

Биологический модуль электронной базы данных Информационной системы по противоопухолевым веществам

Представлена структура биологического модуля компьютерной базы данных формата ISISBase Информационной системы ГУ РОНЦ им. Н. Н. Блохина РАМН по противоопухолевым веществам, описаны организационные и методические аспекты ее заполнения, редактирования, поиска и использования.

ВВЕДЕНИЕ

Фактографическая электронная База данных (ЭБД) формата ISISBase корпоративной Информационной системы (ИС) ГУ Российский Онкологический научный центр (РОНЦ) им. Н. Н. Блохина РАМН содержит формализованные данные по химической структуре, номенклатурным характеристикам, физико-химическим свойствам и биологической активности около 11 000 веществ, изученных в РОНЦ за период с 1952 г. по настоящее время в качестве потенциальных противоопухолевых лекарств [1, 2].

Данные по биологической активности важная часть информации, содержащейся в описываемой ЭБД. Именно эти данные, полученные в результате трудоемких и достаточно дорогих экспериментов, в совокупности с данными по химическому строению и физико-химическим свойствам изученных веществ, являются не только справочной информацией, но и представляют научную ценность как материалы для аналитических исследований в области экспериментальной химиотерапии рака. Одно из важных направлений таких исследований — изучение связи структуры и биологической активности веществ. Конечная цель его — предсказание биологической активности вещества по химической структуре и конструирование веществ с заданной биологической активностью.

Существенным преимуществом данных по биологической активности ЭБД РОНЦ является то, что в большинстве случаев они получены по унифицированным методикам в стандартах экспериментальных условиях одного и того же института, что повышает их ценность для использования в качестве обучающих массивов компьютерных систем прогнозирования биологической активности по химической структуре вещества. Количественный характер данных также повышает их ценность при использовании в аналитических исследованиях.

Для сравнения, надежные, извлеченные из публикаций, данные обучающего массива наиболее продвинутой отечественной компьютерной системы прогноза биологической активности веществ

PASS, во-первых, получены в разное время, разными методами и в разных учреждениях, во-вторых, имеют только качественный (да/нет) характер [3, 4].

Качество прогнозов с помощью системы PASS, несомненно, будет улучшено при использовании в обучающем массиве данных, полученных в стандартных условиях одного учреждения. Степень унифицированности условий получения данных можно изменять, например, включая в обучающие выборки вещества, изученные только одним исследователем в одной лаборатории, или вещества, изученные в разных учреждениях, но одним и тем же методом.

Использование количественных данных, характеризующих биологические эффекты, при формировании обучающих массивов системы PASS и любой другой компьютерной системы прогноза активности по химической структуре веществ позволяет перейти от качественного прогноза к полуквантитативному и количественному и изменить критерии активности. Например, в обучающую выборку в качестве веществ, обладающих противоопухолевой активностью, можно включать из ЭБД РОНЦ вещества, ингибирующие рост экспериментальных опухолей не на 50% и более, что соответствует принятому в настоящее время критерию активности, а на 70% и более, 80% и более.

Также существенно и то, что биологические данные извлечены в основном из неопубликованных или имеющих ограниченное распространение документов.

Планируемые направления использования ЭБД обусловили как выбор типов данных, характеризующих биологическую активность, так и форму их представления в ЭБД.

Для представления в ЭБД отображены результаты исследования специфической для противораковых лекарств активности, а именно: цитотоксической активности в культурах опухолевых клеток, называемой иногда противоопухолевой активностью *in vitro*, и противоопухолевой активностью в опытах на животных с различными экспериментальными опухолями. Представлены также данные

о токсичности, полученные в опытах на интактных животных, как имеющие большое значение для принятия решения о возможности вещества быть лекарством. Вместе с данными о биологической активности представлены данные об экспериментальных моделях и условиях опыта, использованных при их получении. Кроме количественных данных по специфической активности, информация об активности представлена также в качественном виде.

В ЭБД представлены также в небольшом объеме качественные данные по ряду других видов биологической активности — антибактериальной, противовирусной, антифаговой, фагоиндуцирующей, эстрогенной, андрогенной, антиэстрогенной, антиандрогенной, кортикостероидной. Кроме результатов направленного изучения токсичности, есть данные о токсических дозах и токсических эффектах, наблюдавшихся при изучении противоопухолевой активности вещества на животных с экспериментальными опухолями.

Логическая схема, структура и система Лингвистического обеспечения (ЛО) биологического модуля ЭБД формата ISISBase разработаны по результатам анализа первичных документов с результатами изучения биологической активности потенциальных противоопухолевых препаратов, содержащихся в архиве ИС, а также с учетом возможных направлений использования ЭБД в компьютерных аналитических исследованиях в области экспериментальной химиотерапии опухолей. При этом для информационного анализа предметной области использованы основополагающие публикации по экспериментальной химиотерапии

опухолей, описывающие менявшиеся с течением времени подходы к разработке противоопухолевых лекарственных средств, экспериментальные модели их тестирования, методы и критерии оценки активности веществ [5–12]. Учен также опыт разработки ручной информационно-поисковой фактографической системы на перфокартах [13] и ЭБД форматов DBASE.

ЛОГИЧЕСКАЯ СХЕМА И СТРУКТУРА БИОЛОГИЧЕСКОГО МОДУЛЯ ЭБД

В результате систематизации признаков потенциального противоопухолевого вещества, характеризующих его биологическую активность, разработана иерархическая двухуровневая логическая схема биологических данных, на основе которой затем разработана структура биологического модуля ЭБД.

Выделены три блока данных по биологической активности, соответствующие трем представленным в ЭБД основным видам активности: цитотоксическая активность, противоопухолевая активность, токсичность. Для каждого из этих видов активности в ЭБД представлены оптимальные количественные результаты исследований, адекватно характеризующие уровень специфической активности, характеристики использованной при их получении биологической модели и условия проведения эксперимента. Логическая схема данных по биологической активности представлена в табл. 1.

Иерархическая двухуровневая структура ЭБД представлена в табл. 2.

Таблица 1

Логическая иерархическая схема данных по биологической активности противоопухолевого вещества

0 уровень	Биологическая активность					
1 уровень	Цитотоксическая активность	Противоопухолевая активность	Токсичность	Классы активности	Стадия изучения	Комментарий
2 уровень	Линия клеток Концентрация вещества Размерность концентрации Время экспозиции Параметр оценки цитотоксической активности Торможение параметра ЭК50 (мкг/мл) ЭК50 (мкМ) Исполнитель Ссылка на источник данных	Опухоль Локализация опухоли Животное-хозяин Доза вещества Растворитель Режим лечения Способ введения Увеличение продолжительности жизни Уменьшение продолжительности жизни Торможение роста опухоли Увеличение роста опухоли Излечение Полные ремиссии Частичные ремиссии Стабилизация роста Стимуляция роста Исполнитель Ссылка на источник данных	Вид животных Режим введения Путь введения Категория токсической дозы Значение токсической дозы Комментарий Исполнитель Ссылка на источник данных	Класс активности		

Структура биологического модуля ЭБД

№ п/п	Имя поля	Содержание поля (в соответствии с признаком объекта)	Положение поля в иерархической структуре ЭБД	Тип поля
	Cytotoxicity	Блок признаков цитотоксической активности	1-й уровень (root)	родительское поле
1	<i>cellline</i>	Линия клеток	2-й уровень	текст переменной длины
2	<i>concentration</i>	Концентрация вещества	2-й уровень	текст переменной длины
3	<i>unit</i>	Размерность концентрации	2-й уровень	текст переменной длины
4	<i>time</i>	Время экспозиции	2-й уровень	текст переменной длины
5	<i>parameter</i>	Параметр оценки цитотоксичности	2-й уровень	текст переменной длины
6	<i>inhibition</i>	Торможение параметра	2-й уровень	реальное число
7	<i>IC50 (mcg/ml)</i>	ЭК50 (мкг/мл)	2-й уровень	реальное число
8	<i>IC50 (mM)</i>	ЭК50 (мкг/мл)	2-й уровень	реальное число
9	<i>researcher_cytotox</i>	Исполнитель	2-й уровень	текст переменной длины
10	<i>Reference_cytotox</i>	Ссылка на источник данных	2-й уровень	текст переменной длины
	Antitumor activity	Блок признаков противоопухолевой активности	1-й уровень (root)	родительское поле
11	<i>tumor</i>	Опухоль	2-й уровень	текст переменной длины
12	<i>tumor localization</i>	Локализация опухоли	2-й уровень	текст переменной длины
13	<i>tumor host</i>	Животное хозяйство	2-й уровень	текст переменной длины
14	<i>dose</i>	Доза вещества	2-й уровень	текст переменной длины
15	<i>vehicle</i>	Растворитель	2-й уровень	текст переменной длины
16	<i>schedule_atc</i>	Режим лечения	2-й уровень	текст переменной длины

№ п/п	Имя поля	Содержание поля (в соответствии с признаком объекта)	Положение поля в иерархической структуре ЭБД	Тип поля
17	<i>route_ata</i>	Способ введения	2-й уровень	текст переменной длины
18	<i>ILS</i>	Увеличение продолжительности жизни	2-й уровень	реальное число
19	<i>DLS</i>	Уменьшение продолжительности жизни	2-й уровень	реальное число
20	<i>ITG</i>	Торможение роста опухоли	2-й уровень	реальное число
21	<i>STG</i>	Увеличение роста опухоли	2-й уровень	реальное число
22	<i>curation</i>	Излечение	2-й уровень	реальное число
23	<i>full_remission</i>	Полные ремиссии	2-й уровень	реальное число
24	<i>partial_remission</i>	Частичные ремиссии	2-й уровень	реальное число
25	<i>stabilization</i>	Стабилизация роста	2-й уровень	реальное число
26	<i>growth</i>	Стимуляция роста	2-й уровень	реальное число
27	<i>researcher_ata</i>	Исполнитель	2-й уровень	текст переменной длины
28	<i>reference_ata</i>	Ссылка на источник данных	2-й уровень	текст переменной длины
	Toxicity	Блок признаков токсичности	1-й уровень (root)	родительское поле
29	<i>animal_tox</i>	Вид животных	2-й уровень	текст переменной длины
30	<i>schedule_tox</i>	Режим введения	2-й уровень	текст переменной длины
31	<i>route_tox</i>	Путь введения	2-й уровень	текст переменной длины
32	<i>tox_dose_category</i>	Категория токсической дозы	2-й уровень	текст переменной длины
33	<i>tox_dose_value</i>	Значение токсической дозы	2-й уровень	текст переменной длины
34	<i>comment_tox</i>	Комментарий	2-й уровень	текст переменной длины
35	<i>researcher_tox</i>	Исполнитель	2-й уровень	текст переменной длины

№ п/п	Имя поля	Содержание поля (в соответствии с признаком объекта)	Положение поля в иерархической структуре ЭБД	Тип поля
36	<i>reference_toi</i>	Ссылка на источник данных	2-й уровень	текст переменной длины
	<i>Activity</i>	Классы активности	1-й уровень (root)	родительское поле
37	<i>bioactivity_class</i>	Класс активности	2-й уровень	текст переменной длины
38	<i>researching_phase</i>	Стадия изучения	1-й уровень (root)	текст переменной длины
39	<i>bioactivity_comment</i>	Комментарий	1-й уровень (root)	текст переменной длины

По типу данных основная часть полей — текстовые поля переменной длины. Поля с количественными результатами изучения биологической активности — числовые, данные в них представлены в виде реальных чисел.

Ниже приведено содержание полей биологического модуля.

Поля 1 10 являются дочерними по отношению к полю “**Cytotoxicity**” и содержат данные по результатам изучения цитотоксической активности вещества в культуре опухолевых клеток *in vitro*. Цитотоксический эффект оценивался по влиянию препарата, содержащегося в среде культивирования клеток, на их жизнеспособность, которая фиксировалась по включению меченых предшественников в макромолекулы (преимущественно меченого тимидина в ДНК) или по способности клеток к прижизненной окраске 3-[4,5-диметилтиазол-2-ил]-2,5-дифенилтетразолия бромидом (МТТ-тест).

Поле “*cellline*” содержит название линии клеток, на которых получены представленные в конкретной записи данные по цитотоксичности.

Поле “*concentration*” содержит значение концентрации препарата в среде культивирования клеток, при которой зарегистрирован приводимый в записи максимальный цитотоксический эффект по использованному в данном случае параметру его оценки.

Поле “*unit*” содержит сокращенное название единицы измерения концентрации препарата в среде культивирования.

Поле “*time*” содержит значение времени обработки клеток испытуемым препаратом.

Поле “*parameter*” содержит обозначение параметра оценки цитотоксической активности

Поле “*inhibition*” содержит количественное значение относительного торможения (в процентах) биологического процесса, использованного в качестве параметра оценки цитотоксичности, соответствующее указанному в данной записи концентрации препарата в культуральной среде и времени инкубации клеток в присутствии испытуемого препарата.

Поле “*IC50 (mkg/ml)*” содержит значение измеренной в мкг/мл концентрации испытуемого препарата, вызывающей 50% торможения биологического процесса, использованного в качестве параметра оценки цитотоксичности.

Поле “*IC50 (mkM)*” содержит значение измеренной в мкМ концентрации испытуемого препарата, вызывающей 50% торможения биологического процесса, использованного в качестве параметра оценки цитотоксичности.

Поле “*researcher_cytotox*” содержит фамилию и инициалы исследователя, получившего представленные в данной записи показатели цитотоксичности и несущего ответственность за их достоверность.

Поле “*reference_cytotox*” содержит ссылку на первичный документ (номер лабораторного журнала, публикация или другое), из которого извлечены данные, помещенные в соответствующую запись ЭБД.

Поля 11 28 являются дочерними по отношению к полю 1-го уровня “**Antitumor_activity**” и содержат данные по результатам изучения противоопухолевой активности препарата в опытах на мышах и крысах с лейкозами, асцитными или солидными опухолями, перевиваемыми или индуцированными химическими канцерогенами. Противоопухолевый эффект оценивался по способности изучаемого агента увеличивать продолжительность жизни леченых животных по сравнению с контрольными, тормозить рост перевиваемых опухолей, индуцировать полные и частичные ремиссии или стабилизацию роста опухолей у животных с индуцированными опухолями. Для каждого представленного в ЭБД показателя противоопухолевой активности представлены те характеристики опухолевой модели и условий проведения эксперимента, которые могут влиять на величину противоопухолевого эффекта.

Поле “*tumor*” содержит сокращенное нормализованное название опухолевой модели, на которой изучалась противоопухолевая активность *in vivo*

Поле “*tumor_localiz*” содержит данные о локализации экспериментальной опухоли в организме

животного, которая в случае перевиваемых опухолей совпадает со способом прививки.

Поле *"tumor_host"* содержит название вида животного с указанием генетической линии, на котором изучалась противоопухолевая активность препарата.

Поле *"dose"* содержит значение и единицу измерения дозы, при которой получены значения противоопухолевого эффекта, представленные в соответствующей записи ЭБД.

Поле *"vehicle"* содержит название растворителя, в котором препарат вводился животному при изучении противоопухолевой активности.

Поле *"schedule_atc"* содержит описание режима введения препарата животному при изучении противоопухолевой активности, представленное по форме *"интервал времени между введениями X число введений"*.

Поле *"route_atc"* содержит указание на путь введения препарата животному с опухолью при изучении противоопухолевой активности.

Поле *"ILS"* (increase of life span) содержит среднее значение увеличения продолжительности жизни животных с перевиваемой опухолью под действием препарата (в процентах). Рост опухоли оценивался по увеличению ее объема.

Поле *"DLS"* (decrease of life span) содержит среднее значение уменьшения продолжительности жизни животных с перевиваемой опухолью под действием препарата (в процентах). Наличие данных в этом поле может свидетельствовать о способности вещества стимулировать развитие опухолевого процесса или о проявлении им в испытанной дозе токсических свойств.

Поле *"ITG"* (inhibition of tumor growth) содержит среднее значение торможения роста перевиваемой опухоли под действием препарата (в процентах).

Поле *"STG"* (stimulation of tumor growth) содержит среднее значение стимуляции роста перевиваемой опухоли под действием препарата в процентах. Наличие данных в этом поле может свидетельствовать о способности агента стимулировать развитие опухолевого процесса.

Поле *"curation"* содержит в виде реального числа относительное (в процентах) количество животных в опытной группе, у которых опухоли исчезли после лечения и не рецидивировали в течение срока наблюдения (90 дней).

Следующие четыре поля содержат данные, полученные при тестировании противоопухолевой активности на животных с индуцированными опухолями. В этом случае используются показатели оценки противоопухолевого эффекта, соответствующие показателям оценки эффекта при клинических испытаниях новых противоопухолевых лекарств на людях.

Поле *"full_remission"* содержит в виде реального числа относительное (в процентах) количество животных в опытной группе, у которых после курса введения агента наблюдается уменьшение величины опухоли до 0-2 мм², сохраняющееся на протяжении не менее двух измерений.

Поле *"partial_remission"* содержит в виде реального числа относительное (в процентах) количество животных в опытной группе, у которых после курса введения агента наблюдается уменьшение объема опухоли более чем на 25%, сохраняющееся на протяжении не менее двух измерений.

Поле *"stabilization"* содержит в виде реального числа относительное (в процентах) количество животных в опытной группе, у которых после курса введения агента наблюдается остановка роста опухоли или уменьшение ее объема не более чем на 25%, сохраняющееся на протяжении не менее двух измерений.

Поле *"growth"* содержит в виде реального числа относительное (в процентах) количество животных в опытной группе, у которых после курса введения агента наблюдается увеличение объема опухоли.

Поле *"researcher_atc"* содержит фамилию и инициалы исследователя, получившего представленные в соответствующей записи ЭБД данные по противоопухолевой активности и несущего ответственность за их достоверность.

Поле *"reference_atc"* содержит ссылку на первичный документ (номер лабораторного журнала, публикация или другое), из которого извлечены данные, помещенные в соответствующую запись ЭБД.

Поля 29-36 являются дочерними по отношению к полю 1-го уровня *"Toxicity"* и содержат данные по результатам изучения общей токсичности препарата.

Поле *"animal_tox"* содержит название вида животного, на котором изучалась токсичность препарата.

Поле *"schedule_tox"* содержит описание режима введения препарата животному при изучении токсичности, представленное по форме *"интервал времени между введениями X число введений"*.

Поле *"route_tox"* содержит указание на путь введения, при котором изучалась токсичность препарата.

Поле *"tox_dose_category"* содержит название категории токсической дозы, значение которой представлено в данной записи.

Поле *"tox_dose_value"* содержит в виде текста значение токсической дозы вместе с ее размерностью.

Поле *"comment_tox"* содержит комментарии к данным по токсичности, преимущественно указание на наличие или отсутствие опухоли у животных, на которых получены данные по токсичности, в связи с тем, что токсические дозы препарата могут быть разными у интактных животных и животных с опухолями.

Поле *"researcher_tox"* содержит фамилию и инициалы исследователя, получившего представленные в соответствующей записи ЭБД данные по токсичности и несущего ответственность за их достоверность.

Поле *"reference_tox"* содержит ссылку на первичный документ (номер лабораторного журнала, публикация или другое), из которого извлечены данные, помещенные в соответствующую запись ЭБД.

Поле *"bioactivity_class"* является дочерним по отношению к полю 1-го уровня *"Activity"*. Для каждого агента может быть столько разных записей в поле *"bioactivity_class"*, сколько для него известно различных классов биологической активности. Характеристики биологической активности в этом поле имеют качественный характер. Качественные

характеристики о наличии или отсутствии у препарата цитотоксичности и противоопухолевой активности в большинстве случаев присвоены препарату путем экспертной оценки количественных данных в соответствии с принятыми критериями активности. Препарат квалифицируется как обладающий цитотоксической активностью в случае, если для него показана способность ингибировать выбранный параметр оценки цитотоксичности на 25% или более, и/или если IC_{50} равна или меньше $10^{-4}M$. В противном случае препарат квалифицируется как не обладающий цитотоксической активностью. Препарат квалифицируется как обладающий противоопухолевой активностью в случае, если для него показана способность увеличивать продолжительность жизни животных с опухолью на 25% или более, и/или тормозить рост солидных опухолей на 50% или более. В противном случае препарат квалифицируется как не обладающий противоопухолевой активностью. В некоторых случаях качественное указание на наличие у препарата того или иного класса биологической активности может быть извлечено непосредственно из первичного документа, не содержащего количественных данных по активности.

Поле *“researching_phase”* содержит указание на стадию исследования препарата — первичное тестирование, углубленное экспериментальное изуче-

ние, клинические испытания, использование в клинической практике.

В связи с тем, что одно из планируемых направлений использования ЭБД РОНЦ — формирование обучающих массивов для системы прогноза биологической активности по структуре вещества PASS, качественные характеристики биологической активности в поле *“bioactivity_class”* и название стадии исследования препарата в поле *“researching_phase”* приводятся на английском языке, принятом в этой системе.

Поле *“bioactivity_comment”* содержит дополнительную информацию, имеющую существенное значение для характеристики данных по биологической активности препарата.

ЛИНГВИСТИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ЭБД

Существенная часть ЭБД, обеспечивающая ее полноценное функционирование, — система ЛО [14]. Основная часть системы ЛО биологического модуля ЭБД РОНЦ — вербально-дескрипторные словари, или списки дескрипторов, которые используются при формализации, вводе, редактировании и поиске данных. Список словарей представлен в табл. 3.

Таблица 3

Словари, входящие в состав ЛО ЭБД

№ п/п	Имя текстового файла-словаря	Название словаря	Тип словаря	Имя полей ЭБД, при заполнении которых используется словарь
1	bioactivity_class.txt	Класс биологической активности	свободный	bioactivity_class
2	bioactivity_comment.txt	Комментарий к данным по биологической активности	свободный	bioactivity_comment
3	cellline.txt	Линия клеток	свободный	cellline
4	parameter.txt	Параметр	свободный	parameter
5	reference_bioact.txt	Ссылка на источник данных по биологической активности	свободный	reference_cytotox reference_ata reference_tox
6	researcher_ata_tox.txt	Исследователь, получивший данные по противоопухолевой активности или токсичности	свободный	researcher_ata researcher_tox
7	researcher_cytotox.txt	Исследователь, получивший данные по цитотоксичности	свободный	researcher_cytotox
8	researching_phase.txt	Стадия изучения	свободный	researching_phase
9	route.txt	Путь введения	свободный	route_ata
10	schedule.txt	Схема введения	свободный	schedule_ata

№ п/п	Имя текстового файла-словаря	Название словаря	Тип словаря	Имя полей ЭБД, при заполнении которых используется словарь
11	time_expos.txt	Время экспозиции	свободный	time
12	tox_animal.txt	Животное — объект изучения токсичности	свободный	animal_tox
13	tox_comment.txt	Комментарий к данным по токсичности	свободный	comment_tox
14	tox_dose_category.txt	Категория токсической дозы	свободный	dose_category_tox
15	tumor.txt	Опухолевая модель	свободный	tumor
16	tumor_host.txt	Животное — носитель опухоли	свободный	tumor_host
17	tumor_localiz.txt	Локализация опухоли	свободный	tumor_localiz
18	unit.txt	Единица измерения	свободный	unit
19	vehicle.txt	Растворитель при биологических испытаниях	свободный	vehicle

Большинство словарей составлено на основании терминов, встречающихся в первичных документах первого контура ИС. Для первичных документов, содержащих результаты исследования биологической активности препаратов, характерно наличие различных синонимов и разных написаний одних и тех же названий клеточных линий, опухолей, линий животных. Дескрипторные словари позволяют при формализации данных выбрать из ряда синонимов унифицированный вариант термина. Входящие в словари дескрипторы отобраны на основе анализа предметной области «Экспериментальная химиотерапия опухолей». При выборе унифицированных вариантов терминов в соответствии с принятыми нормами, кроме основополагающих публикаций на русском языке по поиску и разработке противоопухолевых лекарств [6, 7], использован Список ключевых слов Реферативного журнала «Онкология» ВИНТИ [15]. При составлении словаря «Класс биологической активности» использован список названий классов биологической активности на английском языке, используемый в системе PASS [16]. Для максимального сокращения трудоемкости формализации данных и заполнения электронной ЭБД в качестве дескрипторов использованы, по возможности, краткие эквиваленты.

Все словари биологического модуля ЭБД имеют свободный тип. Несмотря на важность соблюдения единообразия в написании терминов для обеспечения полноты и точности информационного поиска, часто возникает необходимость добавления в словарь новых дескрипторов. В словарях фиксированного типа любое изменение словаря было бы возможно только при изменении структуры ЭБД, что замедляло бы работу. Словари в процессе ведения ЭБД изменяются, исключаются устаревшие термины, добавляются новые.

Основное назначение словарей — обеспечение адекватного информационного поиска. Кроме того,

использование словарей позволяет ускорить ввод данных и уменьшить количество ошибок при вводе. Так, дескрипторы словаря «Ссылка на источник данных по биологической активности» фактически представляют собой шаблоны с названиями типов первичных документов, в которые необходимо лишь подставлять номера и/или годы создания документов, из которых выбраны данные.

ОСОБЕННОСТИ ВВОДА, РЕДАКТИРОВАНИЯ И ПОИСКА ДАННЫХ БИОЛОГИЧЕСКОГО МОДУЛЯ ЭБД

Важным моментом при заполнении и редактировании полей биологического модуля ЭБД является использование адекватных предметных форматов, содержащих все поля ЭБД, и интерфейсов пользователя, или экранных форм.

При подготовке предметных форматов используются специально разработанные инструкции по формализации данных, в которых для каждого поля описано содержание, приведены дескрипторные словари, правила использования дескрипторных словарей и представления данных.

Разработано несколько экранных форм с полями биологического модуля. Все они содержат ячейки для ввода данных, однозначно идентифицирующих препарат (единый идентификационный номер, некоторые шифры, сведения о конфиденциальности), и одну или несколько табличных частей, удобных для ввода и просмотра данных полей двухуровневой иерархической ЭБД. Единая форма для ввода и редактирования данных всех полей биологического модуля позволяет одновременно работать со всеми полями.

Однако создание нескольких (в ряде случаев более 10) записей для одного и того же поля 2-го уровня, полное заполнение данными большинства полей

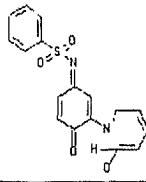
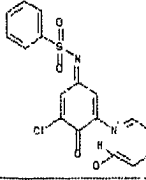
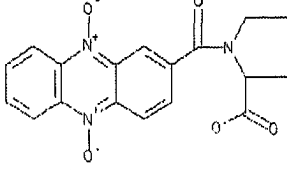
File Edit Options Object Database Search List Window Help				
Forms Query Browse Update				
<Root> 514 of 3866 Search Domain: All				
№№ записи	IDcomp	авторский шифр	структурная формула	класс активности
1136	1969062	ДХТИ-2		Not antineoplastic
1137	1969063	ДХТИ-3		Antineoplastic
1139	1969065	133/23		Not antineoplastic

Рис. 1. Экранная форма, позволяющая одновременно просматривать структурные формулы вещества и данные по наличию у него той или иной активности

File Edit Options Object Database Search List Window Help									
Forms Query Browse Update									
<Root> 1 of 7 Search Domain: All									
IDcomp	биол. тип	биол. вид лат.	биол. вид русск.	хар-ка препарата	линия клеток	время экспозиции	параметр оценки цитотоксичности	IC50 мкМ	класс активности
1993023	растение	Bunias orientalis L.	свербига восточная	пентаноловый экстракт	CaOv	24 час	ДНК/Т	50.0000	Cytostatic
1993024	растение	Bunias orientalis L.	свербига восточная	хлороформенный экстракт	CaOv	24 час	ДНК/Т	100.0000	Not cytostatic
1993025	растение	Bunias orientalis L.	свербига восточная	этилацетатный экстракт	CaOv	24 час	ДНК/Т	100.0000	Not cytostatic
1993026	растение	Bunias orientalis L.	свербига восточная	бутаноловый экстракт	CaOv	24 час	ДНК/Т	200.0000	Not cytostatic
1993028	растение	Comarum palustre L.	сабельник болотный	пентаноловый экстракт	CaOv	24 час	ДНК/Т	50.0000	Cytostatic
1993029	растение	Comarum palustre L.	сабельник болотный	экстракт СС14	CaOv	24 час	ДНК/Т	25.0000	Cytostatic
1993033	растение	Comarum palustre L.	сабельник болотный	этилацетатный экстракт	CaOv	24 час	ДНК/Т	20.0000	Cytostatic

Рис. 2. Экранная форма для просмотра результатов поиска природных экстрактов с результатами изучения цитотоксической активности

обуславливают необходимость использования мелкого шрифта и прокрутки экрана для визуализации всех данных, что затрудняет их комплексное восприятие. Поэтому созданы отдельные экранные формы для работы с полями данных по цитотоксической

активности, противоопухолевой активности и токсичности. Это позволяет более рационально использовать экранное пространство и облегчает работу. Использование той или иной экранной формы определяется конкретной задачей.

Эти же экранные формы могут использоваться при поиске данных биологического модуля.

Система дает возможность одновременной визуализации полученной по одному запросу фактографической информации о биологической активности веществ и их химических формул. При просмотре результатов поиска по биологическим данным удобно использовать табличные формы с ячейками для структурных формул или таксономических данных об источнике выделения препаратов природного происхождения. В этом случае исследователь имеет возможность одновременного просмотра и анализа химических, номенклатурных и биологических данных для конкретного агента. Примеры таких экранных форм приведены на рис. 1 и 2.

Существенным положительным свойством созданной ЭБД является возможность сочетания в одном запросе данных по активности препаратов, структурных формул веществ и номенклатурных характеристик и физико-химических свойств. Числовой тип полей с данными по цитотоксической и противоопухолевой активности позволяет готовить выборки веществ с определенным уровнем активности.

ЗАКЛЮЧЕНИЕ

Описанные структура и ЛО ЭБД формата ISISBase позволяют одновременно с данными по химическому строению вещества хранить формализованные качественные и количественные данные по нескольким видам биологической активности, полученные в стандартизованных экспериментальных условиях. Возможность одновременно экспорта полученных в результате информационного поиска структурных формул и данных по биологической активности в файл коммуникативного формата *.sdf* позволяет создавать разнообразные обучающие массивы для компьютерных систем прогноза активности веществ по их структуре.

Данные по биологической активности позволяют анализировать содержание веществ с той или иной биологической активностью в различных растениях и живых организмах, способность синтетических веществ тех или иных химических классов проявлять ту или иную биологическую активность, оценивать адекватность экспериментальных моделей тестируемой активности.

СПИСОК ЛИТЕРАТУРЫ

1. Апрышко Г. Н. Информационная система по противоопухолевым агентам // Российский биотерапевтический журнал. — 2002. — № 2. — С. 7–10.
2. Апрышко Г. Н. Информационная система РОНЦ им. Н. Н. Блохина РАМН по противоопухолевым агентам // НТИ. Сер. 2. — 2007. — № 1. — С. 18–22.
3. Филимонов Д. А., Поройков В. В. Прогноз спектра биологической активности органических соединений // Российский Химический Журнал. — 2006. — Т. 50. — № 2. — С. 66–75.
4. Poroykov V., Filimonov D. PASS: Prediction of Biological Activity Spectra for Substances // Predictive Toxicology // Ed. by C. Helma. — N. Y.: Marcel Dekker, 2005. — P. 459–478.
5. Ларионов Л. Ф. Химиотерапия злокачественных опухолей. — М.: Медгиз, 1962. — 464 с.
6. Система создания противоопухолевых препаратов в СССР и США / Под ред. Н. Н. Блохина и Ч. Г. Зуброда. — М.: Медицина, 1977. — 352 с.
7. Экспериментальная оценка противоопухолевых препаратов в СССР и США / Под ред. З. П. Софьиной, А. Б. Сыркина, А. Голдина, А. Кляйна. — М.: Медицина, 1980. — 179 с.
8. Anticancer drug development guide: Preclinical screening, clinical trials, and approval / Ed. V. A. Teicher. — Humana Press, 1997. — 311 p.
9. Михайлова Л. М., Сыркин А. Б., Гарин А. М., Барышников А. Ю. Методические указания по изучению общетоксического действия противоопухолевых фармакологических веществ. Руководство по экспериментальному (доклиническому) изучению новых фармакологических веществ. — М.: ИИА “Ремедиум”, 2000. — С. 90–104.
10. Трешалина Е. М., Жукова О. С., Герасимова Г. К., Гарин А. М., Смирнова А. С. Методические указания по изучению противоопухолевой активности фармакологических средств. Руководство по экспериментальному (доклиническому) изучению новых фармакологических веществ. — М.: ИИА “Ремедиум”, 2000. — С. 319–325.
11. Смирнова З. С., Герасимова Г. К., Гарин А. М., Соколова А. С. Методические указания по изучению специфической активности фармакологических веществ, предлагаемых для лечения гормонозависимых опухолей. Руководство по экспериментальному (доклиническому) изучению новых фармакологических веществ. — М.: ИИА “Ремедиум”, 2000. — С. 326–334.
12. Герасимова Г. К., Апрышко Г. Н. Развитие систем и методов отбора веществ с противоопухолевыми свойствами в РОНЦ им. Н. Н. Блохина РАМН // Экспериментальная онкология на рубеже веков / Под ред. М. И. Давыдова, А. Ю. Барышникова. — М.: Издательская группа РОНЦ им. Н. Н. Блохина РАМН, 2003. — С. 59–84.
13. Апрышко Г. Н. Система “Слежение” банка данных “Химиотерапия опухолей” НИИ ЭДитО ВОНЦ АМН СССР // Химиотерапия опухолей в СССР. 1987. — Вып. 47. — С. 217–232.
14. Апрышко Г. Н., Решетникова В. В. Лингвистическое обеспечение электронной базы данных РОНЦ РАМН по противоопухолевым препаратам. М., 2006. — 60 с. — Деп. в ВИНТИ РАН 19.07.06, № 976-В2006.
15. Список ключевых слов к базе данных 28 “онкология (терапия опухолей)”, 29 “онкология экспериментальная”, 30 “онкология клиническая”. М.: ВИНТИ, 1988. — С. 72.
16. <http://1195.178.207.233/PASS>

Материал принят к опубликованию 29.12.06