



УДК 547.458.2.02:577.114.012.7

© 1992 г.

*Н. К. Кочетков, Е. В. Виноградов, Ю. А. Книрель,
А. С. Шашков, Г. М. Липкинд*

ПРОГРАММА SCAN ДЛЯ СТРУКТУРНОГО АНАЛИЗА ЛИНЕЙНЫХ ПОЛИСАХАРИДОВ НА ОСНОВЕ ДАННЫХ ^{13}C -ЯМР-СПЕКТРОВ С ИСПОЛЬЗОВАНИЕМ ПЕРСОНАЛЬНЫХ КОМПЬЮТЕРОВ

Институт органической химии им. Н. Д. Зелинского РАН, Москва

Создана компьютерная программа SCAN для структурного анализа линейных регулярных полисахаридов по данным ^{13}C -ЯМР-спектров, представляющая собой модификацию ранее разработанной программы (Lipkind G. M., Shashkov A. S., Knirel Y. A., Vinogradov E. V., Kochetkov N. K. // Carbohydr. Res. 1988. V. 175. № 1. P. 59–75), предназначенную для работы на персональных компьютерах, совместимых с IBM PC. Программа SCAN использована для установления строения 24 полисахаридов бактериального происхождения. Рассмотрены оптимальные варианты сочетания компьютерного метода с другими методами анализа углеводов и обсуждены границы применимости программы SCAN.

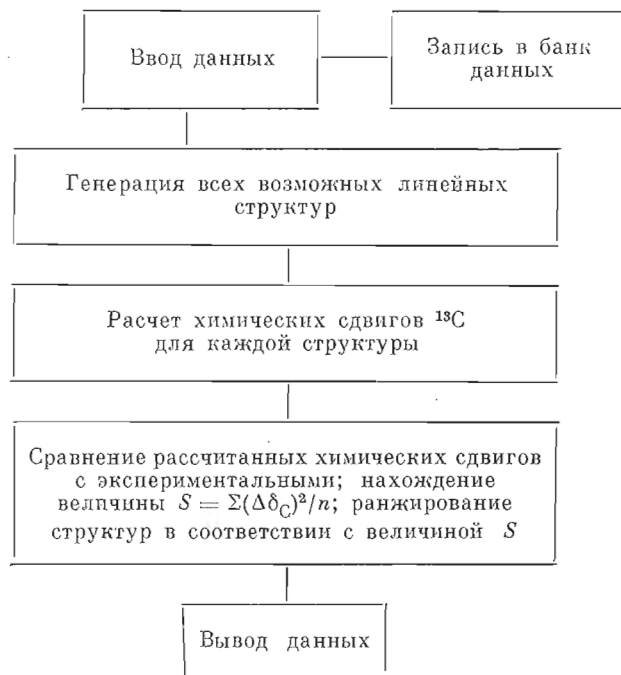
В последние годы развиваются методы структурного анализа углеводов, основанные на компьютерной обработке ^{13}C -ЯМР-спектров [1–6]. Так, разработанная нами ранее компьютерная программа для анализа линейных регулярных полисахаридов основана на расчете ^{13}C -ЯМР-спектров для всех возможных первичных структур по аддитивной схеме, исходя из химических сдвигов ^{13}C свободных моносахаридов и эффектов гликозилирования, с последующим нахождением структуры, для которой рассчитанный спектр наиболее близок экспериментальному [3]. Эта программа была успешно применена для установления строения ряда бактериальных полисахаридов [3, 5]. В настоящем сообщении описана новая программа SCAN (Structural Computerized Analysis), которая представляет собой модификацию нашей более ранней программы [3], предназначенную для работы на персональных компьютерах, совместимых с IBM PC.

В программе SCAN в основном сохранен ранее разработанный алгоритм решения задачи для линейных регулярных полисахаридов (схема), включающих в повторяющееся звено от 1 до 4 моносахаридных остатков. (Возможен структурный анализ линейных полисахаридов и с большими повторяющимися звеньями, однако они встречаются у бактерий относительно редко, а их анализ требует заметно большего компьютерного времени.)

Генерация теоретически возможных структур линейных полисахаридов с данным моносахаридным составом проводится путем перебора всех аномерных конфигураций, положений замещения и последовательности моносахаридных остатков. Дополнительно предусмотрена возможность ограничения числа рассматриваемых вариантов путем указания положений замещения моносахаридных остатков, установленных, как правило, в результате анализа методом метилирования.

Расчет теоретических спектров полисахаридов базируется на концепции дисахаридного приближения [6], согласно которой химические сдвиги ^{13}C в дисахаридах и в аналогичных дисахаридных фрагментах линей-

Диаграмма потока структурного анализа полисахаридов с помощью программы SCAN



ных полисахаридов одинаковы или различаются на величину, которой можно пренебречь при расчете (как правило, $\leq 0,5$ м.д.). Используемая для расчета база данных включает в себя эффекты гликозилирования на C1, зависящие от природы соседнего моносахаридного остатка и места присоединения, и усредненные химические сдвиги ^{13}C для α - и β -пиранозных форм* моносахаридных остатков, замещенных α - и β -гликопиранозильными остатками в положения 2, 3, 4 или 6, с поправками на C2 и C5 (табл. 1).

Список моносахаридов, для которых может быть проведен расчет, расширен по сравнению с работой [3] и теперь включает в себя все гексозы, 6-дезоксигексозы, гексуроновые кислоты, 2-ациламино-2-дезоксигексозы и 2-ациламино-2,6-дидезоксигексозы, сколько-нибудь часто встречающиеся в составе бактериальных полисахаридов. Использование в базе данных усредненных химических сдвигов C2—C6 вместо применявшихся в работе [3] эффектов гликозилирования позволяет избежать многочисленных операций сложения, что имеет существенное значение при работе с персональными компьютерами, обладающими невысоким быстродействием. Многие химические сдвиги, составляющие новую базу данных, уточнены на основании появившихся в последнее время в литературе данных по химическим сдвигам ^{13}C -ЯМР в дисахаридных фрагментах синтетических и природных олигосахаридов и полисахаридов; процесс уточнения базы данных продолжается и в настоящее время.

Химические сдвиги ^{13}C замещенных моносахаридных остатков и эффекты гликозилирования на C1 дифференцированы в зависимости от того, находится ли соседний моносахаридный остаток в той же или противоположной энантиомерной форме, имеет ли он *α*- (*глюко*-) или

* По мере накопления данных по химическим сдвигам в базу данных могут быть включены и фуранозные формы моносахаридов.

Матрицы эффектов гликозилирования на С1'(ΔС1') и химических сдвигов С1-С6 для замещенных остатков глюкозы (м.д.)

Гликозилирующий моносахарид	Конфигурация и тип связи *	ΔС1'	С1	С2	С3	С4	С5	С6	
D-Гексоза (глюко/галакто)	α-(1-2)-α	3,4	91,5	76,7	72,7	70,9	73,2	61,9	
	α-(1-3)-α	6,7	93,3	71,5	80,8	70,8	73,2	61,9	
	α-(1-4)-α	7,6	93,3	72,7	74,3	78,3	71,9	61,9	
	α-(1-6)-α	5,6	93,3	72,7	74,0	70,9	71,4	66,9	
	β-(1-2)-α	7,6	92,6	81,7	72,7	70,9	73,2	61,9	
	β-(1-3)-α	6,7	93,3	72,0	83,5	69,4	73,2	61,9	
	β-(1-4)-α	6,3	93,3	72,7	72,6	79,9	71,9	61,1	
	β-(1-6)-α	6,7	93,3	72,7	74,0	70,9	72,2	69,7	
	α-(1-2)-β	5,6	97,4	79,1	75,7	70,9	77,2	62,1	
	α-(1-3)-β	6,7	97,1	73,2	83,8	70,8	77,2	62,1	
	α-(1-4)-β	7,6	97,1	74,4	77,3	78,3	75,9	62,1	
	α-(1-6)-β	5,6	97,1	74,4	77,0	70,9	75,4	66,7	
	β-(1-2)-β	7,0	95,5	81,9	77,0	70,9	77,2	62,1	
	β-(1-3)-β	6,7	97,1	73,7	86,5	69,4	77,2	62,1	
	β-(1-4)-β	6,3	97,1	74,4	75,6	79,9	75,9	61,3	
	β-(1-6)-β	6,7	97,1	74,4	77,0	70,9	76,2	69,9	
	D-Гексоза (манно)	α-(1-2)-α	3,4	91,5	76,7	72,7	70,9	73,2	61,9
		α-(1-3)-α	6,7	93,3	71,5	80,8	71,8	73,2	61,9
		α-(1-4)-α	7,1	93,3	72,7	74,3	78,3	71,9	61,9
		α-(1-6)-α	5,6	93,3	72,7	74,0	70,9	71,4	66,9
β-(1-2)-α		7,6	92,6	81,7	72,7	70,9	73,2	61,9	
β-(1-3)-α		6,7	93,3	72,0	83,5	69,4	73,2	61,9	
β-(1-4)-α		7,0	93,3	72,7	72,6	79,9	71,9	61,1	
β-(1-6)-α		6,7	93,3	72,7	74,0	70,9	72,2	69,7	
α-(1-2)-β		5,6	97,4	79,1	75,7	70,9	77,2	62,1	
α-(1-3)-β		6,7	97,1	73,2	83,8	71,8	77,2	62,1	
α-(1-4)-β		7,1	97,1	74,4	77,3	78,3	75,9	62,1	
α-(1-6)-β		5,6	97,1	74,4	77,0	70,9	75,4	67,1	
β-(1-2)-β		7,0	95,5	82,4	77,0	70,9	77,2	62,1	
β-(1-3)-β		6,7	97,1	73,7	86,5	69,4	77,2	62,1	
β-(1-4)-β		7,0	97,1	74,4	75,6	79,9	75,9	61,3	
β-(1-6)-β		6,7	97,1	74,4	77,0	70,9	76,2	69,9	
2-Ациламино-2-дезоксид-гексоза		α-(1-2)-α	3,4	91,5	76,7	72,7	70,9	73,2	61,9
		α-(1-3)-α	6,7	93,3	71,5	80,8	70,8	73,2	61,9
		α-(1-4)-α	7,6	93,3	72,7	74,3	78,3	71,9	61,9
		α-(1-6)-α	5,6	93,3	72,7	74,0	70,9	71,4	66,9
	β-(1-2)-α	7,6	92,6	81,7	72,7	70,9	73,2	61,9	
	β-(1-3)-α	6,7	93,3	72,0	83,5	69,4	73,2	61,9	
	β-(1-4)-α	6,3	93,3	72,7	72,6	79,9	71,9	61,1	
	β-(1-6)-α	6,7	93,3	72,7	74,0	70,9	72,2	67,9	
	α-(1-2)-β	5,6	97,4	79,1	75,7	70,9	77,2	62,1	
	α-(1-3)-β	6,7	97,1	73,2	83,8	70,8	77,2	62,1	
	α-(1-4)-β	7,6	97,1	74,4	77,3	78,3	75,9	62,1	
	α-(1-6)-β	5,6	97,1	74,4	77,0	70,9	75,4	66,7	
	β-(1-2)-β	7,0	95,5	80,3	78,0	70,9	77,2	62,1	
	β-(1-3)-β	6,7	97,1	73,7	86,5	69,4	77,2	62,1	
	β-(1-4)-β	6,3	97,1	74,4	75,6	79,9	75,9	61,3	
	β-(1-6)-β	6,7	97,1	74,4	77,0	70,9	76,2	68,1	
	L-Гексоза (глюко/галакто)	α-(1-2)-α	8,3	92,8	81,3	73,1	70,9	73,2	61,9
		α-(1-3)-α	7,1	93,3	71,5	80,6	69,6	73,2	61,9
		α-(1-4)-α	6,4	93,3	72,7	72,7	78,4	72,6	61,9
		α-(1-6)-α	6,7	93,3	72,7	74,0	70,9	72,0	68,4

Таблица 1 (окончание)

Гликозилирующий моносахарид	Конфигурация и тип связи *	ΔC1'	C1	C2	C3	C4	C5	C6
<i>L</i> -Гексоза (<i>манно</i>)	β-(1-2)-α	4,0	91,5	78,7	72,7	70,9	73,2	61,9
	β-(1-3)-α	6,7	93,3	74,5	83,5	70,8	73,2	61,9
	β-(1-4)-α	7,2	93,3	72,7	74,1	77,6	71,8	61,9
	β-(1-6)-α	6,1	93,3	72,7	74,0	70,9	72,2	68,9
	α-(1-2)-β	7,1	95,8	79,6	77,7	70,9	77,2	62,1
	α-(1-3)-β	7,1	97,1	73,2	83,6	69,6	77,2	62,1
	α-(1-4)-β	6,4	97,1	74,4	75,7	78,4	76,6	61,5
	α-(1-6)-β	6,7	97,1	74,4	77,0	70,9	76,0	68,6
	β-(1-2)-β	7,0	97,4	81,9	77,0	70,9	77,2	62,1
	β-(1-3)-β	6,7	97,1	73,2	86,5	70,8	77,2	62,1
	β-(1-4)-β	7,2	97,1	74,4	77,1	77,6	75,8	62,1
	β-(1-6)-β	6,1	97,1	74,4	77,0	70,9	76,2	69,1
	α-(1-2)-α	8,3	92,8	81,3	73,1	70,9	73,2	61,9
	α-(1-3)-α	7,1	93,3	71,5	80,4	69,6	73,2	61,9
	α-(1-4)-α	6,4	93,3	72,7	72,7	78,9	72,6	61,9
	α-(1-6)-α	6,7	93,3	72,7	74,0	70,9	72,0	68,4
	β-(1-2)-α	4,0	91,5	78,7	72,7	70,9	73,2	61,9
	β-(1-3)-α	6,7	93,3	71,5	83,5	71,8	73,2	61,9
	β-(1-4)-α	7,2	93,3	72,7	74,1	77,6	71,8	61,9
	β-(1-6)-α	6,1	93,3	72,7	74,0	70,9	72,2	68,9
	α-(1-2)-β	7,1	95,8	81,2	77,7	70,9	77,2	62,1
	α-(1-3)-β	7,1	97,1	73,2	83,4	69,6	77,2	62,1
	α-(1-4)-β	6,4	97,1	75,3	75,7	78,9	76,6	62,1
	α-(1-6)-β	6,7	97,1	74,4	77,0	70,9	76,0	68,6
β-(1-2)-β	7,0	97,4	81,0	76,8	70,9	77,2	62,1	
β-(1-3)-β	6,7	97,1	73,2	86,5	71,8	77,2	62,1	
β-(1-4)-β	7,2	97,1	74,4	77,1	77,6	75,8	62,1	
β-(1-6)-β	6,1	97,1	74,4	77,0	70,9	76,2	69,1	
2-Ациламино-2-дезоксип- <i>L</i> -гексоза	α-(1-2)-α	8,3	92,8	81,3	73,1	70,9	73,2	61,9
	α-(1-3)-α	7,1	93,3	71,5	80,6	69,6	73,2	61,9
	α-(1-4)-α	6,4	93,3	72,7	72,7	78,4	72,1	60,9
	α-(1-6)-α	6,7	93,3	72,7	74,0	70,9	72,0	66,9
	β-(1-2)-α	4,0	91,5	78,7	72,7	70,9	73,2	61,9
	β-(1-3)-α	6,7	93,3	71,5	83,5	70,8	73,2	61,9
	β-(1-4)-α	7,2	93,3	72,7	74,1	77,6	71,8	61,9
	β-(1-6)-α	6,1	93,3	72,7	74,0	70,9	72,2	68,9
	α-(1-2)-β	7,1	95,8	79,6	77,7	70,9	77,2	62,1
	α-(1-3)-β	7,1	97,1	73,2	83,6	69,6	77,2	62,1
	α-(1-4)-β	6,4	97,1	74,4	75,7	78,4	76,6	61,5
	α-(1-6)-β	6,7	97,1	74,4	77,0	70,9	76,0	67,1
	β-(1-2)-β	7,0	97,4	81,9	75,7	70,9	77,2	62,1
	β-(1-3)-β	6,7	97,1	73,2	86,5	70,8	77,2	62,1
	β-(1-4)-β	7,2	97,1	74,4	77,1	77,6	75,8	62,1
	β-(1-6)-β	6,1	97,1	74,4	77,0	70,9	76,2	69,1

* Слева направо указаны конфигурация гликозилирующего моносахаридного остатка, положение замещения и конфигурация остатка глюкозы.

манно-конфигурацию (экваториальный или аксиальный заместитель при C2) и является ли он производным гексозы или 2-ациламино-2-дезоксигексозы, что отвечает присутствию гидроксильной или ациламиногруппы в положении 2 соответственно; из-за недостатка данных для N-ацилгексозаминов не сделано различия в конфигурации при C2. Таким образом, для каждого из моносахаридов в базе данных содержится шесть матриц

[1 : 16] с эффектами гликозилирования на C1' и шесть матриц [1 : 6, 1 : 16] с химическими сдвигами для C1—C6; в качестве примера в табл. 1 приведены все такие матрицы для глюкозы*.

Чтобы рассчитать химические сдвиги для каждого из замещенных моносахаридных остатков, берут определенную строку из матрицы с химическими сдвигами и прибавляют к химическому сдвигу C1 соответствующую величину из матрицы с эффектами гликозилирования. Химический сдвиг сигнала C6 (карбоксильной группы) остатков уоновых кислот в расчете не участвует.

Данные экспериментального спектра полисахарида вводятся без химических сдвигов неуглеводных заместителей (таких, как N-ацильные группы аминсахаров и заместители карбоксильных групп уоновых кислот), так, чтобы число химических сдвигов совпадало с их числом в рассчитанных спектрах ($6n$, где n — число моносахаридных остатков в повторяющемся звене).

Количественная оценка сходства рассчитанного для каждой из теоретически возможных структур и экспериментального спектров проводится так же, как в работе [3]: вычисляют отнесенную к числу моносахаридных остатков сумму квадратов попарных разностей химических сдвигов, упорядоченных по убыванию их величин ($S = \sum (\Delta\delta)^2/n$). В качестве кандидатов на реальную структуру рассматриваются все структуры, характеризующиеся величиной $S \leq 1,5$.

Программа SCAN может быть также использована для расчета ^{13}C -ЯМР-спектров полисахаридов заданной структуры. Кроме того, эта программа позволяет создавать на магнитном накопителе и обеспечивает доступ к банку данных химических сдвигов ^{13}C углеводов, содержащему, в частности, данные экспериментальных спектров всех исследованных расчетным методом полисахаридов, а также, возможно, других углеводов.

Программа SCAN была использована для анализа ряда регулярных линейных полисахаридов с установленным нами моносахаридным составом ([3, 6—16] и неопубликованные данные), являющихся поверхностными антигенами различных бактерий (табл. 2).

Для 14 исследованных полисахаридов расчет в каждом случае выявил только одну структуру с относительно небольшой величиной $S = 0,2—0,6$ (примеры 1—14). Для всех остальных теоретически возможных структур в примерах 1—14 величина S превышала 1,5, и, следовательно, они не удовлетворяли экспериментальному ^{13}C -ЯМР-спектру. Таким образом, структуры полисахаридов этой группы установлены с применением программы SCAN в качестве единственного метода анализа. В эту группу попали все исследованные полисахариды с дисахаридным повторяющимся звеном, более половины полисахаридов с трисахаридным повторяющимся звеном и 2 из 10 полисахаридов с тетрасахаридным повторяющимся звеном.

Для каждого из полисахаридов второй группы (примеры 15—18) расчет привел по меньшей мере к двум структурам, более или менее удовлетворяющим экспериментальному ^{13}C -ЯМР-спектру ($S \leq 1,5$). Эти структуры различались положением замещения по крайней мере одного моносахаридного остатка, и только одна из них находилась в соответствии с данными проведенного нами метилирования. Таким образом, чтобы установить строение полисахаридов этой группы, достаточно было применить программу SCAN в сочетании с анализом методом метилирования. Отме-

* Данные для других моносахаридов — маннозы, галактозы, рамнозы, фукозы, глюкуроновой и галактуроновой кислот, N-ацелированных глюкозамина, маннозамина, галактозамина, хиновозамина и фукозамина — могут быть получены у авторов по специальному запросу.

**Результаты структурного анализа бактериальных полисахаридов
по программе SCAN**

Номер	Структура * и источник выделения	S
1	-4)- α -L-Rha-(1 \rightarrow 3)- β -D-Man-(1- <i>Pseudomonas gladioli</i> pv. <i>alliiicola</i>	0,2
2	-3)- α -L-Rha-(1 \rightarrow 4)- β -D-Glc-(1- <i>Acetobacter m. thanolicus</i> MB58	0,3
3	-3)- β -D-GlcA-(1 \rightarrow 4)- β -D-Glc-(1- <i>Corynebacterium s.p. donicum</i>	0,3
4	-3)- β -D-GlcNAc-(1 \rightarrow 4)- β -D-GlcA-(1- <i>Pasteurella multocida</i>	0,5
5	-4)- α -D-Glc-(1 \rightarrow 3)- α -L-GlcNAc-(1- <i>Pseudomonas cepacia</i> 01	0,6
6	-4)- α -L-Rha-(1 \rightarrow 3)- β -D-ManNAc-(1- <i>Pseudomonas aeruginosa</i> X (классификация Мейгерта)	0,3
7	-6)- α -D-GlcNAc-(1 \rightarrow 4)- α -D-GalNAc-(1- <i>Pseudomonas solanacearum</i> ICMP 4157	0,6
8	-4)- α -D-Rha-(1 \rightarrow 3)- α -L-Rha-(1 \rightarrow 3)- β -D-Rha-(1- <i>Pseudomonas syringae</i> pv. <i>lachrymans</i>	0,3
9	-3)- β -D-Gal-(1 \rightarrow 4)- β -D-Glc-(1 \rightarrow 3)- β -D-GalNAc-(1- <i>Hafnia alvei</i> 039 **	0,3
10	-3)- α -L-FucNAc(1 \rightarrow 3)- β -D-GlcNAc-(1 \rightarrow 2)- β -D-Gal-(1- <i>Salmonella arizonae</i> 059, <i>Citrobacter</i> 035	0,2
11	-3)- α -L-FucNAc-(1 \rightarrow 3)- β -D-FucNAc-(1 \rightarrow 2)- β -D-Glc-(1- <i>Pseudomonas aeruginosa</i> 07 (классификация Лани)	0,2
12	-4)- α -D-GalNAc-(1 \rightarrow 3)- β -D-GalNAc-(1 \rightarrow 3)- β -D-Gal-(1 \rightarrow \rightarrow 4)- α -D-Glc-(1- <i>Salmonella arizonae</i> 063 **	0,2
13	-2)- β -D-Man-(1 \rightarrow 3)- α -D-ManNAc-(1 \rightarrow 3)- β -L-Rha-(1 \rightarrow \rightarrow 4)- α -D-GlcNAc-(1- <i>Shigella dysenteriae</i> тип 10	0,3
14	-2)- β -D-Glc-(1 \rightarrow 6)- α -D-GlcNAc-(1 \rightarrow 3)- α -L-QuiNAc-(1 \rightarrow \rightarrow 3)- β -D-GlcNAc-(1- <i>Protzus vulgaris</i> OX19	0,9
15	-3)- α -D-Rha-(1 \rightarrow 2)- β -D-Rha-(1 \rightarrow 3)- α -D-Rha-(1- -2)- α -D-Rha-(1 \rightarrow 2)- β -D-Rha-(1 \rightarrow 3)- α -D-Rha-(1- <i>Pseudomonas cepacia</i> 3181	0,3 1,1
16	-3)- α -D-Gal-(1 \rightarrow 3)- α -D-Gal-(1 \rightarrow 3)- α -L-Rha-(1- -2)- α -D-Gal-(1 \rightarrow 3)- α -D-Gal-(1 \rightarrow 3)- α -L-Rha-(1- <i>Azotobacter beijerinckii</i>	0,1 0,5
17	-3)- α -D-Gal-(1 \rightarrow 4)- α -D-GalNAc-(1 \rightarrow 3)- α -L-FucNAc-(1 \rightarrow \rightarrow 3)- β -D-GlcNAc-(1- -4)- α -D-Gal-(1 \rightarrow 4)- α -D-GalNAc-(1 \rightarrow 3)- α -L-FucNAc-(1 \rightarrow \rightarrow 3)- β -D-GlcNAc-(1- -2)- α -D-Gal-(1 \rightarrow 4)- α -D-GalNAc-(1 \rightarrow 3)- α -L-FucNAc-(1 \rightarrow \rightarrow 3)- β -D-GlcNAc-(1- <i>Protzus vulgaris</i> 019 (<i>Protzus hauseri</i> 019)	0,6 0,9 1,0
18	-2)- α -L-Rha-(1 \rightarrow 3)- α -L-Rha-(1 \rightarrow 3)- α -L-Rha-(1 \rightarrow \rightarrow 3)- α -L-Rha-(1- -2)- α -L-Rha-(1 \rightarrow 2)- α -L-Rha-(1 \rightarrow 3)- α -L-Rha-(1 \rightarrow \rightarrow 3)- α -L-Rha-(1- -2)- α -L-Rha-(1 \rightarrow 3)- α -L-Rha-(1 \rightarrow 2)- α -L-Rha-(1 \rightarrow \rightarrow 3)- α -L-Rha-(1- <i>Vibrio fluvialis</i> **	0,3 0,6 0,6
19	-6)- α -D-Glc-(1 \rightarrow 2)- α -D-Gal-(1 \rightarrow 6)- α -D-Gal-(1- -6)- α -D-Glc-(1 \rightarrow 6)- α -D-Gal-(1 \rightarrow 2)- α -D-Gal-(1- <i>Acetobacter m. thanolicus</i> MB5	0,1 0,1
20	-3)- β -D-Glc-(1 \rightarrow 2)- α -D-GlcA-(1 \rightarrow 3)- α -D-Man-(1 \rightarrow \rightarrow 3)- α -D-Glc-(1- -3)- β -D-Glc-(1 \rightarrow 2)- α -D-GlcA-(1 \rightarrow 3)- α -D-Glc-(1 \rightarrow \rightarrow 3)- α -D-Man-(1- <i>Klebsiella ozaenae</i> K4 **	0,6 0,8

Номер	Структура * и источник выделения	S
21	-2)- α -L-Rha-(1 \rightarrow 2)- α -L-Rha-(1 \rightarrow 3)- α -L-Rha-(1 \rightarrow 3)- α -D-GlcNAc-(1-	0,7
	-2)- α -L-Rha-(1 \rightarrow 3)- α -L-Rha-(1 \rightarrow 2)- α -L-Rha-(1 \rightarrow 3)- α -D-GlcNAc-(1-	1,3
22	<i>Pseudomonas solanacearum</i> ICMP 5879	
	-4)- α -D-Man-(1 \rightarrow 2)- α -D-Man-(1 \rightarrow 2)- β -D-Man-(1 \rightarrow 3)- α -D-GalNAc-(1-	1,0
23	-2)- α -D-Man-(1 \rightarrow 4)- α -D-Man-(1 \rightarrow 2)- β -D-Man-(1 \rightarrow 3)- α -D-GalNAc-(1-	1,3
	<i>Citrobacter</i> 023	
23	-3)- β -D-Gal-(1 \rightarrow 3)- β -D-GlcNAc-(1 \rightarrow 3)- α -D-Gal-(1 \rightarrow 6)- β -D-Glc-(1-	0,6
	-3)- β -D-Gal-(1 \rightarrow 3)- α -D-Gal-(1 \rightarrow 6)- β -D-Glc-(1 \rightarrow 3)- β -D-GlcNAc-(1-	0,6
24	<i>Proteus mirabilis</i> 03	
	-2)- β -D-Glc-(1 \rightarrow 4)- β -D-Glc-(1 \rightarrow 3)- β -D-GlcNAc-(1 \rightarrow 4)- α -D-GalA-(1-	4,8
	<i>Proteus penneri</i> 42	

* В тех случаях, когда анализ по программе SCAN выявил более одной структуры, реальная структура полисахарида приведена первой.

** Исследована структура основной цепи разветвленного полисахарида.

тим, что для первой группы полисахаридов (примеры 1–14) наличие данных метилирования служит дополнительным подтверждением их линейного характера, а указывая положения замещения моносахаридных остатков, можно достичь определенной экономии времени компьютерного анализа.

Примеры 19–23 показывают, что в отдельных случаях более одной структуры могут достаточно хорошо удовлетворять экспериментальному ^{13}C -ЯМР-спектру и одновременно соответствовать данным метилирования. Эти структуры различаются последовательностью остатков различных моносахаридов (примеры 20 и 23) или одинаковых моносахаридов с различным положением замещения (примеры 19, 21 и 22). В таких случаях нахождение реальной структуры требует применения дополнительных методов анализа: химических (избирательное расщепление гликозидных связей) или, при достаточно хорошо разрешенном ^1H -ЯМР-спектре, ЯМР-спектроскопических (определение ядерных эффектов Оверхаузера, корреляционная гетероядерная $^{13}\text{C}/^1\text{H}$ -спектроскопия). Но так как выбор реальной структуры проводится из очень узкого круга структур (в примерах 19–23 из двух структур), можно не проводить полного установления строения олигосахаридных фрагментов, полученных при избирательном расщеплении, а ограничиться определением их моносахаридного состава (например, состава альдобиоуроновой кислоты $\text{GlcA} \rightarrow \text{Man}$, полученной парциальным гидролизом полисахарида из *Klebsiella ozaenae* K4 [7], пример 19) или идентификацией терминального невосстанавливающего моносахаридного остатка (галактозы в олигозиде $\text{Gal} \rightarrow \text{GlcNAc} \rightarrow \text{Gal} \rightarrow \text{Gro}$ — продукте распада по Смитту полисахарида из *Proteus mirabilis* O3 [8], пример 23).

Особый случай составляет полисахарид *Proteus penneri* 42, исследованный в примере 24. Для него расчет не выявил ни одной структуры с величиной $S < 3,5$. Строение этого полисахарида было установлено независимым методом [17], и было найдено, что его реальная структура

характеризуется величиной $S=4,8$. Сравнение данных рассчитанных и экспериментального спектров показало, что наибольший вклад в величину S вносят отклонения химических сдвигов остатка глюкозы, замещенного в положение 2, и соседнего, гликозилируемого им, остатка глюкозы.

Аналогичные отклонения были отмечены ранее также для трисахаридных фрагментов полисахаридов из *Streptococcus pneumoniae* 18A [5] и *Salmonella arizonae* 045 [18], имеющих структуры α -D-Galp-(1→2)- β -D-Glcp-(1→4)- α -D-GlcpNAc и α -L-Fucp-(1→2)- β -D-Galp-(1→3)- β -D-GlcpNAc соответственно.

Как видно, во всех этих трисахаридных фрагментах терминальный моносахаридный остаток присоединен 1,2-гликозидной связью, и можно предположить, что наблюдающиеся для них отклонения химических сдвигов ^{13}C от аддитивности имеют такую же конформационную природу, что и в разветвленных трисахаридах с вицинальным замещением моносахаридного остатка, лежащего в узле разветвления [19, 20]. Связь этих отклонений с изменениями межзвеньевых протон-протонных взаимодействий была продемонстрирована экспериментально с помощью ядерных эффектов Оверхаузера для фрагмента полисахарида *S. arizonae* 045 [18]. Таким образом, с точки зрения отклонения химических сдвигов от аддитивности липейные трисахариды, включающие в себя 1,2-гликозидную связь, представляют собой частный случай трисахаридов, имеющих вицинально дизамещенный моносахаридный остаток (формально дизамещенным в положения 1 и 2 является у них центральный моносахаридный остаток).

Эта особенность требует осторожности при применении программы SCAN и аналогичных расчетных методов для анализа углеводов, которые, по данным метилирования, содержат моносахаридные остатки, замещенные в положение 2. В общем случае для включения таких объектов в сферу применимости этих методов необходим переход от концепции дисахаридного приближения [6] к концепции трисахаридного приближения. На практике это должно выразиться во введении в соответствующие матрицы базы данных дополнительных строчек, содержащих химические сдвиги с необходимыми поправками, которые аналогичны поправкам, применяемым при анализе разветвленных углеводов [21], и которые должны быть определены из данных спектров соответствующих модельных трисахаридов.

Таким образом, использование компьютерной программы SCAN в сочетании с методом метилирования — эффективный метод структурного анализа линейных регулярных полисахаридов, позволяющий в большинстве случаев однозначно устанавливать их строение или сводить решение этой проблемы к выбору из двух или нескольких структур. Использование такого подхода дает возможность значительно сократить по сравнению с традиционными приемами исследования углеводов время анализа и расход анализируемого материала. Модификация программы структурного анализа в применении к персональным компьютерам делает ее доступной практически для любой исследовательской лаборатории.

Программа SCAN с незначительными модификациями может быть использована также для структурного анализа линейных олигосахаридов и олигосахаридных цепей олигозидов с произвольным агликоном, включая гликопептиды. Результатам анализа углеводов олигосахаридной природы предполагается посвятить отдельное сообщение.

Экспериментальная часть

Программа SCAN написана на языке Turbo Pascal 5.5 с использованием пакета Turbo Numerical Toolbox для создания меню. Все данные

хранятся и вычисления проводятся в целых числах. Расчеты выполнены на компьютере IBM PC/AT 286 (Tandon).

При вводе данных задаются название файла (не более 8 знаков), комментарий (не более 80 знаков — например, источник выделения полисахарида, фамилия исследователя, моносакхаридный состав), неотнесенные химические сдвиги экспериментального ^{13}C -ЯМР-спектра (с точностью до 0,1 м.д., умноженные на 10), число моносакхаридных остатков в повторяющемся звене (1–4), моносакхаридные компоненты в кодах и, факультативно, положение замещения каждого из моносакхаридных остатков. Название файла, комментарий и экспериментальные химические сдвиги автоматически заносятся в банк данных; если файл с таким названием уже существует, об этом выдается сообщение и предлагается к нему доступ.

При выводе данных сообщаются структуры с наименьшей величиной S , упорядоченные по возрастанию этой величины вплоть до заранее заданного определенного значения (обычно $S_{\max}=5$), а также соответствующие этим структурам значения S . По команде для каждой из этих структур на экран или на печать могут быть выведены рассчитанные химические сдвиги, отнесенные химические сдвиги экспериментального спектра и их попарные разности.

^{13}C -ЯМР-спектры сняты на приборе Bruker AM-300 с рабочей частотой 75,47 МГц в D_2O при 60°C с использованием в качестве внутреннего стандарта метанола (850,1 м.д.).

Авторы благодарят проф. В. Н. Шибаева за плодотворную дискуссию.

СПИСОК ЛИТЕРАТУРЫ

1. Липкинд Г. М., Шашков А. С., Кочетков Н. К. // Биоорган. химия. 1987. Т. 13. № 6. С. 833–841.
2. Jansson P.-E., Kenne L., Widmalm G. // Carbohydr. Res. 1987. V. 168. № 1. P. 67–77.
3. Lipkind G. M., Shashkov A. S., Knirel Y. A., Vinogradov E. V., Kochetkov N. K. // Carbohydr. Res. 1988. V. 175. № 1. P. 59–75.
4. Шибаев В. Н. // Биоорган. химия. 1988. Т. 14. № 7. С. 926–937.
5. Jansson P.-E., Kenne L., Widmalm G. // Carbohydr. Res. 1989. V. 188. P. 169–191.
6. Kochetkov N. K., Shashkov A. S., Lipkind G. M., Knirel Y. A. // Sov. Sci. Rev. B. Chem. 1989. V. 13. Part. 2. P. 1–73.
7. Knirel Y. A., Kocharova N. A., Shashkov A. S., Kochetkov N. K., Mamontova V. A., Solov'eva N. F. // Carbohydr. Res. 1989. V. 188. № 2. P. 145–155.
8. Vinogradov E. V., Kaca V., Shashkov A. S., Krajewska-Pietrasik D., Rózalski A., Knirel Y. A., Kochetkov N. K. // Eur. J. Biochem. 1990. V. 188. № 3. P. 645–651.
9. Книрель Ю. А., Кочарова Н. А., Шашков А. С., Варбанец Л. Д., Кочетков Н. К., Станиславский Е. С., Машилова Г. М. // Биоорган. химия. 1986. Т. 12. № 9. С. 1268–1273.
10. Назаренко Е. Л., Горшкова Р. П., Оводов Ю. С., Шашков А. С., Книрель Ю. А. // Биоорган. химия. 1989. Т. 15. № 8. С. 1100–1106.
11. Варбанец Л. Д., Кочарова Н. А., Книрель Ю. А., Мурац В. А., Москаленко Н. В., Броварская О. С. // Микробиол. ж. 1990. Т. 52. № 4. С. 27–33.
12. Katzenellenbogen E., Romanowska E., Kocharova N. A., Shashkov A. S., Lipkind G. M., Knirel Y. A., Kochetkov N. K. // Carbohydr. Res. 1990. V. 203. № 2. P. 219–227.
13. Кочарова Н. А., Варбанец Л. Д., Шашков А. С., Мурац В. А., Книрель Ю. А. // Биоорган. химия. 1991. Т. 17. № 1. С. 107–110.
14. Shashkov A. S., Vinogradov E. V., Daeva E. D., Knirel Y. A., Zdrovenko G. M., Gubanova N. Y., Yakovleva L. M., Zakharova I. Y. // Carbohydr. Res. 1991. V. 212. № 2. P. 301–305.
15. Vinogradov E. V., Daeva E. D., Shashkov A. S., Knirel Y. A., Zdrovenko G. M., Yakovleva L. M., Gubanova N. Y., Solyanik L. P. // Carbohydr. Res. 1991. V. 212. № 2. P. 313–320.
16. Vinogradov E. V., Kaca W., Rózalski A., Shashkov A. S., Cedzynski M., Knirel Y. A., Kochetkov N. K. // Eur. J. Biochem. 1991. V. 200. № 1. P. 195–201.
17. Vinogradov E. V., Sidorczyk Z., Shashkov A. S., Zych A., Rózalski A., Knirel Y. A., Kochetkov N. K. // Eur. J. Biochem. In press.
18. Vinogradov E. V., Shashkov A. S., Knirel Y. A., Nifant'ev N. E., Kochetkov N. K., Dabrowski J., Stanislawski E. S., Kholodkova E. V. // Carbohydr. Res. In press.

19. Lemieux R. U., Bock K., Delbaere L. T. J., Koto S., Rao V. S. // Can. J. Chem. 1980. V. 58. P. 631-653.
20. Lipkind G. M., Nifant'ev N. E., Shashkov A. S., Kochetkov N. K. // Can. J. Chem. 1990. V. 68. № 7. P. 1238-1250.
21. Lipkind G. M., Shashkov A. S., Kochetkov N. K. // Carbohydr. Res. 1990. V. 198. № 2. P. 399-402.

Поступила в редакцию
18.VI.1991

N. K. KOCHETKOV, E. V. VINOGRADOV, Yu. A. KNIREL, A. S. SHASHKOV,
G. M. LIPKIND

**A PROGRAMME SCAN FOR THE ^{13}C NMR-BASED STRUCTURAL
ANALYSIS OF LINEAR POLYSACCHARIDES WITH USING
THE PERSONAL COMPUTER**

*N. D. Zelinsky Institute of Organic Chemistry,
Russian Academy of Sciences, Moscow*

A programme SCAN was elaborated for the ^{13}C NMR - based structural analysis of regular polysaccharides, which represents our earlier programme (Lipkind G. M. et al. // Carbohydr. Res. 1988. V. 175. № 1. P. 59-75) modified for IBM-PC-compatible personal computers. SCAN was successfully applied for the structural elucidation of 24 polysaccharides of bacterial origin. Optimal combinations of the computer-assisted method with other approaches for analysis of carbohydrates and scopes of application of the programme are discussed.