

## БИОИНФОРМАТИКА

УДК 544.165

© Коллектив авторов

### КОМПЬЮТЕРНОЕ КОНСТРУИРОВАНИЕ ПОЛИКЕТИДОВ С ЗАДАНЫМИ СВОЙСТВАМИ

*А.П. Сергейко<sup>1\*</sup>, А.В. Степанчикова<sup>1</sup>, Б.Н. Соболев<sup>1</sup>, С.Б. Зотчев<sup>2</sup>,  
А.А. Лагунин<sup>1</sup>, Д.А. Филимонов<sup>1</sup>, В.В. Поройков<sup>1</sup>*

<sup>1</sup>ГУ НИИ биомедицинской химии им. В.Н. Ореховича РАМН, 119121, Москва,  
ул. Погодинская, д. 10; тел.: (495) 247-30-29, эл.почта: sergeikonp@mail.ru

<sup>2</sup>Норвежский университет наук и технологий, г. Тронхейм, Норвегия

Предложен подход к рациональному конструированию новых поликетидов с заданным спектром биологической активности. Разработана компьютерная программа BioGenPharm, которая на основе заданных пользователем входных параметров генерирует комбинаторные библиотеки поликетидов, осуществляет прогноз спектра биологической активности для сгенерированных структур и отбор среди них молекул с требуемыми свойствами. Для прогноза спектра биологической активности поликетидов использовался алгоритм PASS (<http://www.ibmc.msk.ru/PASS>). Оценка его прогностической способности для предсказания биологической активности поликетидов была выполнена на тестовой выборке, содержащей 242 природных макролида из базы данных Dictionary of Natural Products. Она показала, что средняя точность прогноза составляет 75,5%. Рассмотрена проблема выбора пороговых значений вероятности наличия активности (Pa), обеспечивающих оптимальные сочетания таких параметров как чувствительность, специфичность и согласие, характеризующих качество прогноза. В качестве примера применения описанного метода была сгенерирована виртуальная библиотека аналогов эритромицина и отобраны вещества, для которых снижена вероятность гепатотоксического действия.

**Ключевые слова:** поликетиды, макролиды, виртуальная библиотека, поликетидсинтазы, прогноз биологической активности, отбор веществ с заданными свойствами.

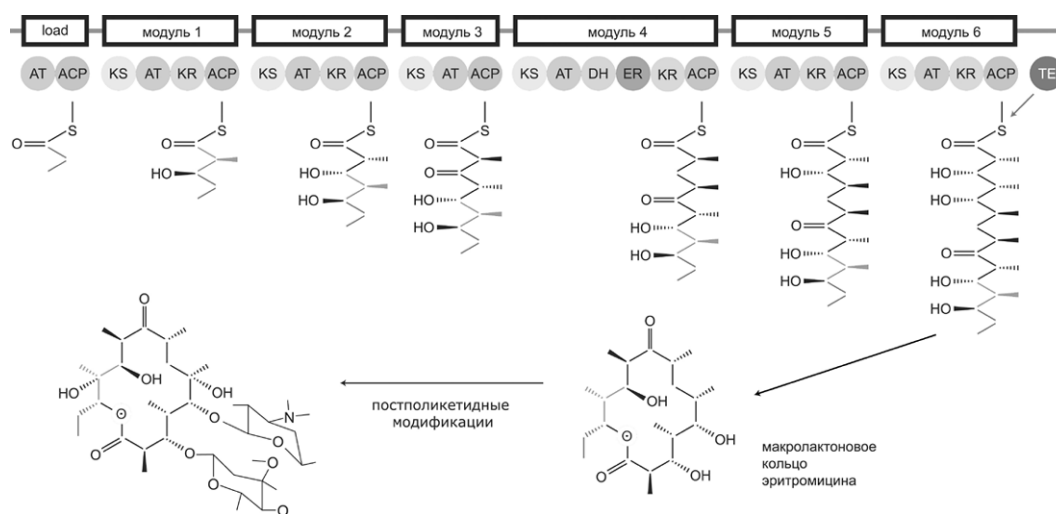
**ВВЕДЕНИЕ.** В последние десятилетия расшифрованы пути биосинтеза микроорганизмами многих биологически значимых соединений, в частности, поликетидов [1]. Поликетиды представляют собой группу разнообразных по структуре и функциям вторичных метаболитов бактерий, грибов, водорослей, губок и некоторых других организмов. Подобно жирным кислотам, поликетиды синтезируются из небольших углеродных фрагментов посредством повторяющихся реакций конденсации [1]. Этот процесс катализируется ферментативными комплексами, которые названы поликетидсинтазами. Известно несколько типов поликетидсинтаз (PKS): PKS I типа (модульные), PKS II типа (итеративные), PKS III типа (халькононовые синтазы высших растений).

Схема работы PKS типа I, предложенная Donadio и др., состоит в том, что PKS организованы повторяющимися модулями, каждый из которых отвечает за один цикл конденсации [2]. Модули состоят из доменов. Минимальный для одного модуля набор включает в себя кетосинтазный (KS), ацилтрансферазный (AT) и ацилпереносящий (ACP) домены. Домен AT осуществляет первую стадию

\* - адресат для переписки

## КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ПЕПТИДОВ

сборки поликетид – перенос активированных “строительных блоков” на АСР домен соответствующего модуля. К домену АСР в результате работы домена АТ присоединяется малонат или метилмалонат. Домен KS катализирует реакцию конденсации, в результате которой молекула поликетид удлинняется на два атома углерода и прикрепляется к АСР домену. Помимо этих основных доменов, модули PKS могут содержать дополнительные домены: кеторедуктазный (KR), дегидратазный (DH), енол-редуктазный (ER) (рис. 1). Наличие или отсутствие этих доменов в модуле определяет степень восстановления той части, которая добавлена предшествующим модулем в поликетидную цепь. Их отсутствие приводит к наличию β-кетогруппы в молекуле поликетид; присутствие KR, KR+DH, KR+DH+ER даёт соответственно гидроксильную группу, двойную связь, насыщенную связь.



**Рисунок 1.**

Строение и механизм работы модульных поликетидсинтаз на примере синтеза эритромицина.

Три гена EryAI-III кодируют три белка PKS: DEBS1 (загрузочный, 1, 2 модули), DEBS2 (3, 4 модули), DEBS3 (5, 6 модули, ТЕ домен). Таким образом, PKS состоит из загрузочного модуля, 6-ти модулей расширения, домена ТЕ. Каждый модуль включает от 3-х до 6-ти доменов: АТ – ацилтрансферазный, АСР – ацилпереносящий, KS – кетосинтазный, KR – кеторедуктазный, DH – дегидратазный, ER – енолредуктазный.

Когда поликетидная цепь полностью синтезирована, она отделяется от PKS с помощью тиоэстеразного домена (Te), затем происходит ее циклизация “хвост к голове” (рис. 1). Образовавшееся макролактоновое кольцо обычно подвергается пост-поликетидным модификациям (гидроксилирование, гликозилирование, ацилирование и др.), в результате чего окончательно определяются структура и биологические функции синтезируемой молекулы.

Поликетиды применяют в медицине в качестве антибиотиков (эритромицин, тетрациклин, кларитромицин и др.), противогрибковых (амфотерицин, нистатин), иммуносупрессантов (FK506), противоопухолевых (доксорубин и другие антрациклины), и ряда других препаратов [3]. В настоящее время известно около 10 000 поликетидов [4]. Нуклеотидные последовательности генных кластеров многих модульных поликетидсинтаз, в частности, отвечающих за синтез эритромицина, нистатина, пимарицина, амфотерицина, кандицидина, полностью определены и разработаны методы генетических манипуляций с

продуцирующими их бактериями [5]. Полностью охарактеризованы также генные кластеры, отвечающие за синтез авермектина [6], спинозина [7, 8], мицинамицина [9], сорафена [10]. Знание строения и механизма функционирования модульных поликетидсинтаз позволяет получать новые вещества путем манипуляций с генами PKS [11, 12]. Этот подход активно развивается в последние годы и получил название “комбинаторный биосинтез”. С помощью методов комбинаторного биосинтеза можно изменять субстратную специфичность АТ домена и число дополнительных (KR, DH, ER) доменов в модуле, удалять или заменять модули, влиять на пост-поликетидные модификации макролактонового кольца [11, 13, 14]. Теоретически можно получить огромное количество новых поликетидов [15], однако экспериментальное получение такого числа соединений невозможно. Предлагаемый нами подход направлен на решение этой проблемы с помощью компьютерных методов конструирования поликетидов с заданными свойствами.

Ранее нами была разработана компьютерная программа Biogenerator [16], которая генерирует виртуальные комбинаторные библиотеки структур поликетидов и последовательности доменов PKS, необходимые для их синтеза. Для сгенерированных структур выполнялся прогноз спектра биологической активности с помощью компьютерной программы PASS, прогнозирующей около 2800 видов биологической активности со средней точностью более 90% [17]. Применимость PASS для прогноза биологической активности макролидов была подтверждена модельным экспериментом: для 285 полусинтетических аналогов эритромицина из базы данных MDDR [18] средняя точность прогноза составила около 96% [19]. Отбор соединений с заданным спектром биологической активности осуществлялся с помощью программы PharmaExpert [20]. Было показано, что последовательное применение упомянутых программ позволяет отбирать среди сгенерированных структур поликетиды с требуемым профилем биологической активности [19].

В то же время возможны ситуации, когда количество сгенерированных структур очень велико. Например, для поликетидов с размером макролактонового кольца 38 атомов углерода, число возможных структур составляет  $2^{54}$ . Очевидно, что при работе с такими большими объемами информации возникают проблемы, связанные с необходимостью ее записи и хранения. С целью оптимизации расчетов мы разработали компьютерную программу BioGenPharm [21], интегрирующую функции программ Biogenerator, PASS и PharmaExpert, которая позволяет генерировать и отбирать виртуальные структуры поликетидов без записи промежуточных результатов. В данной работе представлены описание программы BioGenPharm, валидация используемого алгоритма прогноза биологической активности на тестовой выборке природных макролидов, и иллюстрация применения предложенного подхода на примере анализа виртуальной библиотеки аналогов эритромицина.

**МЕТОДИКА.** Входные параметры программы BioGenPharm можно условно разделить на две группы: параметры генерации и параметры отбора. Параметрами генерации молекул служат: а) тип стартового “строительного блока” (ацетат или пропионат); б) размер макролактонового кольца (от 6 до 38 атомов углерода); в) наличие или отсутствие полукетала в определенной позиции макролактонового кольца; г) позиции пост-поликетидных модификаций и их тип (гликозилирование, гидроксилирование, окисление метильной группы). Для гликозилирования могут быть использованы следующие сахара: дезозамин, микаминоза, микозамин, олиоза, оливоза, даунозамин, олеандроза, микароза, кладиноза. Возможен выбор субстратной специфичности АТ-домена каждого модуля (малонил-специфический, метилмалонил-специфический) и фиксация доменного состава модулей. Для отбора соединений можно задавать следующие параметры: список видов биологической активности, которые будут прогнозироваться; пороговые значения  $P_a$  для каждого вида активности (см. ниже). Наличие или отсутствие тех или иных видов активности в прогнозе характеризует требуемый профиль биологической активности.

Результатом работы программы является файл, который содержит структуры молекул и описания соответствующих им PKS в формате SDfile (Elsevier MDL) [18]. Описание PKS включает порядок следования и доменный состав модулей в гипотетической поликетидсинтазе. В выходной файл записываются только те структуры поликетидов, прогнозируемые спектры биологической активности которых удовлетворяют заданным параметрам.

Отбор молекул основывается на результатах прогноза спектра биологической активности, полученных с помощью алгоритма прогноза, используемого в программе PASS. Прогноз PASS основан на анализе взаимосвязей “структура-активность” с использованием обучающей выборки, содержащей более 68000 разнородных химических соединений с различными видами биологической активности [17]. Спектром биологической активности называется весь комплекс биологических эффектов, которые вещество способно вызывать при взаимодействии с биологическими объектами. Результат прогноза спектра биологической активности представляется в виде упорядоченного списка названий соответствующих активностей и вероятностей  $P_a$  – “быть активным” и  $P_i$  – “быть неактивным”.

Оценка применимости программы PASS для предсказания биологической активности поликетидов проводилась на тестовой выборке из 242 природных макролидов с экспериментально исследованной биологической активностью. Выборка была составлена на основе базы данных Dictionary of Natural Products (DNP) [22]. В тестовую выборку включали вещества, которые относятся к группе “Macrolide polyketides” и “Polyenes” по классификации этой базы данных.

Для оценки точности прогноза был использован инвариантный критерий точности прогноза (IAP):

$$IAP = \frac{N(P_a(s_a) > P_a(s_i))}{Na - Ni}$$

где  $N(P_a(s_a) > P_a(s_i))$  – число случаев, когда  $P_a$  для активных соединений больше, чем  $P_a$  для неактивных соединений;  $Na$  и  $Ni$  – число активных и неактивных соединений в выборке.

Выбор порогового значения  $P_a$  для отдельных видов активности проводился путём оценки согласия, чувствительности, специфичности. Согласие – это отношение числа случаев правильного предсказания к общему числу предсказаний. Чувствительность – доля правильных предсказаний наличия данного вида активности. Специфичность – доля правильных предсказаний отсутствия данного вида активности.

**РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ.** Результаты предсказания биологической активности для тестовой выборки из 242 структур природных макролидов представлены в таблице 1. Таблица содержит данные о количестве соединений в тестовой выборке и обучающей выборке PASS, проявляющих тот или иной вид активности в эксперименте, а также оценку точности прогноза – IAP – для каждой активности. Вычисление IAP для обучающей выборки проводилось в режиме скользящего контроля с исключением по одному. В обучающей выборке PASS содержится 33 структуры из тестовой выборки, однако при выполнении прогноза для каждого из этих соединений соответствующие данные исключались из обучающей выборки.

Из представленных в таблице данных видно, что для обучающей выборки все виды активности прогнозируются с высокой точностью: от 77,5% (антагонист интерлейкина 1) до 99,7% (стимулятор протеинкиназы C). Для тестовой выборки виды активности, для которых известно достаточное количество соединений, прогнозируются с приемлемой точностью: от 63,4% (антибиотик) до 99,8% (противопаразитарная). Исключением является иммуносупрессорная активность. Оценка точности прогноза для выборки, содержащей небольшое количество соединений, не является статистически достоверной и может варьировать в широких пределах. Это подтверждается приведёнными в таблице 1 данными: точность прогноза варьирует от 0,8% (ингибитор синтеза белка) до 100% (ингибитор синтеза холестерина).

Таблица 1. Результаты прогноза экспериментально подтвержденных видов активности для соединений из тестовой выборки.

Название активности	N	IAP	N(PASS)	IAP(PASS)
Антибактериальная	62	0,7456	5570	0,9190
Антибиотик	61	0,6335	3920	0,9569
Антибиотик макролидного ряда	58	0,6963	720	0,9963
Токсическое действие	50	0,8492	1626	0,8182
Цитотоксическая	49	0,8636	293	0,9415
Противогрибковая	43	0,6427	1634	0,8839
Противопаразитарная	36	0,9981	548	0,8939
Противоопухолевая	36	0,9260	8886	0,8440
Противоглистная	15	0,9430	422	0,8966
Иммуносупрессорная	12	0,4844	1081	0,8116
Противовирусная	10	0,8043	3968	0,8767
Инсектицидная	6	0,8482	127	0,9232
Ингибитор синтеза холестерина	4	1,0000	31	0,8264
Противопротозойная	4	0,7269	915	0,8763
Противоглистная (Нематоды)	2	0,9604	56	0,8494
Противомикоплазмозная	2	0,9375	36	0,9252
Акарицид	2	0,9063	39	0,9060
Ингибитор ангиогенеза	2	0,6167	584	0,8598
Ингибитор синтеза РНК	2	0,3458	108	0,8769
Агонист ангиотензина	2	0,3188	108	0,7801
Антиапоптотическая	2	0,2104	721	0,9326
Ингибитор цитохроксидазы	1	1,0000	781	0,9242
Ингибитор ДНК-закисляющей РНК-полимеразы	1	0,9917	301	0,9056
Ингибитор высвобождения цитокинов	1	0,9793	5	0,9773
Стимулятор протеинкиназы C	1	0,9793	17	0,9965
Стимулятор протеинкиназы	1	0,9793	18	0,9792
Антигиперхолестеринемическая	1	0,9627	251	0,8781
Антагонист холинэстеразы В	1	0,9585	131	0,9767
Антагонист интерлейкина	1	0,9585	208	0,8387
Ингибитор АТФазы	1	0,9461	66	0,8083
Антагонист холинэстеразы	1	0,9461	220	0,9778
Антагонист интерлейкина 1	1	0,9129	105	0,7753
Противоопухолевый антибиотик	1	0,9087	719	0,9498
Противопротозойная (Кокцидии)	1	0,8008	104	0,9163
Противовирусная (Герпес)	1	0,7967	111	0,8293
Ингибитор транспортной $H^+K^+$ АТФазы	1	0,7178	175	0,9535
Гипотензивная	1	0,6307	1685	0,9080
Лечение атеросклероза	1	0,6141	766	0,8182
Ингибитор транспортной $Na^+K^+$ АТФазы	1	0,2365	18	0,8917
Ингибитор синтеза ДНК	1	0,1826	282	0,8489
Ингибитор синтеза белка	1	0,0083	361	0,9116
Средняя значимость		0,7553		0,8942

Примечание: N – количество соединений с экспериментально подтвержденной активностью данного вида в тестовой выборке; IAP – инвариантная точность прогноза для тестовой выборки; N (PASS) – количество соединений с экспериментально подтвержденной активностью данного вида в обучающей выборке PASS; IAP (PASS) – инвариантная точность прогноза по скользящему контролю с исключением по одному для обучающей выборки.



Из таблицы 1 следует, что средняя точность прогноза по всем прогнозируемым видам активности для тестовой выборки, состоящей только из макролидов, составляет 75,5%. Средняя точность прогноза для обучающей выборки, включающей соединения разных классов – 89,4%. Таким образом, даже для тестовой выборки, состоящей из близких по структуре соединений, PASS позволяет прогнозировать различия в спектре активности с приемлемой точностью.

Отбор виртуальных молекул макролидов, генерируемых в процессе работы программы BioGenPharm, осуществляется на основе списка прогнозируемых видов активности и пороговых значений  $P_a$  для каждого вида активности. От выбора порогового значения  $P_a$  зависит отнесение виртуального соединения к группе активных или неактивных. По умолчанию, к активным относят соединения, для которых  $P_a > P_i$ , однако не для всех видов активности этот порог является оптимальным. Проблему выбора наилучшего порогового значения  $P_a$  для каждого вида активности можно решить путем анализа результатов прогноза для веществ из тестовой выборки. На основе данных из DNP мы провели оценку чувствительности, специфичности и согласия в зависимости от порога  $P_a$  для нескольких видов активности.

Для антибактериальной активности увеличение порогового значения  $P_a$  от 50% до 100% приводит к повышению точности прогноза отсутствия данного вида активности и уменьшению числа ложноположительных прогнозов (рис. 2а). Специфичность увеличивается от 0,19 ( $P_a = 50,6\%$ ) до 1 ( $P_a = 99,1\%$ ). Одновременно с этим происходит снижение точности предсказания. Чувствительность уменьшается от 0,95 ( $P_a = 50,6\%$ ) до 0,01 ( $P_a = 99,1\%$ ). Принимая во внимание все три графика, можно предположить, что выбор  $P_a = 75\%$  в качестве порогового значения является оптимальным для прогноза данного вида активности. В этом случае, согласие равно 0,67, специфичность – 0,67, чувствительность – 0,66.

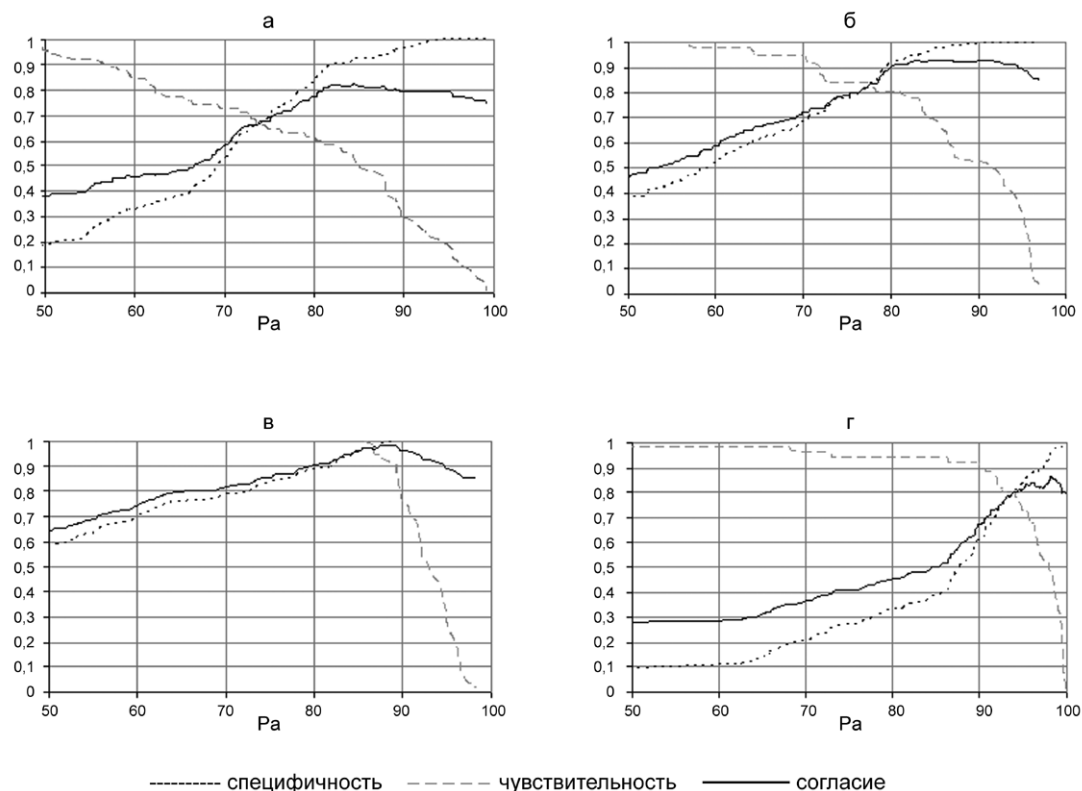
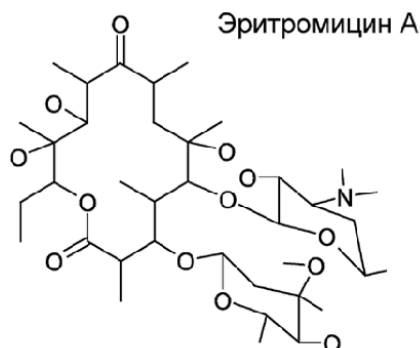


Рисунок 2.

Точность предсказания различных видов активности для тестовой выборки в зависимости от порогового значения  $P_a$ ; а – антибактериальная активность, б – противоопухолевая активность, в – противопаразитарная активность, г – токсическое действие. Согласие – общее число правильных прогнозов/общее число прогнозов, чувствительность – доля правильных предсказаний наличия данного вида активности, специфичность – доля правильных предсказаний отсутствия данного вида активности.

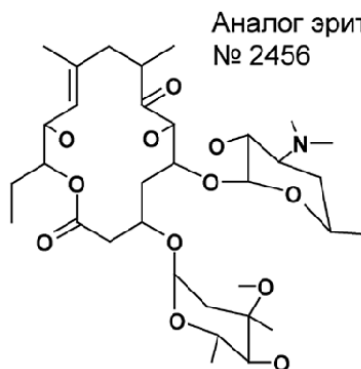


Эритромицин А

[Loading: Propionate]  
 [KS-ATmet-\*\*-\*\*-KR-ACP]  
 [KS-ATmet-\*\*-\*\*-KR-ACP]  
 [KS-ATmet-\*\*-\*\*-ACP]  
 [KS-ATmet-DH-ER-KR-ACP]  
 [KS-ATmet-\*\*-\*\*-KR-ACP]  
 [KS-ATmet-\*\*-\*\*-KR-ACP]

## PASS ACTIVITY SPECTRUM at Pa &gt; 0.8

Pa	Pi	
0.991	0.001	Antibiotic Macrolide-like
0.990	0.003	Cytochrome P450 inhibitor
<b>0.982</b>	<b>0.000</b>	<b>Antibacterial</b>
0.975	0.000	Antibiotic
0.969	0.001	Mucolytic
0.968	0.002	Protein 50S ribosomal subunit inhibitor
0.961	0.000	Transcription factor NF kappa B inhibitor
0.957	0.003	Antiprotozoal (Toxoplasma)
0.955	0.003	Protein synthesis inhibitor
0.951	0.003	HERG channel antagonist
0.943	0.005	QT interval prolongation
0.937	0.000	Antirickettsial
<b>0.941</b>	<b>0.004</b>	<b>Hepatotoxic</b>
0.932	0.001	Expectorant
0.922	0.002	Transcription factor inhibitor
0.870	0.003	Gastrointestinal motility stimulant
0.853	0.002	Prokinetic
0.844	0.014	Toxic
0.819	0.005	Antidyskinetic
0.811	0.001	Tumour necrosis factor agonist
0.810	0.004	Antiparasitic

Аналог эритромицина  
№ 2456

[Loading: Propionate]  
 [KS-ATmal-\*\*-\*\*-KR-ACP]  
 [KS-ATmet-DH-\*\*-KR-ACP]  
 [KS-ATmet-DH-ER-KR-ACP]  
 [KS-ATmal-\*\*-\*\*-ACP]  
 [KS-ATmal-\*\*-\*\*-KR-ACP]  
 [KS-ATmal-\*\*-\*\*-KR-ACP]

## PASS ACTIVITY SPECTRUM at Pa &gt; 0.8

Pa	Pi	
0.904	0.007	Antiprotozoal (Toxoplasma)
0.888	0.001	Transcription factor NF kappa B inhibitor
0.885	0.009	Toxic
<b>0.869</b>	<b>0.001</b>	<b>Antibacterial</b>
0.868	0.004	Transcription factor inhibitor
0.845	0.001	Antibiotic
0.839	0.006	HERG channel antagonist
...		
0.588	0.002	Antibiotic Macrolide-like
...		
<b>0.565</b>	<b>0.109</b>	<b>Hepatotoxic</b>

## Рисунок 3.

Структурные формулы и прогнозируемые PASS виды биологической активности для эритромицина А и его аналога №2456, полученного с помощью программы BioGenPharm. Состав модулей PKS типа I, необходимой для синтеза аналога №2456, получен с помощью программы BioGenPharm. ATmal – малонил-специфический ацилтрансферазный домен,

ATmet – метилмалонил-специфический ацилтрансферазный домен.

Остальные пояснения – см. в тексте.

## КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ПЕПТИДОВ

Таблица 2. Результаты прогноза антибактериальной и гепатотоксической активностей для 17-ти аналогов эритромицина (эр.), отобранных при генерации и скрининге виртуальной библиотеки.

№	$Pa^r$	$Pi^r$	$Pa^a$	$Pi^a$	$Pa^a / Pa^r$
401(эр.)	0,941	0,004	0,982	0,000	1,043571
2454	0,599	0,091	0,905	0,001	1,510851
2452	0,598	0,092	0,890	0,001	1,488294
1734	0,593	0,094	0,845	0,002	1,424958
1566	0,593	0,094	0,845	0,002	1,424958
1732	0,592	0,095	0,830	0,002	1,402027
1564	0,592	0,095	0,804	0,002	1,358108
2310	0,587	0,097	0,870	0,001	1,482112
2472	0,586	0,098	0,875	0,001	1,493174
2308	0,586	0,098	0,851	0,001	1,452218
1840	0,584	0,099	0,826	0,002	1,414384
2360	0,583	0,099	0,873	0,001	1,497427
1648	0,581	0,101	0,849	0,001	1,461274
2464	0,580	0,101	0,882	0,001	1,52069
2456	0,565	0,109	0,869	0,001	1,538053
1568	0,559	0,112	0,810	0,002	1,449016
1736	0,559	0,112	0,810	0,002	1,449016
2312	0,552	0,115	0,830	0,002	1,503623

Примечание: № – номер структуры в виртуальной библиотеке;  $Pa^r$  – вероятность наличия гепатотоксической активности;  $Pi^r$  – вероятность отсутствия гепатотоксической активности;  $Pa^a$  – вероятность наличия антибактериальной активности;  $Pi^a$  – вероятность отсутствия антибактериальной активности;  $Pa^a/Pa^r$  – отношение  $Pa^a$  к  $Pa^r$

Для противоопухолевой активности оптимальное пороговое значение  $Pa = 80\%$ , при этом согласие = 0,90, специфичность = 0,92, чувствительность = 0,81 (рис. 2б).

Для противопаразитарной активности оптимальное пороговое значение  $Pa = 85\%$ , при этом согласие = 0,97, специфичность = 0,96, чувствительность = 1 (рис. 2в).

Для прогноза токсического действия оптимальное пороговое значение  $Pa = 95\%$ , при этом согласие = 0,81, специфичность = 0,84, чувствительность = 0,72 (рис. 2г).

Для иллюстрации применения разработанного метода, мы сгенерировали виртуальную библиотеку аналогов эритромицина и провели ее компьютерный скрининг. Эритромицин сравнительно хорошо переносится пациентами, наиболее серьезным побочным эффектом, наблюдаемым при приеме всех форм эритромицина, является холестаз [23]. Прием эритромицина может приводить к повышению уровня сывороточных трансаминаз и развитию холестатического гепатита. Механизм развития повреждения печени в результате приема эритромицина не ясен: показано, что эритромицин не ингибирует транспорт желчных кислот (таурохолевой кислоты) [24]. Мы провели поиск аналогов эритромицина, для которых вероятность проявления гепатотоксической активности наименьшая.



Генерацию аналогов эритромицина выполняли при следующем наборе параметров: пропионат в качестве стартового блока, 6 модулей расширения, гликозилирование кладинозой по 3-му атому углерода, гликозилирование десозамином по 5-му атому углерода, гидроксильное по 6-му и 12-му атомам углерода. Отбор соединений проводили при значениях  $P_a > 0,8$  для активности “Антибактериальная” и  $P_a < 0,6$  для активности “Гепатотоксическая”. Инвариантная точность прогноза гепатотоксической активности составляет 75,9%, антибактериальной – 91,9%.

В результате было отобрано 17 аналогов эритромицина, удовлетворяющих заданным критериям. В таблице 2 приведены значения  $P_a$  и  $P_i$  для антибактериальной и гепатотоксической активностей эритромицина и его аналогов. Из этой таблицы видно, что наименьшее значение  $P_a$  для гепатотоксической активности при наибольшем значении  $P_a$  для антибактериальной активности прогнозируется для молекулы № 2456. В соответствии с результатами предсказания, вероятность проявления гепатотоксичности для молекулы №2456 составляет 56,5%, что существенно ниже, чем у эритромицина (94,1%). Таким образом, при наличии возможности выбора аналогов эритромицина, наиболее перспективным является аналог №2456 (рис. 3). Необходимая для синтеза этой молекулы информация о доменном составе PKS (рис. 3) получена с помощью программы BioGenPharm и является основой для конструирования соответствующего микроорганизма-продуцента.

**ЗАКЛЮЧЕНИЕ.** В работе представлен новый подход к поиску поликетидов с заданными свойствами. В результате апробации данного подхода на известных из литературы данных о макролидах продемонстрирована приемлемая точность компьютерных предсказаний, что позволяет применять предложенный подход для поиска новых поликетидов, синтезируемых PKS типа I.

Работа выполнена при частичной финансовой поддержке Норвежского исследовательского фонда, SINVENT AS и Biosergen AS (Норвегия), РФФИ (проекты 05-07-90123 и 05-03-08077), МНТЦ (грант №3197).

## ЛИТЕРАТУРА

1. Hopwood D.A., Sherman D.H. (1990) *Annu. Rev. Genet.*, **24**, 37–66.
2. Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L. (1991) *Science*, **252**(5006), 675–679.
3. Omura S. (2002) *Macrolide antibiotics: chemistry, biology, and practice*, 2nd edn. Academic Press: New York.
4. Kosan Biosciences, <http://www.kosan.com/technology-platform.html>.
5. Aparicio J.F., Caffrey P., Gil J.A., Zotchev S.B. (2003) *Appl. Microbiol. Biotechnol.*, **61**, 179–188.
6. Ikeda H., Nonomiya T., Usami M., Ohta T., Omura S. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 9509–9514.
7. Waldron C., Madduri K., Crawford K., Merlo D.J., Treadway P., Broughton M.C., Baltz R. (2000) *Antonie Van Leeuwenhoek*, **78**(3–4), 385–390.
8. Waldron C., Matsushima P., Rostek P.R.Jr., Broughton M.C., Turner J., Madduri K., Crawford K.P., Merlo D.J., Baltz R.H. (2001) *Chem. Biol.*, **8**, 487–499.
9. Anzai Y., Saito N., Tanaka M., Kinoshita K., Koyama Y., Kato F. (2003) *FEMS Microbiol. Lett.*, **218**, 135–141.
10. Schupp T., Toupet C., Cluzel B., Neff S., Hill S., Beck J.J., Ligon J.M. (1995) *J. Bacteriol.*, **177**, 3673–3679.
11. McDaniel R., Thamchaipenet A., Gustafsson C., Fu H., Betlach M., Ashley G. (1999) *Proc. Natl. Acad. Sci. USA*, **96**, 1846–1851.
12. Bentley R., Bennett J.W. (1999) *Annu. Rev. Microbiol.*, **53**, 411–446.

13. *Reeves C.D., Murli S., Ashley G.W., Piagentini M., Hutchinson C.R., McDaniel R.* (2001) *Biochemistry*, **40**, 15464-15470.
14. *McDaniel R., Welch M., Hutchinson C.R.* (2005) *Chem. Rev.*, **105**, 543-558.
15. *Gonzalez-Lergier J., Broadbelt L.J., Hatzimanikatis V.* (2005) *J. Am. Chem. Soc.*, **127**, 9930-9938.
16. *Зотчев С.Б., Соболев Б.Н., Степанчикова А.В., Сергейко А.П., Поройков В.В.* (2007) Свидетельство об официальной регистрации программы для ЭВМ Biogenerator № 2007610742, Москва, Федеральная служба по интеллектуальной собственности, патентам и товарным знакам.
17. *Филимонов Д.А., Поройков В.В.* (2006) *Рос. хим. ж.*, **50**(2), 66-75.
18. <http://www.mdl.com>.
19. *Zotchev S.B., Stepanchikova A.V., Sergeyko A.P., Sobolev B.N., Filimonov D.A., Poroikov V.V.* (2006) *J. Med. Chem.*, **49**, 2077-2087.
20. *Poroikov V., Lagunin A., Filimonov D.* (2004) in: *QSAR and Molecular Modelling in Rational Design of Bioactive Molecules*. (E. Aki Sener and I. Yalcin eds.) CADD & D Society, Ankara (Turkey), pp. 514-515.
21. *Зотчев С.Б., Соболев Б.Н., Степанчикова А.В., Сергейко А.П., Филимонов Д.А., Лагунин А.А., Глориозова Т.А., Поройков В.В.* (2007) Свидетельство об официальной регистрации программы для ЭВМ Biogenpharm № 2007610741, Москва, Федеральная служба по интеллектуальной собственности, патентам и товарным знакам.
22. <http://www.chemnetbase.com/index.html>.
23. *Aronson J.K.* (2006) *Meyler's side effects of drugs*, 15-th ed. Elsevier.
24. *Kostrubsky V.E., Strom S.C., Hanson J., Urda E., Rose K., Burliegh J., Zocharski P., Cai H., Sinclair J.F., Sahi J.* (2003) *Toxicol. Sci.*, **76**, 220-228.

Поступила: 29. 11. 2006.

## COMPUTER-AIDED DESIGN OF POLYKETIDES WITH THE REQUIRED PROPERTIES

*A.P. Sergeyko<sup>1</sup>, A.V. Stepanchikova<sup>1</sup>, B.N. Sobolev<sup>1</sup>, S.B. Zotchev<sup>2</sup>, A.A. Lagunin<sup>1</sup>,  
D.A. Filimonov<sup>1</sup>, V.V. Poroikov<sup>1</sup>*

<sup>1</sup>Institute of Biomedical Chemistry of Russian Academy of Medical Science, Pogodinskaya ul., 10, Moscow, 119121 Russia, Phone: (495) 247-30-29, e-mail: sergeikonp@mail.ru

<sup>2</sup>Norwegian University of Science and Technology, Trondheim, Norway

We propose an approach to rational design of new polyketides with the required spectrum of biological activity. We developed BioGenPharm software for generation of polyketide combinatorial libraries, prediction of activity spectra for the generated structures and selection of molecules with the required properties on the basis of user defined input parameters and selection criteria. For prediction of polyketide activity spectra we used PASS algorithm (<http://www.ibmc.msk.ru/PASS>). Validation of PASS prediction ability for polyketides was performed vs. the evaluation set containing 242 natural macrolides from the Dictionary of Natural Products. The mean prediction accuracy was 75,5%. The problem of choice of cutting points for probability of the presence of activity (Pa), which provide optimal combination of such parameters as sensitivity, specificity, concordance was considered. Applicability of the described method has been illustrated by generation of a virtual library of the erythromycin analogues and selection substances for which the probability of hepatotoxic action is low.

**Key words:** polyketides, macrolides, virtual library, polyketide synthase, biological activity prediction, selection substances with required properties.