

ИСПОЛЬЗОВАНИЕ АССОЦИАТИВНОГО АНАЛИЗА ДЛЯ ОБРАБОТКИ НАУЧНЫХ ПУБЛИКАЦИЙ В ОБЛАСТИ СИСТЕМ ДОСТАВКИ ЛЕКАРСТВ

О.А. Угольников, ассистент, *Ю.А. Демич, научный сотрудник,
А.В. Лисица, заместитель директора по научной работе, *В.Ю. Кистанова,
ведущий специалист, *В.Ф. Корнюшко, заведующий кафедрой, **А.И. Арчаков,
директор, В.И. Швеиц, заведующий кафедрой
кафедра Биотехнологии и бионанотехнологии, МИТХТ им. М.В. Ломоносова
*Научно-исследовательский институт биомедицинской химии им. В.Н.Ореховича РАМН
**кафедра Информационных технологий, МИТХТ им. М.В. Ломоносова
***Научно-исследовательская компания «Куб»
e-mail: ovman@yandex.ru

В данной работе был использован ассоциативный анализ для обработки более 600 тыс. научных статей, посвященных системам доставки лекарств. Ассоциативную взаимосвязь между двумя объектами, которыми являлись наименования лекарств, устанавливали в том случае, если они совместно встречались в реферате статьи. Совокупность взаимосвязей, полученных для 2358 фармацевтически активных соединений, отображали в виде сетевой диаграммы (семантической сети). В составе диаграммы были выявлены кластеры веществ, характеризующиеся общностью химической структуры или сходными биологическими свойствами. Показано, что в составе доминирующего по количеству узлов кластера преобладают вещества, используемые для адресной доставки лекарственной субстанции при помощи какой-либо транспортной системы.

In this paper the associative analysis of terms' co-occurrences in the abstracts of scientific articles (more than 600 thousands publications) in the field of drug delivery system was used. The associations between 2358 biologically active chemicals were depicted as a network diagram (semantic network). In the diagram clusters were found, which included compounds similar either in chemical structure or in biological properties. Most of the members of a certain cluster were investigated in relation to the problems of drug delivery.

Ключевые слова: доставка лекарств, семантический анализ, PubChem, PubMed.

Key words: drug delivery, semantic analysis, PubChem, PubMed

Введение

Разработка систем доставки лекарств является актуальной и быстро развивающейся областью биотехнологии. В основе данного направления лежит комплекс биотехнологических подходов, позволяющих встраивать молекулы лекарственного вещества в состав носителя [1]. В результате встраивания изменяется фармакокинетика и фармакодинамика препарата. При этом происходит повышение эффективности действия и снижение побочных эффектов лекарственного вещества как за счет увеличения биодоступности, так и за счет избирательной концентрации вещества в пораженных тканях и органах [2].

Высокий уровень активности научно-исследовательских коллективов, работающих в области создания систем транспорта лекарств, отражается в значительном количестве опубликованных работ. В биомедицинской электронной библиотеке PubMed за последние 20 лет размещено более 600 тыс. публикаций, непосредственно описывающих создание и испытание систем транспорта, и более 1.5 млн. публикаций, имеющих косвенное отношение к разработке систем транспорта и посвященных биологическим свойствам лекарственных компози-

ций. Такой объем информации невозможно систематизировать без привлечения автоматических средств интеллектуальной обработки данных.

В данной работе обработка опубликованной информации проводится с использованием ассоциативного анализа [3]. Суть метода заключается в выявлении совместного упоминания терминов в текстах рефератов статей. Этот метод был успешно применен для решения задач, связанных с выявлением взаимодействующих белков [3], установлением связи генов и белков с заболеваниями [4], в создании автоматической классификации веществ природного происхождения [5]. Опираясь на указанные примеры, можно полагать, что ассоциативный анализ окажется эффективным при систематизации опубликованных данных с целью создания базы знаний, содержащей сведения об экспериментах по включению различных химических соединений в системы доставки лекарств.

Материалы и методы исследования

В качестве источника публикаций использовали доступную в сети Интернет электронную библиотеку PubMed/MEDLINE [6], в которой на 01 октября 2009 года содержалось более 19 млн. рефератов научных статей в области биомедицины и биотехнологии. Авто-

матическое извлечение информации из ресурсов сети Интернет и ее обработку проводили с использованием программного пакета «Biblio-Engine» (ООО «Куб», Москва). Поисковыми средствами сервера PubMed отобрали 646 тыс. рефератов, соответствующих тематике доставки лекарств. Публикацию считали соответствующей тематике, если в ней встречался хотя бы один термин из сформированного экспертом перечня ключевых слов.

Выборку из 6500 наименований химических соединений формировали согласно разделу «PharmacologicalAction» контролируемого словаря терминов MeSH (Medical Subjects Headings), загруженного с сайта Национального центра биотехнологической информации США [7]. С использованием сервера PubChem [8] для каждого химического соединения были получены регистрационные номера CID и варианты синонимичных названий и аббревиатур. Контекстным поиском из 6500 химических соединений отобрали 2358 веществ, наименования или номера CID которых встречались в ранее созданной выборке рефератов по тематике доставки лекарств.

По методике, предложенной в работе [9], для каждого (из отображенных 2358) химического соединения проводили библиографическое картирование [10], то есть поиском PubMed отбирали релевантные публикации, в которых встречается наименование вещества, либо одно из его синонимичных названий, либо регистрационный номер CID. Для каждой релевантной публикации определяли 5 родственных по смыслу документов, используя данные из поля «Related Articles» системы PubMed, и включали их в состав родственного библиографического профиля.

Ассоциативные связи рассчитывали между каждой парой химических соединений. Принимая за k количество ссылок, совпадающих в библиографических профилях родственных публикаций двух соединений, индекс сходства r вычисляли по формуле [11]:

$$r = k / (m + n - k), \quad (1)$$

где m и n – количество статей в библиографических профилях одного и другого соединения. Для построения сетевой диаграммы отбирали только те пары соединений, для которых выполнялось условие: $r > 0.018$.

Для визуализации результатов использовали программу построения сетевых диаграмм GVedit [12]. Программа размещает на диаграмме наименования соединений так, чтобы наиболее оптимальным образом отразить существующие между этими соединениями попарные ассоциативные взаимосвязи.

Результаты и обсуждение

Химические соединения отбирали для работы из контролируемого словаря терминов MeSH, в

раздел «PharmacologicalAction» которого входят вещества, обладающие фармакологическим действием. Семантические взаимосвязи между каждой парой химических соединений устанавливали с использованием рефератов родственных публикаций, родственность между которыми вычисляется сервером PubMed автоматически с использованием встроенного алгоритма частотного анализа [13]. Всего в анализ было включено более 1.5 млн. родственных публикаций, соответствующих 646 тыс. рефератов релевантных статей.

Для каждой пары соединений индекс ассоциативного сходства вычисляли с использованием формулы (1) для нормализации количества совпадающих для двух соединений статей с учетом количества статей, найденных для каждого соединения в отдельности. Это позволяет выровнять значения ассоциативного сходства между хорошо изученными и недавно появившимися соединениями.

При выбранном пороге $r > 0.018$ между 1087 соединениями (из 2358) наблюдали выраженную степень ассоциативной связности. Такие пары отображали в виде ненаправленного циклического графа – сетевой диаграммы. Узлами сетевой диаграммы являются наименования химических соединений, а ребра отражают степень ассоциативной связи между каждой парой соединений.

В составе полученной сетевой диаграммы выявили 18 кластеров (подграфов). Кластерами считали такие подграфы, в которых узлы соединены ребрами, а между узлами из разных подграфов ребер нет.

Сведения о полученных кластерах приведены в табл. 1. Видно, что в составе семантической сети образовался один доминирующий по количеству узлов кластер №1, а также существенно меньшие кластеры, насчитывающие от 7 до 20 узлов. В табл. 1 для каждого кластера, кроме первого, указана принадлежность к одной из терапевтических групп. Во всех кластерах присутствуют выраженные ассоциативные связи, обусловленные сходством биологической активности входящих в них соединений. Например, в кластер №2 вошло 20 соединений, которые образуют 27 ассоциативных связей между веществами транквилизирующего действия. Сходные результаты были получены ранее в работе [5], в которой также было отмечено, что природные вещества распределяются по семантическим кластерам в соответствии со спецификой своего биологического действия.

Кластер №1, содержащий 766 соединений, анализировали более детально. Прежде всего, было отмечено, что в этот кластер вошли разнородные по своей химической структуре и биологическим свойствам соединения. С наибольшей частотой в кластере №1 встречаются

статьи, относящиеся к антибиотикам и соединениям с противоопухолевой или противовоспалительной активностью. Например, на рис. 1 показан центральный фрагмент кластера №1,

где отмечены химиотерапевтические антинеопластические агенты: доксорубин, винкристин, цисплатин и др., а также антибиотики: рифампицин, изониазид, рифабутин.

Таблица 1. Характеристика терапевтических свойств соединений в составе кластеров, выявленных на семантической диаграмме.

№ кластера	Кол-во узлов	Примечание
1	766	
2	20	транквилизирующие средства
3	15	антибактериальные и антипротозойные средства
4	14	гипогликемические средства
5	10	допаминовые средства
6	8	антиконвульсивные и успокаивающие средства
7	8	допаминовые антагонисты
8	8	анаболические средства и андрогены
9	8	никотиновые антагонисты
10	8	адренергические альфа-антагонисты
11	8	витамины группы А
12	7	опиатные алкалоиды
13	7	антинеопластичные средства и ингибиторы ароматазы
14	7	антинеопластичные средства и андрогенные антагонисты
15	7	гормоны гипофиза и их производные
16	7	ингибиторы протеазы
17	7	нейромедиаторные средства
18	7	антиинфекционные средства и вспомогательные вещества

Исходя из рис. 1, можно отметить, что в кластер №1 объединены соединения, для многих из которых выполнены экспериментальные работы по встраиванию в систему транспорта. В частности, липосомальные формы доксорубина, кварцетина, цисплатина, фторурацила, доцетаксела, дофамина, рифампицина, изониазида, рифабутина, даунорубина, амикацина, винкристина, цисплатина, нистатина являются либо коммерческими препаратами, либо проходят испытания различных уровней. Это означает, что наиболее крупный кластер отражает тематическую направленность, согласно которой производили отбор публикаций из библиотеки PubMed.

В составе кластера №1 было проанализировано выборочно 35 соединений. Из них 24 соединения (69%) опубликованы в статьях в связи с повышением эффективности их доставки с использованием систем транспорта. Немногим менее половины соединений (11 наименований) исследовались при встраивании в липосомы. Кроме того, в качестве носителей фигурировали полимерные наночастицы и

мицеллы. В некоторых случаях были представлены сложные нетипичные системы, например, система доставки пропанолаола основана на наночастицах-на-микросферах [14], а система доставки преднизона — на магнитомицеллах [15]. Наряду с системами, являющимися продуктами биотехнологии, в двух случаях доставку осуществляли с использованием естественных механизмов проникновения веществ в организм: за счет осмотического давления [16] и в виде раствора в диметилсульфоксиде (ДМСО) [17].

В табл. 2 приведены сведения по 20 (из 35 проанализированных) соединениям из кластера №1 с высоким индексом ассоциативной связи с веществами, встраиваемыми в систему транспорта. Видно, что значения индекса распределились от 0.037 для бримонида до 0.587 для лопрессора. При этом в таблицу вошли 7 соединений, для которых не удалось установить по соответствующим публикациям, что они были экспериментально проверены в направлении систем доставки лекарственных субстанций. В частности, к таким соединениям отно-

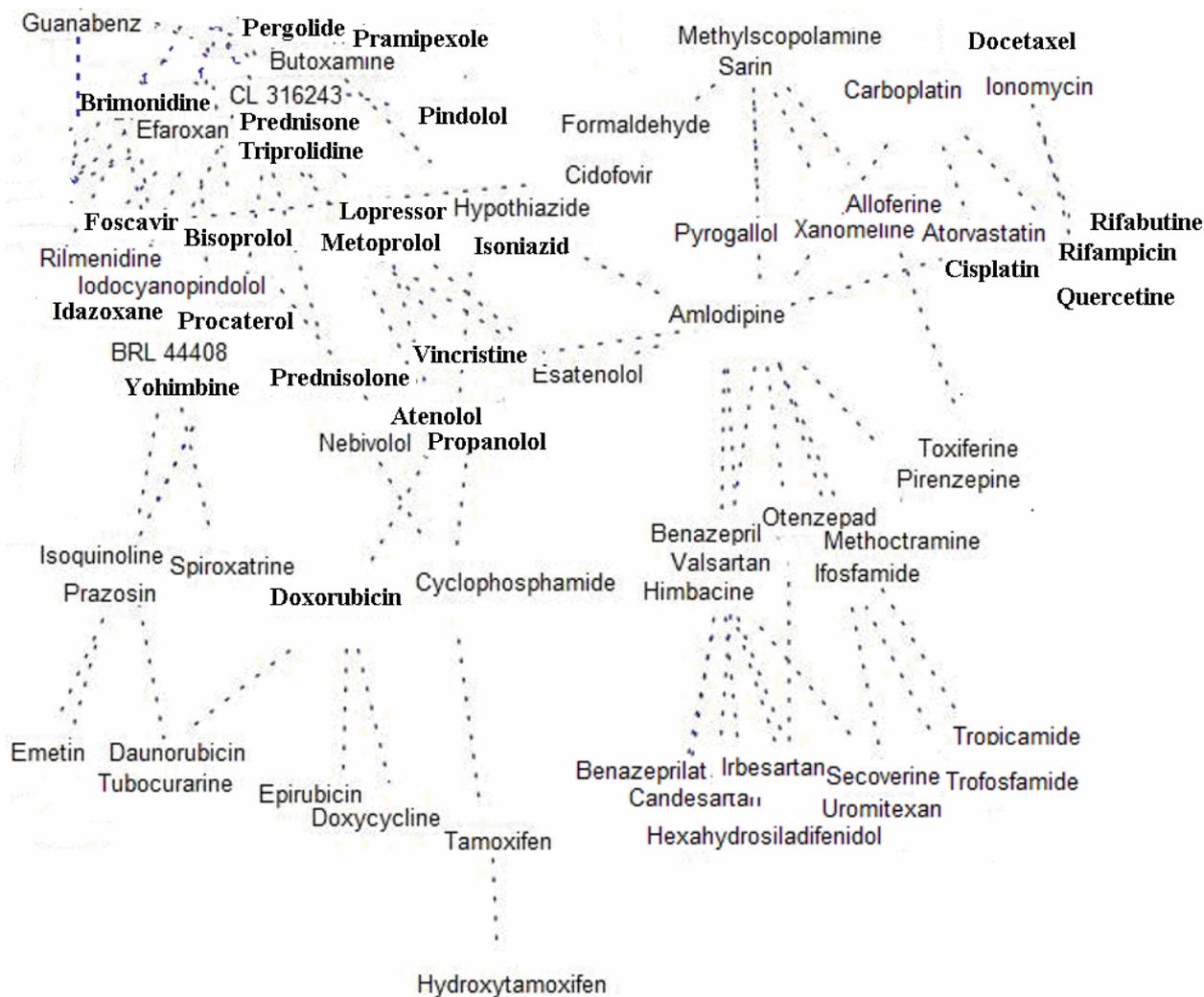


Рис. 1. Центральный фрагмент сетевой диаграммы кластера №1. Пунктиром соединены соединения, индекс ассоциативной связи между которыми превышает 0.018. Полу жирным шрифтом отмечены соединения, исследовавшиеся в составе систем доставки лекарств.

ются BRL-44408 и изохинолин, для которых значение индекса составляет довольно высокую величину – 0.175. Высокое значение индекса ассоциативного сходства для веществ, не использовавшихся в системах транспорта, возможно связано с использованием их в качестве контроля. Например, BRL-44408 является широко используемым в научных целях соединением с выраженным антипсихотическим действием – эталоном, по которому оценивают возросшую эффективность лекарства, снабженного системой доставки.

Таким образом, высокие значения индекса ассоциативного сходства необязательно указывают на то, что соединение испытывалось в составе системы транспорта. Это видно из рис. 2, на котором показано, как изменяется относительное количество (в процентах) соединений, свойства которых были изучены при включении в системы доставки, по отношению к общему количеству соединений, с приблизительно таким же индексом семанти-

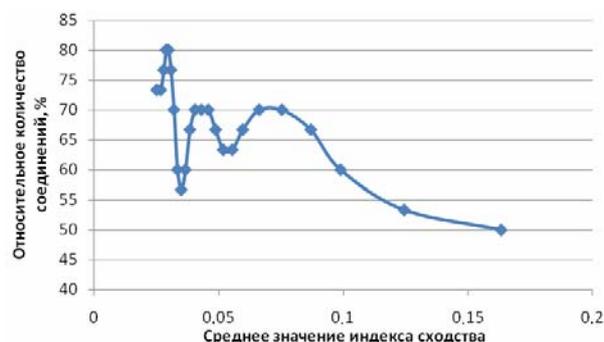


Рис. 2. Зависимость относительного количества соединений, изучавшихся в составе системы доставки лекарств, от среднего значения индекса сходства.

ческого сходства. При низких значениях индекса ассоциативного сходства (от 0.01 до 0.034) наблюдается высокая степень обогащения выборки целевыми соединениями, то есть соединениями, свойства которых анализировались после встраивания в систему транспорта. По мере приближения к высоким значе-

Таблица 2. Системы доставки, применяемые для изменения фармакологических свойств 20 химических соединений, обладающих наибольшим значением ассоциативного сходства.

Химическое соединение	Тип носителя	Ссылка	Релев. статьи	Родств. статьи	Индекс
лопрессор	карбопол 974Р-порошки	[18]	512	2151	0.587
метопролол	липосомы	[14]	512	2151	0.587
BRL 44408	-	-	125	528	0.175
изохинолин	-	-	211	954	0.175
атенолол	липосомы	[14]	134	562	0.166
рифампицин	липосомы	[19]	792	3557	0.149
CL 316243	-	-	116	498	0.095
прамипексол	осмотическая доставка	[16]	172	634	0.085
эфароксан	-	-	147	469	0.07
рилменидин	-	-	160	511	0.07
доксорубицин	липосомы	[20]	1327	4941	0.061
преднизон	магнитомицеллы; полимерные наночастицы	[15]	1046	4506	0.061
преднизолон	магнитомицеллы; полимерные наночастицы; микрочастицы	[15]	1046	4506	0.061
изониазид	липосомы	[21]	1930	7671	0.05
перголид	ПЭГ-везикулы; мицеллы	[22]	344	1303	0.047
кверцетин	липосомы	[23]	532	2184	0.046
йодоциано-пиндолол	-	-	228	1006	0.044
гипотиазид	-	-	1283	5014	0.041
бутоксамин	-	-	102	443	0.038
бримонидин	поликарбоксильные наночастицы	[24]	509	2008	0.037
прокатерол	гелевая система доставки	US Patent 4788064	249	1226	0.034

ниям индекса (0.07 и выше) резко возрастает количество соединений, для которых отсутствуют экспериментальные сведения о встраивании в транспортные системы (т.н. ложноположительные результаты). Наблюдаемая закономерность позволяет предположить, что существуют три области значений семантического сходства. При высоких значениях индекса наблюдается увеличение ложноположительных результатов, при средних – достоверность ассоциативной идентификации лекарственной композиции с системой доставки может достигать 70-80%. Наконец, при низких значениях индекса можно ожидать, что в составе семантической сети будут наблюдаться соединения, для которых целесообразно планировать исследования по оценке эффективности встраивания в систему доставки. Таким образом, полученные результаты позволяют предложить ассоциативный анализ в качестве алгоритмической

основы для создания экспертной системы в области доставки лекарств.

Выводы

Ассоциативные связи между химическими соединениями, установленные путем сопоставления библиографических профилей, отражают распределение в соответствии с фармакологическими группами, а также позволяют выявить сходство по характеру экспериментальных исследований, направленных на улучшение фармакологических свойств биологически активных веществ. Показано, что такое сходство может коррелировать с вероятностью использования веществ в составе системы доставки к тканям и органам, что может использоваться для создания экспертных систем и баз знаний в данной области.

Основные результаты получены в рамках Государственного контракта № П2215 с Федеральным агентством по образованию.

ЛИТЕРАТУРА:

1. Швец, В. И. Фосфолипиды в биотехнологиях / В. И. Швец // Вестник МИТХТ. – 2009. – Т. 4, № 4. – С. 4–25.
2. Гельперина, С. Э. Системы доставки лекарственных веществ на основе полимерных наночастиц / С. Э. Гельперина, В. И. Швец // Биотехнология. – 2009. – № 3. – С. 8–23.
3. Stapley, B. J. Biobibliometrics information retrieval and visualization from co-occurrence of gene names in Medline abstracts / B. J. Stapley, G. Benoit // Pac. Symp. Biocomput. – 2000. – P. 529–540.
4. Extraction of semantic biomedical relations from text using conditional random fields / M. Bundschuh [et al.] // BMC Bioinformatics. – 2008. – Vol. 9. – P. 207.
5. Малько, А. М. Группирование природных соединений с использованием ассоциативного библиометрического анализа / А. М. Малько, В. Ю. Кистанова, С. А. Гусев // Известия ТСХА. – 2009. – Вып. 4. – С. 158–163.
6. www.pubmed.org
7. nlmpubs.nlm.nih.gov/online/mesh/xmlmesh/pa2009.xml
8. <http://pubchem.ncbi.nlm.nih.gov/>
9. Создание семантических сетей белков с использованием Pubmed/Medline / Е. А. Пономаренко [и др.] // Молек. биология. – 2010. – Т. 44. – С. 1–11.
10. iProLINK: an integrated protein resource for literature mining / Z. Z. Hu [et al.] // Comput. Biol. Chem. – 2004. – Vol. 28, № 5-6. – P. 409–416.
11. Rogers, D. J. A computer program for classifying plants / D. J. Rogers, T. T. Tanimoto // Science. – 1960. – Vol. 132. – P. 1115–1118.
12. www.graphviz.org
13. Lin, J. PubMed related articles: a probabilistic topic-based model for content similarity / J. Lin, W. J. Wilbur // BMC Bioinformatics. – 2007. – Vol. 8. – P. 423.
14. Keratinized epithelial transport of beta-blocking agents. III. Evaluation of enhancing effect on percutaneous absorption using model lipid liposomes / T. Kai [et al.] // Biol. Pharm. Bull. – 1993. – Vol. 16, № 8. – P. 801–805.
15. Self-assembled polypeptide-block-poly(vinylpyrrolidone) as prospective drug-delivery systems / L. Wang [et al.] // Colloids Surf. B Biointerfaces. – 2009. – Vol. 74, № 1. – P. 284–292.
16. Wakode, R. Development and evaluation of push-pull based osmotic delivery system for pramipexole / R. Wakode, R. Bhanushali, A. Bajaj / PDA J. Pharm. Sci. Technol. – 2008. – Vol. 62, № 1. – P. 22–31.
17. Horita, A. Skin penetrating property of drugs dissolved in dimethylsulfoxide (DMSO) and other vehicles / A. Horita, L. J. Weber // Life Sci. – 1964. – Vol. 3. – P. 1389–1395.
18. Influence of heat treatment on spray-dried mixtures of Amioca starch and Carbopol 974P used as carriers for nasal drug delivery / D. Coucke [et al.] // Int. J. Pharm. – 2009. – Vol. 378, № 1-2. – P. 45–50.
19. *In situ* formation of liposome of rifampicin: better availability for better treatment / P. K. Gaur [et al.] // Curr. Drug Deliv. – 2009. – Vol. 8. – P. 25–28.
20. Obata, Y. Evaluation of pH-responsive liposomes containing amino acid-based zwitterionic lipids for improving intracellular drug delivery *in vitro* and *in vivo* / Y. Obata, S. Tajima, S. Takeoka // J. Control Release. – 2009. – Vol. 90, № 10. – P. 185–189.
21. Gürsoy, A. Co-encapsulation of isoniazid and rifampicin in liposomes and characterization of liposomes by derivative spectroscopy / A. Gürsoy, E. Kut, S. Ozkirimli // Int. J. Pharm. – 2004. – Vol. 271, № 1-2. – P. 115–123.
22. Honeywell-Nguyen, P.L. The *in vitro* transport of pergolide from surfactant-based elastic vesicles through human skin: a suggested mechanism of action / P. L. Honeywell-Nguyen, J. A. Bouwstra // J. Control Release. – 2003. – Vol. 86, № 1. – P. 145–156.
23. Liposomal quercetin efficiently suppresses growth of solid tumors in murine models / Z. P. Yuan [et al.] // Clin. Cancer Res. – 2006. – Vol. 12, № 10. – P. 3193–3199.
24. Polycarboxylic acid nanoparticles for ophthalmic drug delivery: an *ex vivo* evaluation with human cornea / T. K. De [et al.] // J. Microencapsul. – 2004. – Vol. 21, № 8. – P. 841–855.