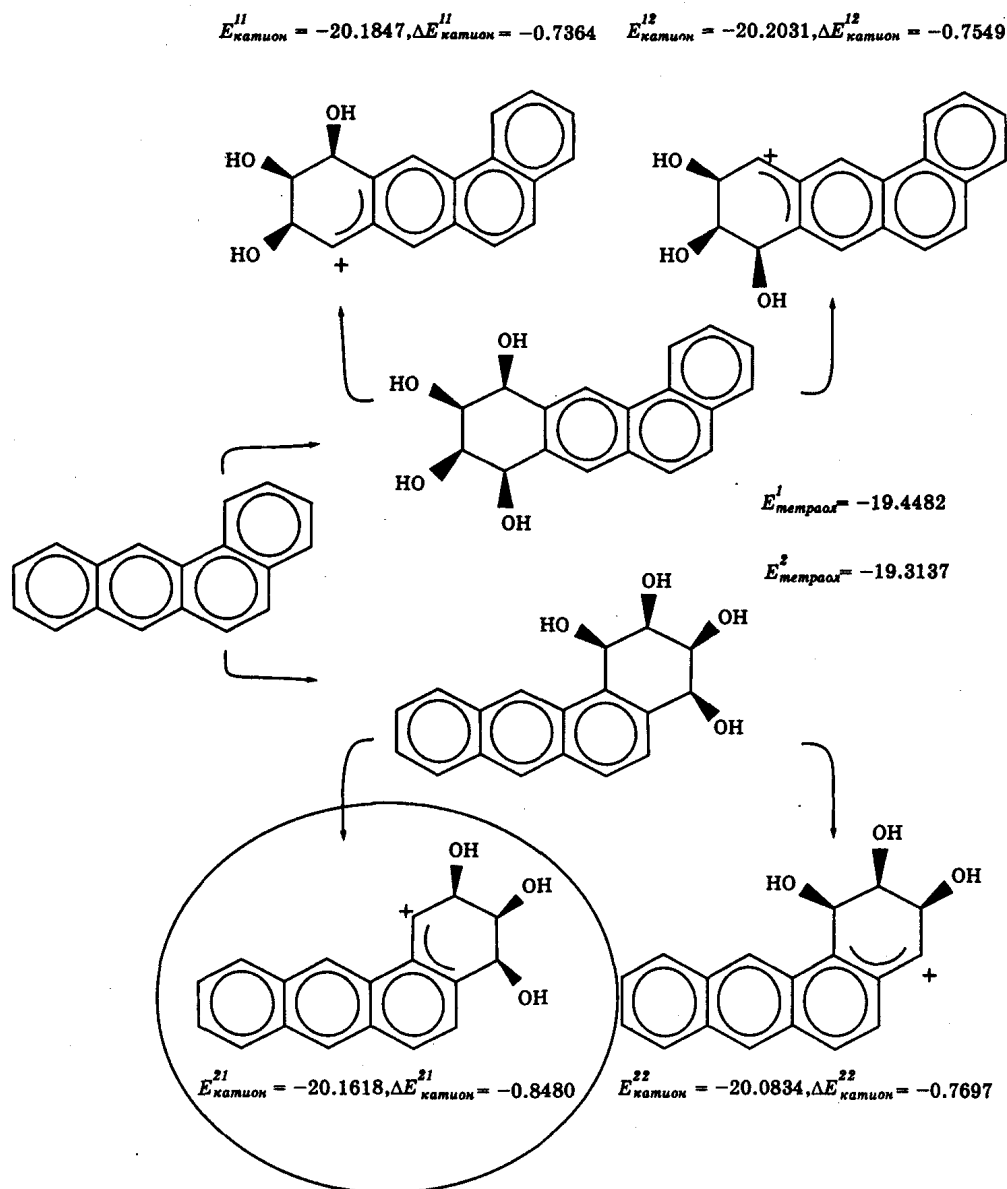


Рис. 2. Определение наиболее стабильного, т. е. конечного метаболита бенз(а)антрацена (обведен кружком)

экспериментально определенный уровень канцерогенной активности исходных ПАУ. Канцерогенная активность ПАУ характеризуется по Бэджеру, т. е. по результатам экспериментов на млекопитающих соединения относятся к одной из пяти групп: неактивные (—), слабо активные (+), умеренно активные (++) , сильно активные (+++) и чрезвычайно активные (++++). При проведении ДСМ-эксперимента соединения с очень высокой и чрезвычайно высокой канцерогенной активностью были объединены в одну группу высокоактивных соединений, соединения довольно и слабо активные также были объединены в одну группу — умеренно активных соединений.

На первом этапе проведен эксперимент на массиве 23 соединений (соединения 1–23 в табл. 1). Три соединения — пентацен (не канцероген), бенз(а)пирен (высоко канцерогенное соединение), бензо(а)нафто(2,1,8-*hi*)нафтацен (слабо канцерогенное соединение) — не включались в тестовую выборку: их активность надо было предсказать.

Эксперимент проводился с использованием двух свойств, которые определяют разбиение соединений по канцерогенности на три группы: отсутствие канцерогенности, умеренная канцерогенность и высокая канцерогенность. Положительные примеры по первому свойству — метаболиты соединений с высокой и умеренной канцерогенной активностью. Отрицательные примеры — метаболиты не канцерогенных соединений. По второму свойству положительные примеры — метаболиты высоко канцерогенных соединений, отрицательные примеры — метаболиты умеренно и не канцерогенных ПАУ.

Были выделены гипотезы (структурные фрагменты молекул метаболитов), определяющие принадлежность соединения к определенной группе канцерогенности. Гипотезы, с помощью которых осуществлен прогноз, и номера тех соединений в обучающей выборке (см. табл. 1), метаболиты которых содержат эти структурные фрагменты, приведены в табл. 2.

Таблица 1

Канцерогенная активность ПАУ [11, 16] и оценка активности в ДСМ-эксперименте

№ п/п	Соединение	Оценка канцерогенной активности по результатам эксперимента	Оценка канцерогенной активности в ДСМ-эксперименте
1	Нафталин	—	не активен
2	Антрацен	—	не активен
3	Бензо(с)фенантрен	+	умеренно активен
4	Тетрацен	—	не активен
5	Хризен	+	умеренно активен
6	Бензо(б)хризен	—	не активен
7	Фенантрен	—	не активен

№ п/п	Соединение	Оценка канцерогенной активности по результатам эксперимента	Оценка канцерогенной активности в ДСМ-эксперименте
8	Пицен	—	не активен
9	Трифенилен	—	не активен
10	Нафто(2,3- <i>b</i>)пирен	++	умеренно активен
11	Пентацен	Прогноз (-)	Прогноз
12	Бензо(е)пирен	+	умеренно активен
13	Дибенз(а,с)антрацен	+	умеренно активен
14	Дибенз(а,і)антрацен	+	умеренно активен
15	Бензо(г)хризен	++	умеренно активен
16	Дибенз(а,н)антрацен	++	умеренно активен
17	Бенз(а)антрацен	+	умеренно активен
18	Дибензо(а,е)пирен	+++	высоко активен
19	Гексацен	—	не активен
20	Бензо(а)пирен	Прогноз (++++)	Прогноз
21	Дибензо(а,і)пирен	++++	высоко активен
22	Трибензо(а,е,і)пирен	++	умеренно активен
23	Дибензо(а,н)пирен	++++	высоко активен
24	Бензо(а)тетрацен	—	не активен
25	Дибензо(а,і)пирен	++++	высоко активен
26	Дибензо(е,і)пирен	+	умеренно активен
27	Трибензо(а,е,н)пирен	+	умеренно активен
28	Бензо(а)пентацен	—	не активен
29	Бензо(с)хризен	+	умеренно активен
30	Трибензо(а,с,н)антрацен	+	умеренно активен
31	Флуорантен	—	не активен
32	Пентафен	—	не активен
33	Бензо(с)пентафен	—	не активен
34	Гексафен	—	не активен
35	Дибензо(с,г)фенантрен	—	не активен
36	Нафто(2,3- <i>e</i>)пирен	—	не активен
37	Бензокоронен	—	не активен
38	Бензо(а)нафто(2,1,8- <i>hi</i>)нафтацен	Прогноз (+)	Прогноз
39	Дибензо(г,р)хризен	Прогноз	Прогноз
40	Дибензо(с,і)тетрацен	Прогноз	Прогноз

Структура конечных метаболитов соединений, представленных на прогноз, структурные фрагменты молекул метаболитов, с помощью которых осуществлен прогноз (гипотезы), и номера тех соединений в обучающей выборке (см. табл. 1), метаболиты которых содержат эти структурные фрагменты (метаболиты являются соединениями-родителями)

№ в табл. 1	Структура конечного метаболита соединения	Гипотезы и номера соединений-родителей	
		Положительные гипотезы	Отрицательные гипотезы
11			 19, 24
20		 10, 12, 21, 25 10, 12, 18, 21, 23, 25 5, 23, 25	
39			 10, 24 19, 24
40		 15, 18, 21, 23 15, 18, 23, 33	 3, 5, 6, 7, 8, 10, 12, 14, 15, 16, 29, 30, 32, 33, 35, 36, 38, 39 27, 33
41			 10, 24 3, 5, 6, 7, 8, 10, 12, 14, 15, 16, 29, 30, 32, 33, 35, 36, 38, 39

С использованием полученных гипотез канцерогенная активность пентацена была предсказана правильно, канцерогенная активность

бенз(а)пирена была определена не полностью — он был отнесен к высоко или умеренно активным соединениям, бензо(а)нафто(2,1,8-*hij*)нафтацен был отнесен к умеренно активным или неактивным соединениям. В целом прогноз согласуется с экспериментальными данными.

Затем массив обучающей выборки ПАУ был расширен до 39 соединений (см. табл. 1). Канцерогенную активность двух неизученных экспериментально ПАУ — дибензо(г,р)хризена и дибензо(а,д)тетрацена надо было предсказать. В результате компьютерного эксперимента первое из этих соединений было отнесено к группе умеренно активных, второе соединение определено не полностью, по результатам прогноза оно должно быть либо умеренно канцерогенным, либо не обладать канцерогенной активностью. В целом результаты совпадают с полученными в работе [16] с использованием регрессионной модели. Канцерогенная активность в [16] характеризовалась индексами Иболла, пропорциональными доле животных в группе, у которых при воздействии вещества возникло заболевание, деленной на латентный период развития опухоли. Уравнение включало три независимых переменных, две из которых — квантово-химические параметры, характеризующие процессы активации и детоксикации ПАУ, а третья отражает размер молекулы. Индексы Иболла соединений обучающей выборки находились в интервале 0–76. Для двух неизученных соединений были предсказаны значения индексов Иболла 18 и 12, что соответствует умеренной активности по принятому в нашей работе отнесению к группам неактивных, умеренно активных и высокоактивных соединений.

Результат свидетельствует об адекватности исходной модели и правомерности применения логико-комбинаторного ДСМ-метода с поиском сходства на структурах метаболитов, а не исходных соединений. Однако точность прогноза может быть повышена при использовании также и поиска сходства на числовых параметрах — энергетических характеристиках метаболитов. Разность энергии карбокатиона и диолэпоксида характеризует реакционную способность метаболита, и канцерогенная активность в ряду ПАУ зависит от этого параметра. Компьютерные эксперименты с использованием такого комбинированного подхода в настоящее время проводятся.

Можно сделать вывод, что для соединений, канцерогенная активность которых определяется биоактивацией (непрямые канцерогены), необходимо сначала определять наиболее активный конечный метаболит, а затем проводить анализ структура-активность с использованием структуры и энергетических характеристик этих метаболитов. Определение конечного метаболита должно проводиться с помощью квантово-химических расчетов. Таким образом, перспективным направлением для прогноза канцерогенности является соединение логико-комбинаторного подхода и квантово-химических расчетов.

СПИСОК ЛИТЕРАТУРЫ

1. Bristol D. W., Washman J. T., Greenwell A. Introduction: The NIEHS Predictive-Toxicology Evaluation Project // Environmental Health Perspectives.— 1996.— Vol. 104 (Supplement 5).— P. 1001-1010.

2. Srinivasan A., King R. D., Muggleton S. H., Stermberg M. J. E. The predictive toxicology evaluation challenge // Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence, 1997.— P. 46–53.
3. Woo Y.-T. Predictive toxicology challenge (PTC) 2000–2001: A toxicologist's view // Proceedings of Predictive Toxicology Challenge (PTC) Workshop at the 5th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2001).— September 6.— Freiburg, 2001.— P. 32–37.
4. Кузнецов С. О. ДСМ-метод как система автоматического обучения // Итоги науки и техники. Сер. Информатика. Т. 15.— М.: ВИНТИ, 1991.— С. 17–53.
5. Финн В. К. Правдоподобные рассуждения в интеллектуальных системах типа ДСМ // Итоги науки и техники. Сер. Информатика. Т. 15.— М.: ВИНТИ, 1991.— С. 54–101.
6. Панкратова Е. С. Пути развития ДСМ-системы в различных моделях канцерогенеза // НТИ. Сер. 2.— 1993.— № 1.— С. 21–22.
7. Панкратова Е. С. Применение ДСМ-метода к задаче распознавания прямых и непрямых канцерогенов // НТИ. Сер. 2.— 1993.— № 3.— С. 14–16.
8. Канцерогенные вещества: Справочник. Материалы Международного агентства по изучению рака / Пер. с англ. А. Ф. Карамышевой.— М.: Медицина, 1987.— 332 с.
9. Шабад Л. М. Бластомогенная опасность химических загрязнений среды обитания человека // Итоги науки и техн. ВИНТИ. Сер. Токсикология. Т. 9.— М.: ВИНТИ, 1978.— С. 7–55.
10. Hoffmann P., Wynder E. L. Organic particulate pollutants chemical analysis and bioassays for carcinogenicity // Air Pollution / Ed. A. C. Stern.— Academic Press, 1977.— Vol. 2.— P. 374–383.
11. Дьячков П. Н. Квантово-химические расчеты в изучении механизма действия и прогнозе токсичности чужеродных соединений // Итоги науки и техники. Сер. Токсикология. Т. 16.— М.: ВИНТИ, 1990.— 280 с.
12. Jerina D. M., Lehr R. E., Yagi H. Mutagenicity of benz(a)pyrene derivatives and the description of a quantum mechanical model which predicts the case of carbonium ion formation from diol epoxides // In vivo metabolic activation and mutagenesis testing / Ed. De Serres F. — 1976.— P. 159–177.
13. Lehr R. E., Kumar S., Levin W., Jerina D. M. The bay region theory of polycyclic aromatic hydrocarbons carcinogenesis // ACS Symposium Series.— 1985.— Vol. 283.— P. 63–84.
14. Максим М. В., Харчевникова Н. В. Квантово-механический модуль системы, реализующей комбинаторно-численный подход к проблеме прогнозирования свойств химических соединений // НТИ. Сер. 2.— 2002.— № 6.— С. 25–31.
15. Блинова В. Г., Добрынин Д. А. Язык ФКСП описания химической структуры соединения // НТИ. Сер. 2.— 2001.— № 6.— С. 14–21.
16. Von Szentpaly L. Carcinogenesis by polycyclic aromatic hydrocarbons: a multilinear regression on new type PMO indices // J. Am. Chem. Soc.— 1984.— Vol. 106.— № 20.— 6021–6028.

Материал поступил в редакцию 19.09.03.