

Компьютерная система прогнозирования свойств антимикотиков

Т. Б. Чистякова¹, Э. Э. Мусаев²,
Р. В. Макарук³, И. А. Смирнов⁴

Санкт-Петербургский государственный
технологический институт (технический университет)

Санкт-Петербург, Россия

¹chistb@mail.ru, ²eldarofficial@gmail.com,
³makaruk_rv@mail.ru, ⁴igorsmirsapr@mail.ru

В. В. Белахов

Технион — Израильский институт технологии

Хайфа, Израиль

chvalery@technion.ac.il

Аннотация. В статье рассматривается компьютерная система прогнозирования свойств антимикотиков, которая позволяет по химической структуре молекулы данного вещества предсказать его токсичность и полезные свойства, а также подбирать из базы данных регламентов синтеза антимикотиков близкие технологические параметры для химического воспроизведения спрогнозированного вещества. В состав системы входят две нейронные сети (прогнозирование токсичности и полезных свойств с помощью молекулярных дескрипторов и химической структуры в нотации SMILES), база правил получения новых антибиотиков и база регламентов синтеза антимикотиков. Данная система позволяет значительно снизить стоимость разработки новых антибиотиков. Работоспособность разработанной компьютерной системы протестирована на примере противогрибковых полиеновых макролидных антибиотиков.

Ключевые слова: компьютерная система; нейронные сети; база знаний, антимикотики; модификация антибиотиков; регламент синтеза

I. ВВЕДЕНИЕ

В настоящее время наблюдается повышение устойчивости патогенных микроорганизмов к действию антибиотиков, что вызывает необходимость постоянного синтеза новых лекарственных препаратов [1, 2]. Создание новых антибиотиков связано с большими финансовыми затратами, а также с вероятностью их синтеза с высокой токсичностью и/или плохой связываемостью с целевым белком [2, 3].

К патогенам с высокой устойчивостью к антибиотикам относятся и различные грибы, вызывающие такие заболевания как инвазивные микозы, которые в свою очередь являются актуальной проблемой практической медицины [2, 3]. Для лечения грибковых заболеваний применяются прежде всего полиеновые макролидные антибиотики (ПМА) [2-4]. Применяемые в медицине ПМА плохо удовлетворяют возложенным на них задачам из-за их ограниченной эффективности [6]. Ввиду этого коллектив авторов сосредоточен на актуальной задаче поиска новых производных ПМА (ППМА) с улучшенными медико-биологическими свойствами [3, 5].

Применение в медицине информационных технологий, в том числе и систем прогнозирования, стало привычным и привело к их стремительному развитию. Это, в свою очередь, позволило создать общедоступные медицинские базы данных. Используя эту информацию, а также опыт разработчиков системы в синтезе и модификации ППМА [3, 5], нами разработана компьютерная система прогнозирования свойств антимикотиков по их химической структуре, включающая две нейронные сети (прогнозирование токсичности и полезных свойств с помощью молекулярных дескрипторов и химической структуры в нотации SMILES), базу правил получения новых антибиотиков и базу регламентов синтеза антимикотиков [6, 7]. Данная система позволяет снизить время и стоимость синтеза ППМА путём прогнозирования токсичности и полезных свойств по заданной структуре позволяет снизить и осуществить поиск наиболее подходящего лабораторного регламента.

II. РЕШЕНИЕ ПРОБЛЕМЫ

С точки зрения использования любой антимикотик можно описать следующим образом: $Am = \{ LD_{50}, Pm, Se, Rm, Ga, Sol \}$, где LD_{50} - уровень острой токсичности (мг/кг, средняя доза вещества, вызывающая 50%-ю гибель лабораторных животных при изучении токсичности соединений), Pm – белки, против которых направлено действие антимикотика, Rm - виды грибков с известной резистентностью к данному виду противогрибкового препарата, Ga - всасываемость из желудочно-кишечного тракта (при пероральном приёме), Sol – растворимость в жидкостях (при введении с помощью инъекций). LD_{50} и Pm являются одними из важнейших характеристик антимикотика.

Полезные свойства антимикотика, как и токсичность (LD_{50}), можно описать как связь его молекулы с белком, или иной молекулой в грибе или теле человека. Таким образом задачи определения токсичности для человека и противогрибковой активности тесно связана [8, 9].

Проблема синтеза новых ППМА состоит в том, чтобы найти такую его структуру, которая связывалась бы с эргостеролом (элемент Pm в векторе Am), но не с фосфолипидами мембраны клеток человека. Другие

антимикотики связываются с белками грибковой клетки, и таким образом останавливают синтез элементов стенки мембраны [2, 9, 10].

Целью разработанной компьютерной системы является прогноз свойств антибиотика *Am*. Используя молекулярную структуру противогрибкового препарата, разработанная система прогнозирует токсичность (LD_{50}) и полезные свойства (*Pm*) синтезируемого и/или модифицируемого антимикотика. Среда включает в себя следующие модули: прогнозирования токсичности по молекулярным дескрипторам и связи молекул с целевыми белками, для прогнозирования полезных свойств (связь с целевым белком гриба) и дополнительно токсичности (связь с белками организма человека), разработки лабораторных регламентов [6, 8, 11]. Кроме того, система включает в себя базу знаний, позволяющую исследователю осуществлять выбор антимикотики и направление его синтеза в зависимости от химической структуры вещества и его прогнозируемых свойств (общая схема системы описана в [6]). Модуль разработки лабораторных регламентов помогает подобрать наиболее подходящий для синтеза нового ППМА и на его основе разработать новый. Что сократит время перехода к натурному эксперименту с полученным в ходе моделирования в компьютерной системе веществом.

Чтобы правильно зафиксировать все аспекты и методологию синтеза сложных веществ, данные, полученные при проведении опытов и сопутствующие им, необходимо учитывать и хранить в унифицированном виде.

Существует ряд обязательных требований к производству лекарственных средств:

- идентичное, стандартное выполнение всех операций;
- прослеживаемость процесса производства каждой серии продукции;
- обеспечение полного соответствия всех этапов производства, исходных и упаковочных материалов, промежуточной и готовой продукции заданным требованиям.

Производство лекарственных средств должно основываться на системе документации, регламентирующей организацию и технологию производства, контроль качества и документальное оформление всех выполняемых операций. Любое действие должно выполняться в соответствии с документом. Исключаются следование устным распоряжениям и выполнение действий по собственному усмотрению [12].

Лабораторный регламент является первым нормативным технологическим документом, разработкой которого завершаются научные исследования в лабораторных условиях по созданию нового лекарственного препарата. Он устанавливает методы изготовления продукции и условия, обеспечивающие

воспроизводимость технологических процессов в лабораторных условиях со стабильными выходами, а также правила техники безопасности.

Описываемые в лабораторном регламенте параметры проводимых технологических процессов и операций для изготовления нового препарата, способы и методы контроля, мониторинга критических точек, допустимые отклонения, а также сроки и условия хранения препарата определяют в процессе научно-исследовательской работы, которая проводится на основе сравнительных, контролируемых исследований с последующей математической обработкой результатов.

В лабораторном регламенте совмещают требования надлежащей лабораторной практики и надлежащей производственной деятельности, поскольку по этому регламенту в лабораторных условиях готовят экспериментальные серии препарата, предназначенные для доклинических исследований.

В соответствии с [12], основными требованиями к документации являются:

- исчерпывающий характер;
- ясно сформулированные наименование и назначение;
- отсутствие лишней информации и ее необоснованного дублирования;
- ясное изложение текста без неопределенности и двусмысленности;
- логичная структура;
- оформление и утверждение в установленном порядке.

На основании анализа структуры лабораторных регламентов разработано фреймовое описание структуры лабораторного регламента [11].

Подсистема прогнозирования свойств антимикотика использует открытые государственные источники химико-фармацевтических данных о лекарственных препаратах. Полученная из этих баз данных информация приводится к двум используемым в среде форматам, которые подаются на вход нейронным сетям и базе правил. Первый формат, это молекулярные дескрипторы, которые в векторном виде имеют следующий вид: $MdA = \{ Md_1, Md_2, \dots, Md_n \}$, где $n = 99$, по количеству дескрипторов, Md_n – молекулярный дескриптор [6]. Второй формат представляет химическую структуру антибиотика в нотации SMILES (Simplified Molecular Input Line Entry System, система упрощенного представления молекул в строке ввода), приведенный к векторной форме и имеет следующий вид: $S_A = \{ Sm_1, Sm_2, \dots, Sm_n \}$, где $n = 15$ (количество свойств в векторном формате Word2Vec), $Sm_n = \{ Sv_1, Sv_2, \dots, Sv_j \}$ – записи в векторном формате нотации SMILES, где $j = 250$ (максимальная длина записи), Sv_j – координата элемента.

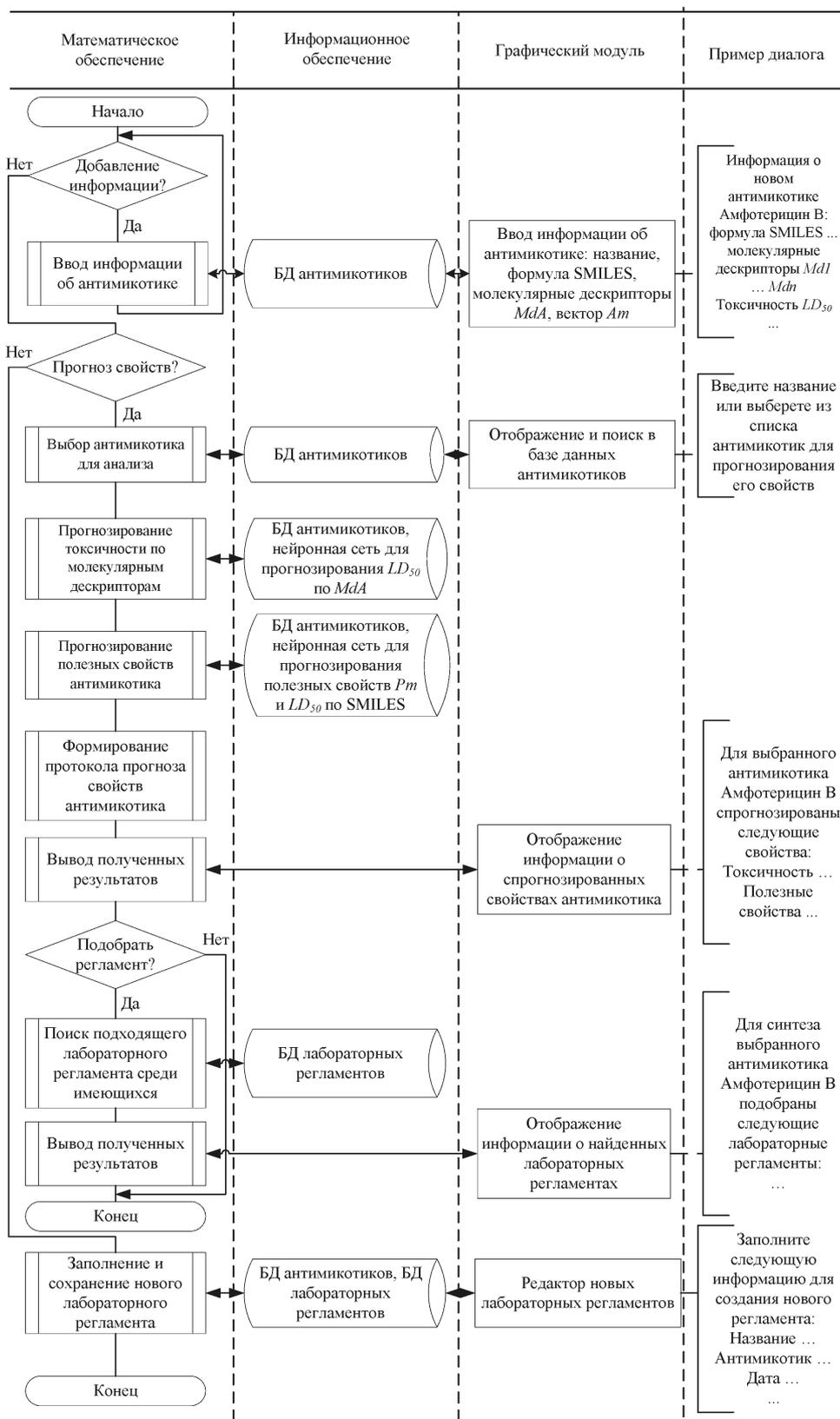


Рис. 1. Алгоритм работы компьютерной системы

Использование формы нотации SMILES [13, 14] вызвано сложностью ввода химических структур молекул в нейронную сеть, а также в связи с тем, что химическая формула в простой форме (например, для противогрибкового антибиотика амфотерицина В, $C_{47}H_{73}NO_{17}$) не предоставляет достаточно информации о связи между атомами, и не является уникальной (множество молекул могут иметь одну и ту же химическую формулу). Ценность канонической нотации SMILES в том, что она предоставляет уникальную, однозначную форму записи.

В среде используются две нейронные сети. Первая прогнозирует токсичность L_A (LD_{50}) по молекулярным дескрипторам Md_A [6]. Вторая прогнозирует связь антибиотика по химической структуре в форме нотации SMILES с белком в: организме человека $Bh_A = \{ Bh_1, Bh_2, \dots, Bh_n \}$, где $n = 21$, по количеству белков в базе TOX21 (<https://ntp.niehs.nih.gov/results/tox21/index.html>), Bh_n – наличие или отсутствие связи с n -ым белком; клетке грибка $Bf_A = \{ Bf_1, Bf_2, \dots, Bf_n \}$, где n – количество белков, повреждение которых убивает грибок, Bf_n – наличие или отсутствие связи с n -ым белком. Вывод обеих нейронных сетей может быть уточнен с помощью базы знаний в области синтеза и модификации антибиотиков.

Для прогнозирования полезных свойств антибиотика (Pm) используется рекуррентная нейронная сеть с нейронами GRU, с существующей архитектурой, которая состоит из одного вводного слоя размером на 50 нейронов, двух рекуррентных слоев размером на 224 и 384 GRU нейрона, и вывод, который для TOX21 состоял из 21 нейрона [8].

Алгоритм работы компьютерной системы прогнозирования свойств антимикотиков представлен на рисунке.

Химические формулы в форме нотации SMILES были получены из базы данных Европейской Лаборатории Молекулярной Биологии (ChEMBL). В качестве исходных данных для обучения и тестирования нейронной сети для прогнозирования полезных свойств антибиотика использовались базы данных ChEMBL (1727081 молекул с известной структурой и свойствами), набор данных TOX21 (10,000 молекул), а также полученные авторами ППМА [3, 5].

III. ЗАКЛЮЧЕНИЕ

Тестирование разработанной компьютерной системы прогнозирования свойств антимикотиков на примере ПМА нистатина ([nystatin, https://www.drugbank.ca/drugs/DB00646](https://www.drugbank.ca/drugs/DB00646)) показало состоятельность такого подхода и возможность его использования для прогнозирования антимикотических свойств и токсичности препарата в результате связи препаратов с белками в клетках гриба и организме

человека, а также подбора лабораторного регламента для синтезируемого вещества.

Разработанная компьютерная система обладает гибкостью и дополняемостью, позволяет обучать разработанные нейронные сети с помощью синтезированных ППМА, добавлять новые антимикотики и лабораторные регламенты, что в результате позволяет снизить стоимость и время синтеза новых ППМА.

СПИСОК ЛИТЕРАТУРЫ

- [1] Yewale V.N. Antimicrobial resistance – A ticking bomb! // Indian Pediatr. 2014. Vol. 51, P. 171-172
- [2] Ya-Lin Chang, Shang-Jie Yu, Joseph Heitman, Melanie Wellington, Ying-Lien Chen. New facets of antifungal therapy // . 2017. Vol. 8, Issue 2, P. 222-236
- [3] Belakhov V., Garabadzhiu A., Chistyakova T. Polyene Macrolide Antibiotic Derivatives: Preparation, Overcoming Drug Resistance, and Prospects for Use in Medical Practice (Review) // Pharmaceutical Chemistry Journal. 2019. Vol. 52, P. 890-901
- [4] Vyas S., Gupta S Optimizing efficacy of amphotericin B through nanomodification // International Journal of Nanomedicine. 2006. Volume 1, Issue 4, P. 417-432
- [5] Belakhov V., Shenin Y. Synthesis and antifungal activity of N-benzyl derivatives of amphotericin B // Pharmaceutical Chemistry Journal. 2007. Vol. 41, P. 362-366
- [6] Chistyakova T.B., Makaruk R.V., Musayev E.E., Belakhov V.V. Computer-Aided Solution for Intellectual Analysis and Judicious Selection of Medically Advanced Antifungals Synthesis Conditions. In Book: Proceedings of XX International Conference on Soft Computing and Measurements. Saint-Petersburg, Publisher: Saint-Petersburg Electrotechnical University. 2017. P. 516-518
- [7] Chistyakova T.B., Makaruk R.V., Musayev E.E., Belakhov V.V. computer system for analyzing and predicting antibiotic properties based on their structure // XXI Mendeleev Congress on General and Applied Chemistry. 2019. Vol. 5, P. 74
- [8] Чистякова Т.Б., Макарук Р.В., Мусаев Э.Э., Белахов В.В. Интеллектуальная среда прогнозирования свойств антибиотиков по их химической структуре // XI Международная конференция по мягким вычислениям и измерениям (SCM-2018). Сб. докладов в 2-х томах. Т. 2. СПб.: 2018. С. 68-71
- [9] Rautenbach M., Troskie A., Vosloo J. Antifungal peptides: To be or not to be membrane active // Biochimie. 2016. Vol. 130, P. 132-145
- [10] Nett J., Andes D. Antifungal Agents: Spectrum of Activity, Pharmacology, and Clinical Indications // Infectious Disease Clinics of North America. 2016. Vol. 30, Issue 1, P. 51-83
- [11] Смирнов И.А., Барбакова В.Е., Чистякова Т.Б., Белахов В.В. Автоматизированная информационная система для разработки лабораторных регламентов по получению производных полиеновых макролидных антибиотиков // Математические методы в технике и технологиях – ММТТ-30: сб. тр. междунар. науч. конф.: в 12 т. Т. 12: в 3 ч. Ч. 2. Санкт-Петербург: Изд.-во Политехн. Ун-та, 2017. С. 51-54
- [12] ГОСТ Р 52550-2006. Производство лекарственных средств. Организационно-технологическая документация. Введ. 2007-01-01. М.: Стандартинформ: Изд-во стандартов, 2006. 51 с.
- [13] Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules // Journal of Chemical Information and Computer Sciences. 1988. Vol. 28, Issue 1, P. 31-36
- [14] Weininger D., Weininger A., Weininger J. Weininger D/Weininger A/Weininger J // Journal of Chemical Information and Computer Sciences. 1989. Vol. 29, Issue 2, P. 97-101