

Компьютерное прогнозирование биологической активности природных соединений и их производных

Поройков В.В., Филимонов Д.А., Лагунин А.А., Глориезова Т.А.

*Институт биомедицинской химии им. В.Н. Ореховича РАМН, Москва, Россия
эл. почта: vladimir.poroikov@ibmc.msk.ru*

Новые гетероциклические системы представляют интерес не только как пример фундаментальных достижений в органической химии, но и как ценный источник инновационных фармацевтических продуктов. Более 70% оригинальных фармацевтических субстанций (New Chemical Entities), введенных в медицинскую практику в 1981-2006 г.г., были разработаны на основе природных соединений [1]. Поскольку подавляющее большинство природных соединений проявляет плейотропные биологические эффекты путем взаимодействия с множеством мишеней в организме [2], оценка профилей их биологической активности на ранних стадиях исследований даст возможность сформулировать критерии для отбора наиболее перспективных базовых структур новых лекарств. Для прогноза биологической активности химических соединений в настоящее время широко используются методы, основанные на структуре макромолекулы-мишени и/или на структуре лигандов [3]. Однако большинство современных компьютерных методов сфокусированы на одной-единственной мишени либо на специфическом фармакологическом эффекте и, тем самым, не могут быть использованы для прогноза плейотропного действия веществ.

Компьютерная программа PASS (Prediction of Activity Spectra for Substances) [4] создана на основе идей, возникших около 37 лет тому назад в рамках Государственной системы регистрации новых химических соединений, синтезированных в СССР [5]. Мы начали работу над этим проектом в 1989 году, основываясь на опыте наших предшественников [6, 7]. В течение прошедших лет мы исследовали тысячи различных химических дескрипторов и сотни вариантов алгоритмов с целью разработки единого подхода, обеспечивающего построение достаточно точных и робастных моделей связей «структура-активность» (ССА) на основе анализа обучающих выборок, содержащих многие тысячи органических молекул, принадлежащих к различным химическим классам и проявляющих множество видов биологической активности. Разработчики PASS постоянно проводят работу по сбору и анализу информации о новых фармацевтических субстанциях и биологически активных соединениях для пополнения обучающей выборки и расширения возможностей прогноза PASS на новые химические классы и новые виды биологической активности. Благодаря этому, современная версия PASS (9.1) прогнозирует 3750 видов биологической активности со средней точностью около 95%. С использованием PASS можно было предсказывать 3300 видов активности в 2007 году, 2500 – в 2005 г., 541 – в 1998 г., и только 114 – в 1996 г. Обучающая выборка PASS 9.1 содержит 205873 известных биологически активных

веществ (лекарственных субстанций; фармакологических веществ, изучаемых в клинике и в доклинических тестах; токсических веществ), в то время как в 1996 году обучающая выборка содержала только около 9500 биологически активных соединений. Алгоритм прогноза детально описан в нескольких публикациях [8-11]; поэтому здесь мы представим только краткую информацию, необходимую для практического применения программы PASS.

Биологическая активность описана в PASS качественным образом («активно»/«неактивно»), что позволяет интегрировать в обучающей выборке PASS информацию, собранную из множества различных источников. Структурная формула молекулы, для которой должен быть получен прогноз PASS, представлена в виде MOL файла (для набора молекул – в виде SDF файла, см.: <http://www.symyx.com>). Прогнозируемый спектр биологической активности представлен в PASS в виде списка активностей с двумя вероятностями P_a – «быть активным», и P_i – «быть неактивным», рассчитанными для каждой активности. Этот список упорядочен по убыванию разностей $P_a - P_i$; таким образом, наиболее вероятные активности находятся в верхней части списка. Только активности с $P_a > P_i$ считаются возможными для анализируемого соединения. Список может быть укорочен при любом пороговом значении, однако «по умолчанию» этим порогом является $P_a > P_i$. Если пользователь выберет достаточно высокие значения P_a в качестве порога, шансы подтвердить активность в эксперименте будут довольно высоки, однако многие виды биологической активности могут быть потеряны. Например, если значение $P_a > 90\%$ используется в качестве порога, около 90% имеющихся активностей может быть потеряно; при $P_a > 80\%$, ожидаемая доля потерянных активностей составляет 80%; и т.д. На основе такого рода критериев пользователь может определить, на какие виды биологической должны быть протестированы изучаемые соединения, исходя из компромисса между новизной ожидаемого фармакологического эффекта и риска получения отрицательных результатов при экспериментальном тестировании.

Прогноз PASS был получен для примерно 250000 молекул из открытой базы данных (БД) Национального института рака (НИР) США [12], и результаты были представлены на веб-сайте НИР (<http://cactus.nci.nih.gov/ncidb2/>) в виде, пригодном для поиска. Поисковые запросы формируются путем комбинации различных терминов с использованием Булевских операторов. Например, по запросу «ингибитор ангиогенеза И $P_a > 0.9$ И $P_i < 0.2$ НЕ кислота НЕ амид» было идентифицировано 85 «хитов». Семь соединений было протестировано в НИР и четыре оказались ингибиторами ангиогенеза на уровне значений $IC_{50} = 10-100 \mu\text{M}$ [12]. Кроме того, на основе результатов тестирования анти-ВИЧ действия веществ из этой БД было показано, что использование прогноза PASS позволяет существенно (до 17 раз) повысить долю «активных» соединений в отобранной подвыборке [12].

С 2000 года функционирует сервис PASS INet (<http://www.ibmс.msk.ru/PASS>), который позволяет любому зарегистрированному пользователю получать бесплатно прогноз PASS через Интернет [13, 14]. Структурная формула направляется на прогноз в виде MOL файла, либо вводится непосредственно в Интернет-браузере с использованием Marvin аплета; результат прогноза выдается на дисплей

пользователя. На 1 января 2010 года число зарегистрированных пользователей составило свыше 5500 человек, и более 120000 прогнозов было получено за истекший период. Основываясь на результатах прогноза, исследователи отбирают наиболее перспективные вещества для химического синтеза и биологического тестирования. Сравнение предсказаний PASS с результатами последующего экспериментального тестирования дает возможность провести независимую валидацию данного подхода на соединениях из различных химических классов, проявляющих разные виды биологической активности. В настоящее время опубликовано уже около 30 публикаций, в которых предсказания PASS были подтверждены экспериментом (в качестве обзора см. [15]).

В настоящей работе мы применили PASS к анализу базы данных IBS2009OCT_NC, версия «октябрь 2009 г.» (<http://www.ibscreen.com>), которая содержит 44607 молекул, включая природные соединения (GNC), их производные и аналоги (DNC) и редкие производные (RNC). Входные данные PASS содержали структурные формулы и идентификаторы соединений, представленные в виде файла формата SDF. Выходной файл содержал предсказанные спектры биологической активности для 44540 соединений. Прогноз не был получен для 67 молекул, содержащих две или более компонент. Пример прогноза PASS для молекулы № STOCK1N-00152, относящейся к категории DNC, приведен на рис. 1. При пороге $P_a > 0.5$ предсказываются только три биологических активности, включая: Apoptosis agonist, Phosphatase inhibitor, Antineoplastic (ovarian cancer). Основываясь на результатах прогноза, данная молекула может рассматриваться как противоопухолевое вещество, обладающее достаточно высокой избирательностью действия.

С помощью компьютерной программы PharmaExpert [16] мы проанализировали статистику видов биологической активности, предсказываемых для молекул из базы данных IBS2009OCT_NC. Двадцать наиболее часто прогнозируемых видов биологической активности при пороге $P_a > 0.7$ включают в себя: Gluconate 2-dehydrogenase (acceptor) inhibitor (15233*), Antimetastatic (12408), CYP2C12 substrate (10767), Phosphatase inhibitor (9690), Ubiquinol-cytochrome-c reductase inhibitor (9564), Membrane permeability inhibitor (9151), Transferase stimulant (8972), Membrane integrity agonist (8855), 15-Oxoprostaglandin 13-reductase inhibitor (8831), NOS2 expression inhibitor (8627), Antineoplastic (ovarian cancer) (8538), Neurotrophic factor enhancer (8231), Sialagogue (8158), Nerve growth factor agonist (8128), Antiinflammatory, pancreatic (8070), Depression (8017), Oxidoreductase inhibitor (7878), Hypercholesterolemic (7597), CYP2C19 inducer (7318), Testosterone 17beta-dehydrogenase (NADP+) inhibitor (7264), CYP2A11 substrate (7238). Четыре из этих активностей относятся к фармакотерапевтическим эффектам, 12 – к биохимическим механизмам действия, 3 – к метаболическим терминам, 1 – к побочным или токсическим эффектам.

* В скобках приведено число соединений, для которых прогнозируется соответствующий вид активности.

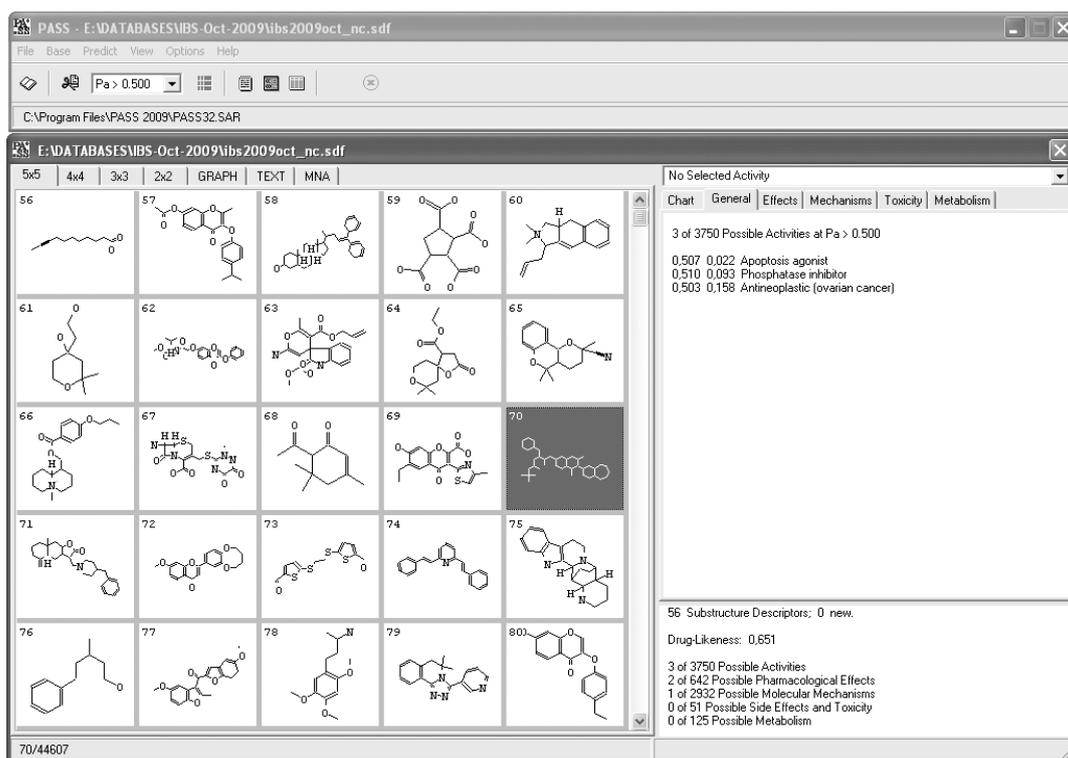


Рис 1. Прогноз PASS для молекулы № STOCK1N-00152.

В соответствии со статистикой прогнозируемых видов активности, при пороге $P_a > 0.5$ 1493 различных биологических активностей прогнозируется не менее чем для 50 соединений, 1783 - не менее чем для 25 соединений, 2140 - не менее чем для 10 соединений, 2364 не менее чем для 5 соединений, 2617 - не менее чем для 2 соединений, 2804 – хотя бы для одного соединения. По меньшей мере для одного соединения прогнозируется 3183 видов активности при пороге $P_a > 0.3$, а при пороге $P_a > P_i$ прогнозируется 3740 видов активности. Таким образом, можно прийти к выводу, что выборка природных соединений и их производных IBS2009OCT_NC является достаточно разнообразной в биологическом пространстве (прогнозируются почти все виды активности, предсказываемые PASS 9.1).

Прогноз PASS рассматривается как достаточно грубая оценка, если в анализируемой структуре имеется более трех новых дескрипторов (по сравнению с соединениями обучающей выборки). Мы проанализировали число таких соединений в базе данных IBS2009OCT_NC set. Оказалось, что 3 соединения содержит по 10 новых дескрипторов, 2 - по 8 новых дескрипторов, 6 - по 7 новых дескрипторов, 16 - по 6 новых дескрипторов, 228 - по 5 новых дескрипторов, 365 - по 4 новых дескриптора. С учетом новизны химической структуры рекомендуется исследовать такие соединения во всех возможных биологических тестах.

Мы оценили новизну трех соединений, содержащих по 10 новых дескрипторов, с использованием сервиса «Chemical Structure Lookup Service» (CSLS, <http://cactus.nci.nih.gov/cgi-bin/lookup/search>). CSLS позволяет проводить структурный поиск среди 74 млн структур из более чем 100 различных баз данных (около 46 млн уникальных структур). Оказалось, что два из трех рассматриваемых соединений (STOCK1N-31841 and STOCK1N-30971) не представлены ни в какой из баз данных, включенных в CSLS, в то время как третье соединение (STOCK1N-30979) также представлено в базах данных ChemNavigator (<http://www.chemnavigator.com>) и ChemSpider (<http://www.chemspider.com>).

С помощью компьютерной программы PharmaExpert [16] мы также идентифицировали соединения, для которых вероятно наличие множественных механизмов противоопухолевого действия (при пороге $P_a > 0.5$ прогнозируется взаимодействие с двумя или более мишенями). Для соединений из базы данных IBS2009OCT_NC set прогнозируется 1384 комбинаций двойных механизмов действия, проявляемых одной или более молекулами. Например, комбинация Histone acetyltransferase inhibitors & Tumor necrosis factor agonists прогнозируется для 43 молекул; Bcl2 antagonists & Phosphatase inhibitors – для 128 молекул; Kinase inhibitors & Membrane integrity antagonists - для 581 molecules; и т.д. Также прогнозируется 5725 различных тройных комбинаций механизмов действия. Например, комбинация Estrogen antagonist & Lipocortins synthesis antagonist & RNA synthesis inhibitor прогнозируется для одной молекулы, Cytidine deaminase inhibitors & DNA directed RNA polymerase inhibitors & Thymidine phosphorylase inhibitors – для 8 молекул, RET inhibitors & Superoxide dismutase inhibitors & Tumor necrosis factor agonists – для 16 молекул, и т.д.

Пример молекулы № STOCK1N-26598, для которой прогнозируются множественные механизмы противоопухолевого действия, приведен на рисунке 2. Как видно из приведенных на рисунке данных, прогнозируемые механизмы действия включают в себя: Tyrosine kinase inhibitor ($P_a=0.70$) & Abl kinase inhibitor ($P_a=0.59$) & Nucleotide metabolism inhibitor ($P_a=0.73$) & RET inhibitor ($P_a=0.71$) & Growth factor antagonist & EGRF kinase inhibitor ($P_a=0.62$) & Cyclic AMP agonist ($P_a=0.53$) & ErbB-1 antagonist ($P_a=0.52$) NOT carcinogen. Как показывает этот пример, с помощью программы PharmaExpert можно отбирать вещества с требуемыми профилями биологической активности среди коммерчески доступных образцов химических соединений. Примеры подтверждения прогноза множественных механизмов действия с использованием PASS в эксперименте приведены в работах [17, 18].

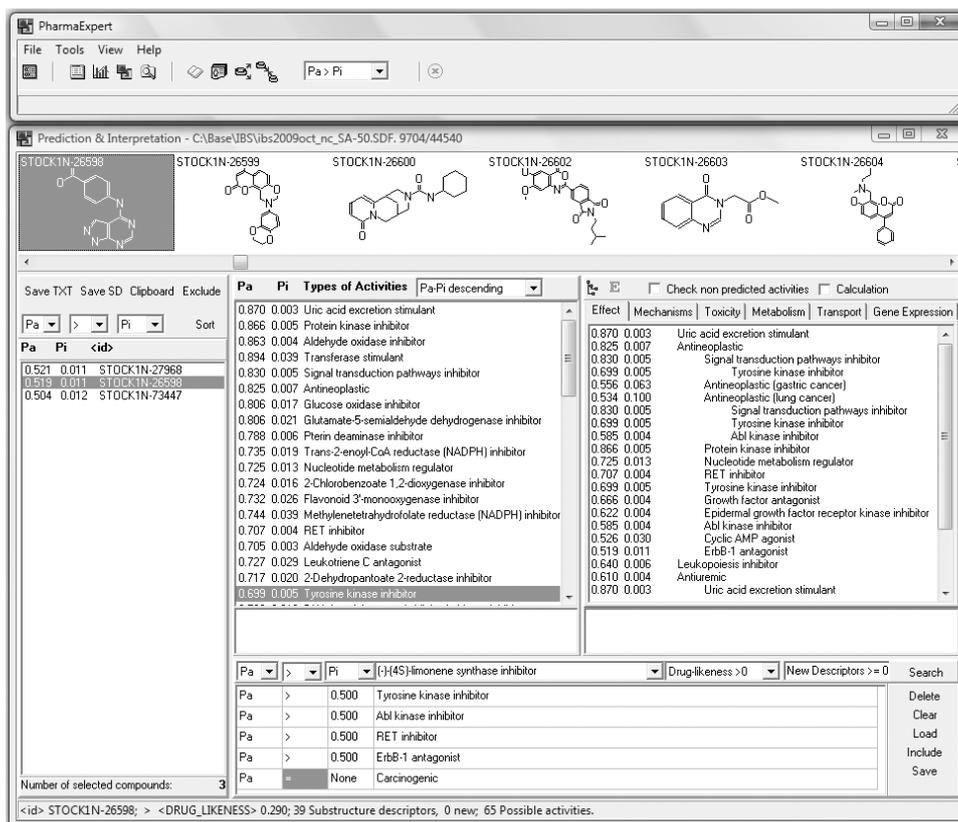


Рис. 2. Поиск молекул с множественными механизмами противоопухолевого действия в базе данных IBS2009OCT_NC.

Заключение

1. Компьютерное предсказание спектров биологической активности с помощью программы PASS для природных соединений и их производных дает возможность отбора молекул, обладающих необходимыми видами биологической активности.
2. Использование компьютерной программы PharmaExpert позволяет отбирать молекулы с требуемыми профилями биологической активности, включая мультитагетные лиганды, а также идентифицировать соединения, содержащие значительное число новых химических дескрипторов.
3. База данных IBS2009OCT_NC, содержащая природные соединения и их производные, является достаточно разнообразной в биологическом пространстве и может быть использована для виртуального скрининга молекул, имеющих самые разные фармакотерапевтические эффекты и молекулярные механизмы действия.

Благодарности. Работа частично поддержана грантами РФФИ 03-07-90282 и 05-07-90123, грантами Европейской комиссии LSHB-СТ-2007-037590 и 200787, грантом МНТЦ 3777.

Литература

1. Newman D.J., Cragg G.M. *J. Natur. Prod.* **2007** 70 (3) 461.
2. Morphy R., Rankovic Z. *Cur. Pharm. Des.* **2009** 15 (6) 586.
3. *The Practice of Medicinal Chemistry*, Wermuth C.G., Ed. New York, Academic Press, 2008, 982 pp.
4. Свидетельство об официальной регистрации программы для ЭВМ PASS № 2006613275 от 15 сентября 2006 г., Москва, Федеральная служба по интеллектуальной собственности, патентам и товарным знакам.
5. Буров Ю.В., Корольченко Л.В., Пороиков В.В. *Бюлл. Всесоюзн. научн. центра безопасн. биолог. активн. веществ.* **1990** (1) 4.
6. Авидон В.В., Аролович В.С., Козлова С.П., Пирузян Л.А. *Хим.-фарм. журн.* **1978** 12 99.
7. Голендер В.Е., Розенблит А.Б. *Логико-комбинаторные методы конструирования лекарств.* Рига: Зинатне, 1983, 352 с.
8. Филимонов Д.А., Пороиков В.В., Караичева Е.И., Казарян Р.К., Будунова А.П., Михайловский Е.М., Рудницких А.В., Гончаренко Л.В., Буров Ю.В. *Эсперим. клин. фармакол.* **1995** 58 (2) 56.
9. Filimonov D.A., Poroikov V.V., in: *Bioactive Compound Design: Possibilities for Industrial Use*, BIOS Scientific Publishers, Oxford (UK), 1996, p.47.
10. Филимонов Д.А., Пороиков В.В. Prediction of biological activity spectra for organic compounds. *Рос. хим. журн.* **2006** 50 (2), 66.
11. Filimonov D.A., Poroikov V.V., in: *Chemoinformatics Approaches to Virtual Screening*. Alexandre Varnek and Alexander Tropsha, Eds. Cambridge (UK): RSC Publishing, 2008, p.182.
12. Poroikov V.V., Filimonov D.A., Ihlenfeldt W.-D., Glorizova T.A., Lagunin A.A., Borodina Yu.V., Stepanchikova A.V., Nicklaus M.C. *J. Chem. Inform. Comput. Sci.* **2003** 43 (1) 228.
13. Lagunin A., Stepanchikova A., Filimonov D., Poroikov V. *Bioinformatics.* **2000** 16 (8) 747.
14. Sady M., Lagunin A., Filimonov D., Poroikov V. *SAR & QSAR Environ. Res.* **2003** 14 (5-6) 339.
15. Geronikaki A., Druzhilovsky D., Zakharov A., Poroikov V. *SAR & QSAR Environ. Res.* **2008** 19 (1-2) 27.
16. Poroikov V., Lagunin A., Filimonov D., in: *QSAR and Molecular Modelling in Rational Design of Bioactive Molecules*. Esin Aki Sener and Ismail Yalcin, Eds., Ankara (Turkey): CADD & D Society, 2005, p. 514.
17. Lagunin A.A., Gomazkov O.A., Filimonov D.A., Gureeva T.A., Dilakyan E.A., Kugaevskaya E.V., Elisseeva Yu.E., Solovyeva N.I., Poroikov V.V. *J. Med. Chem.* **2003** 46 (15) 3326.
18. Geronikaki A.A., Lagunin A.A., Hadjipavlou-Litina D.I., Elefteriou P.T., Filimonov D.A., Poroikov V.V., Alam I., Saxena A.K. *J. Med. Chem.* **2008** 51 (1) 1601.