

УДК 616-056.3:379.0:004.4

КОМПЬЮТЕРНЫЙ АНАЛИЗ ВЗАИМОСВЯЗИ АЛЛЕРГЕННОСТИ МИКРООРГАНИЗМОВ И СРЕДЫ ИХ ОБИТАНИЯ

© 2012 г. А.О. Брагин¹, П.С. Деменков¹, Е.С. Тийс¹, Р. Хофештадт²,
В.А. Иванисенко¹, Н.А. Колчанов^{1,3,4}

¹ Федеральное государственное бюджетное учреждение науки Институт цитологии и генетики
Сибирского отделения Российской академии наук, Новосибирск, Россия,
e-mail: ibragim@bionet.nsc.ru;

² Университет Билефельда, Билефельд, Германия;

³ Новосибирский национальный исследовательский государственный университет,
Новосибирск, Россия;

⁴ НИЦ «Курчатовский институт», Москва, Россия

Поступила в редакцию 1 августа 2012 г. Принята к публикации 31 августа 2012 г.

В течение 20-го столетия наблюдался стремительный рост числа аллергических заболеваний. На данный момент в индустриальных странах от аллергии страдает значительная часть населения, что делает анализ аллергенных свойств белков важной задачей. Ранее выдвигались предположения, что аллергенность белков зависит от их размера, ферментативных свойств, гомологии с белками человека и т. д. Однако анализ взаимосвязи аллергенности белков и среды обитания организма, к которому принадлежат данные белки, до сих пор не проводился. Нами были предсказаны белки-аллергены из протеомов более 500 видов микроорганизмов. Показано, что количество белков-аллергенов в протеомах микроорганизмов статистически значимо связано с патогенностью, ареалом, температурными условиями среды обитания и потребностью в кислороде этих микроорганизмов.

Ключевые слова: предсказание аллергенности белков, аллергенность протеомов микроорганизмов, среда обитания микроорганизмов, экстремофилы.

ВВЕДЕНИЕ

Аллергия является распространенной проблемой со здоровьем в наши дни. От различных аллергических заболеваний в мире страдает более одной трети населения индустриально развитых стран (WAO White Book, 2011). Элиминационная терапия остается одним из наиболее эффективных средств борьбы с появлением симптомов аллергии (Platts-Mills *et al.*, 2000). В связи с этим возникает потребность в анализе аллергенных свойств белков. Теоретические знания о потенциальной аллергенности белков из окружающих нас микроорганизмов имеют большое значение для прогнозирования и выработки стратегий предотвращения аллергических заболеваний.

Кроме экспериментальных методов оценки аллергенности белков существуют еще и компьютерные методы оценки перекрестной реактивности между анализируемым белком и аллергенами. Так, Всемирная организация здравоохранения (WHO) и Продовольственная и сельскохозяйственная организация (FAO) предложили проводить сравнение последовательности анализируемого белка с аминокислотными последовательностями известных аллергенов (FAO/WHO, 2003). Согласно предложенным этими организациями критериям, считалось, что белок имеет перекрестную реактивность с известным аллергеном, если у них имеется идентичный участок длиннее 6 аминокислотных остатков или фрагмент последовательности из 80 аминокислотных остатков анализируемого

белка идентичен последовательности белка-аллергена более чем на 35 %. Но такой подход давал большое количество ложноположительных результатов (Goodman *et al.*, 2008).

Кроме метода оценки аллергенности белков, предложенного WHO и FAO, были разработаны и другие методы предсказания аллергенности. При анализе аллергенных свойств белков было предложено использовать поиск мотивов (Stadler M.B., Stadler B.M., 2003; Kong *et al.*, 2007), вейвлет-преобразования (Li *et al.*, 2004), классификаторы, такие, как метод ближайших соседей (Zorzet *et al.*, 2002), метод опорных векторов (Saha *et al.*, 2006; Muh *et al.*, 2009) и др. Нами также был разработан метод предсказания аллергенности белков с использованием конформационных пептидов (Брагин и др., 2011).

В настоящее время активно ведутся работы по изучению молекулярных механизмов IgE-опосредованных аллергических реакций. Считается, что в норме IgE-антитела играют одну из важных ролей в реакциях иммунитета на паразитарные инфекции, аллергия же является нежелательным побочным эффектом (MacDonald *et al.*, 2002; Bischoff *et al.*, 2007). Обычно выделяют два этапа развития аллергии: сенсибилизация к определенному аллергену живого организма и выделение медиаторов воспаления при повторном контакте с аллергеном. После первого контакта с аллергеном при IgE-опосредованной аллергии начинается активная выработка антител, что приводит к сенсибилизации организма. При повторном контакте с аллергеном происходит дегрануляция тучных клеток и базофилов, при этом выделяются медиаторы, действие которых обуславливает клинические проявления аллергии (Cianferoni, Spergel, 2009; Locksley, 2010). Для дегрануляции тучных клеток требуется контакт аллергена с двумя IgE-антителами на поверхности тучных клеток, поэтому вызывать аллергию могут белки, размер которых позволяет осуществить сшивание двух реагиновых молекул на мембране тучных клеток или базофилов (Huby *et al.*, 2000). Кроме того, размер аллергена должен позволять ему проникать через защитные слои, например, через слизистый слой. Обычно аллергенами являются белки массой от 10 кДа до 70 кДа (Jeebhay *et al.*, 2001; Puc, 2003). Было также

показано, что протеазная активность аллергенов может оказывать адьювантный эффект (Sudha *et al.*, 2009). Кроме того, на аллергенность белков влияют их структурные и физико-химические свойства (Breiteneder, Mills, 2005).

Известно, что у некоторых видов организмов количество аллергенных для человека белков намного выше, чем у других. В связи с этим выдвигались гипотезы о том, что аллергенность белков организма зависит от того, насколько этот организм эволюционно близок к человеку. Так, в работе Т.А. Platts-Mills показано, что более «аллергенными» являются те виды, которые имеют меньшее родство с млекопитающими (Platts-Mills, 2012). В работе J.A. Jenkins с соавт. было показано, что белки животных, вызывающие пищевую аллергию у человека, как правило, не имеют гомологов среди белков человека (Jenkins *et al.*, 2007).

Особое место в изучении аллергенности белков занимают белки различных микроорганизмов. Микроорганизмы широко распространены в природе и используются в биотехнологии (Antranikian *et al.*, 2005), биомедицине и ветеринарии (Irwin, 2010), фармацевтике (Van den Burg, 2003; Antranikian *et al.*, 2005), текстильной промышленности (Pennisi, 1997) и т. д. Поскольку человек находится в тесном контакте с множеством видов различных микроорганизмов, то данное направление исследований является весьма актуальным. Но, несмотря на это, вопрос об аллергенности белков микроорганизмов до сих пор остается малоизученным. Одной из причин слабой изученности аллергенных свойств большинства микроорганизмов, существующих в природных условиях, является трудность их культивирования.

В связи с этим целью настоящей работы были предсказание белков-аллергенов в протеомах микроорганизмов и исследование связи аллергенности протеома с факторами среды обитания микроорганизма. Благодаря достижениям в массовом секвенировании геномов, мы смогли проанализировать многие микроорганизмы, чей геном был полностью секвенирован. Мы рассмотрели микроорганизмы, которые распространены в среде обитания, позволяющей легко контактировать с человеком, а также различные экстремофилы, для которых контакты с человеком маловероятны.

Было проанализировано более 500 видов архей и бактерий, включая экстремофилы. Показано, что аллергенность микроорганизмов (доля аллергенных белков в протеоме) статистически значимо связана с их патогенностью, ареалом, температурными условиями среды обитания и потребностью в кислороде. Оказалось, что фактором, наиболее тесно связанным с аллергенностью, является патогенность микроорганизмов. Таким образом, полученные результаты могут быть использованы для планирования экспериментов по изучению аллергенных свойств белков микроорганизмов и их связи с патогенезом заболеваний.

МАТЕРИАЛЫ И МЕТОДЫ

Предсказание аллергенности проводилось для 546 видов бактерий и архей, геномы которых были полностью секвенированы. Последовательности белков, а также характеристики микроорганизмов, включая их патогенность,

подвижность, среду обитания и т. д. (см. табл. 1), были взяты из баз данных GenBank и BioProject (Benson *et al.*, 2011; Barrett *et al.*, 2012).

Для предсказания аллергенности белков рассматриваемых микроорганизмов использовался двухшаговый подход (рис.). На первом шаге методом выравнивания аминокислотных последовательностей отбирались потенциальные белки-аллергены, которые имели высокую гомологию с известными белками-аллергенами. На втором шаге количество потенциальных белков-аллергенов увеличивалось за счет предсказания аллергенности среди белков, не имеющих высокой гомологии с известными аллергенами, с использованием разработанного нами ранее метода, основанного на анализе конформационных пептидов (Брагин и др., 2011). В методе проводился поиск конформационных и линейных пептидов из аллергенных белков в аминокислотной последовательности анализируемого белка с учетом физико-химических свойств аминокислот.

Таблица 1

Характеристики микроорганизмов

| Группа | Характеристика микроорганизма | Количество рассматриваемых видов с данной характеристикой |
|-------------------------|---|---|
| Царство | Бактерии | 493 |
| | Археи | 53 |
| Потребность в кислороде | Факультативные анаэробы | 156 |
| | Микроаэрофилы | 20 |
| | Аэробы | 198 |
| Патогенность | Патогенность по отношению к млекопитающим | 139 |
| Способ существования | Наземные | 56 |
| | Специализированные | 78 |
| | Множественные | 140 |
| Температурные факторы | Психрофилы | 12 |
| | Гипертермофилы | 38 |
| | Термофилы | 50 |
| | Мезофилы | 432 |
| Подвижность | Неподвижные | 188 |
| | Малоподвижные | 1 |
| | Подвижные | 296 |

Выравнивание проводилось с помощью программы BLAST версии 2.2.26+ (Altschul *et al.*, 1990). Использовались установленные по умолчанию параметры выравнивания программы BLAST. Порог E-value, показывающей уровень значимости сходства последовательностей, для отбора гомологов брался равным 10^{-21} . Значение порога было подобрано таким образом, чтобы ошибка перепредсказания (доля ложнопредсказанных как аллергены белков среди неаллергенных белков) была близка к нулю при максимальном значении порога для E-value. Выборка известных белков-аллергенов формировалась из базы данных SDAP (Ivanciuc *et al.*, 2003). Из базы данных были экстрагированы 960 последовательностей известных белков-аллергенов.

Расчеты проводились на высокопроизводительном кластере центра коллективного пользования «Биоинформатика» СО РАН. Использование кластера обусловлено размером поставленной задачи. Требовалось предсказать

аллергенность более 1,76 млн белков с известной последовательностью из 546 видов микроорганизмов. Использовалось распараллеливание по данным. Все белки были сгруппированы по их принадлежности к видам микроорганизмов. Вычисления для каждой группы проводились параллельно на 24 узлах с 8 ядрами на каждом. Общее время счета составило около 63 ч. В связи с суперпараллельностью задачи было достигнуто практически линейное ускорение.

Аллергенность микроорганизма рассчитывалась как отношение количества белков, предсказанных как аллергены, к общему числу белков в протеоме данного микроорганизма.

Для установления связи между характеристиками микроорганизмов (табл. 1) и их аллергенностью использовались методы множественной линейной регрессии и дисперсионный анализ, представленные в стандартной библиотеке языка R (R Development Core Team, 2011). Характеристики микроорганизмов, представленные в табл. 1, рассматривались как независимые

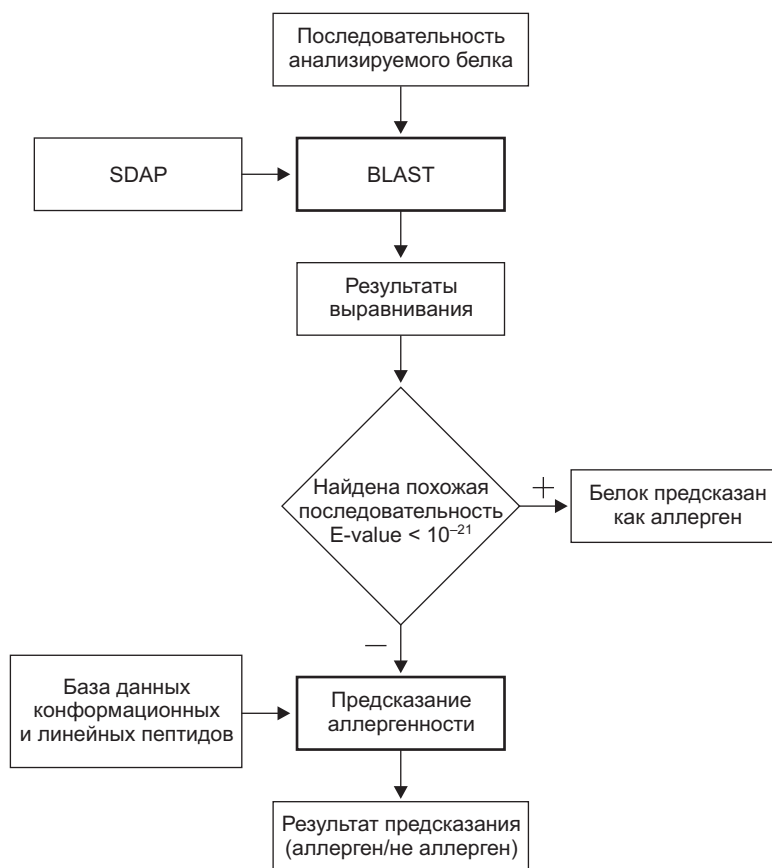


Рис. Блок-схема подхода для анализа аллергенности белков микроорганизмов.

переменные, а аллергенность – как зависимая переменная. Независимые переменные принимали значения 0 или 1. Например, если вид архей или бактерий относился к термофилам, то характеристика «термофильность» этого микроорганизма задавалась как 1, в противном случае – как 0.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Результаты дисперсионного анализа связи между характеристиками микроорганизмов и их аллергенностью приведены в табл. 2.

С помощью дисперсионного анализа были выявлены 4 наиболее значимые характеристики микроорганизмов, связанные с аллергенностью (см табл. 2). Среди них на первом месте по значимости оказалась патогенность микроорганизма. При этом, согласно нашим предсказаниям, патогенные микроорганизмы обладают повышенной аллергенностью. *Staphylococcus aureus* является одним из примеров патогенных микроорганизмов, для которых есть данные, свидетельствующие об их аллергенности. Так, в работе K. Reginald с соавт. показано, что IgE-антитела человека, страдающего атопическим дерматитом, могут взаимодействовать с белками *S. aureus*, что является показателем возможной аллергенности таких белков (Reginald *et al.*, 2011). Известно также, что у людей, страдающих аллергией, могут наблюдаться изменения в составе биоценоза кишечника, в котором могут присутствовать условно-патогенные бактерии родов *Staphylococcus*, *Klebsiella*, *Proteus* (Макарова, Боровик, 2008). В связи с этим интерес могут представлять дальнейшие исследования,

Таблица 2

Характеристики микроорганизмов, статистически значимо связанные с аллергенностью

| Характеристика микроорганизма* | Направление связи с аллергенностью | Статистика Фишера | Значимость |
|--------------------------------|------------------------------------|-------------------|------------|
| Патогенность** | + | 31,7 | 3,447e-08 |
| Наземный способ существования | - | 21,1 | 5,683e-06 |
| Аэробы | - | 12,9 | 0,0003571 |
| Мезофилы | - | 12,6 | 0,0004324 |

* Приведены характеристики с уровнем значимости менее 0,001. ** Патогенность рассматривалась по отношению к млекопитающим.

направленные на выяснение связи между аллергенностью и патогенностью микроорганизмов.

Другие выявленные нами менее значимые факторы (наземный способ существования, принадлежность к аэробам и мезофилам) оказались отрицательно связанными с аллергенностью. Среди микроорганизмов, ведущих наземный способ существования, а также мезофилов и аэробов, согласно предсказанию, аллергенные белки встречаются реже по сравнению с остальными микроорганизмами. Следует отметить, что такие микроорганизмы могут чаще других микроорганизмов, обитающих в экстремальных условиях, контактировать с людьми.

Согласно нашим предсказаниям, наиболее аллергенными среди патогенных для человека микроорганизмов могут считаться *Bacillus*

Таблица 3

Примеры некоторых видов патогенных для человека бактерий, для которых предсказано наибольшее значение показателя аллергенности

| Вид | Заболевание | Показатель аллергенности |
|---|--|--------------------------|
| <i>Bacillus weihenstephanensis</i> KBAB4 | Пищевое отравление | 0,1009 |
| <i>Streptococcus pneumoniae</i> 670-6B | Пневмония | 0,0841 |
| <i>Staphylococcus saprophyticus</i> subsp. saprophyticus ATCC 15305 | Инфекции мочевыводящих путей | 0,0819 |
| <i>Staphylococcus haemolyticus</i> JCSC1435 | Широкий спектр оппортунистических инфекций | 0,0812 |

weihenstephanensis, *Streptococcus pneumonia*, *Staphylococcus saprophyticus* и *Staphylococcus haemolyticus*, которые могут вызывать пищевое отравление (Lapidus *et al.*, 2008), пневмонию (Jauneikaite *et al.*, 2012), инфекции мочевыводящих путей (Kuroda *et al.*, 2005) и оппортунистические инфекции (Takeuchi *et al.*, 2005) соответственно (см. табл. 3).

Полученные результаты могут быть использованы при планировании экспериментов по проверке аллергенных свойств белков микроорганизмов.

БЛАГОДАРНОСТИ

Работа была поддержана грантом РФФИ «Компьютерный анализ взаимосвязи аллергенности микроорганизмов и условий их обитания» № 12-04-31892-мол_a, DAAD Leonard Euler Program grant Nr. 50024820 (AB, ET, TI) и Министерством образования и науки РФ (Госконтракт № 14.740.11.0001). Коллектив авторов выражает благодарность Ольге Владимировне Сайк за помощь в написании статьи.

ЛИТЕРАТУРА

- Брагин А.О., Деменков П.С., Иванисенко В.А. Предсказание аллергенности белков с использованием информации о конформационных пептидах // Вавилов. журн. генет. и селекции. 2011. Т. 15. № 3. С. 462–468.
- Макарова С.Г., Боровик Т.Э. Дисбиоз кишечника у детей с пищевой аллергией: патогенетические аспекты и современные методы коррекции // Вопр. соврем. педиатрии. 2008. Т. 7. № 2. С. 82–92.
- Altschul S.F., Gish W., Miller W. *et al.* Basic local alignment search tool // J. Mol. Biol. 1990. V. 215. No. 3. P. 403–410.
- Antranikian G., Vorgias C.E., Bertoldo C. Extreme environments as a resource for microorganisms and novel biocatalysts // Adv. Biochem. Eng. **Biotechnol.** 2005. V. 96. P. 219–262.
- Barrett T., Clark K., Gevorgyan R. *et al.* BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata // Nucl. Acids Res. 2012. V. 40. (Database issue) D57–63.
- Benson D.A., Karsch-Mizrachi I., Lipman D.J. *et al.* GenBank // Nucl. Acids Res. 2011. V. 39. (Database issue) D32–37.
- Bischoff S.C., Krämer S. Human mast cells, bacteria, and intestinal immunity // Immunol. Rev. 2007. V. 217. No. 1. P. 329–337.
- Breiteneder H., Mills E.N. Molecular properties of food allergens // J. Allergy Clin. Immunol. 2005. V. 115. No. 1. P. 14–23.
- Cianferoni A., Spergel J.M. Food allergy: review, classification and diagnosis // Allergol. Int. 2009. V. 58. P. 457–466.
- FAO/WHO. Codex Principles and Guidelines on Foods Derived from Biotechnology. 2003.
- Goodman R.E., Vieths S., Sampson H.A. *et al.* Allergenicity assessment of genetically modified crops – what makes sense? // Nat. Biotechnol. 2008. V. 26. No. 1. P. 73–81.
- Huby R.D., Dearman R.J., Kimber I. Why are some proteins allergens? // Toxicol. Sci. 2000. V. 55. No. 2. P. 235–246.
- Irwin J.A. Extremophiles and their application to veterinary medicine // Environ. Technol. 2010. V. 31. P. 857–869.
- Ivanciuc O., Schein C.H., Braun W. SDAP: database and computational tools for allergenic proteins // Nucl. Acids Res. 2003. V. 31. No. 1. P. 359–362.
- Jauneikaite E., Jefferies J.M., Hibberd M.L., Clarke S.C. Prevalence of *Streptococcus pneumoniae* serotypes causing invasive and non-invasive disease in South East Asia: a review // Vaccine. 2012. V. 30. No. 24. P. 3503–3514.
- Jeebhay M.F., Robins T.G., Lehrer S.B., Lopata A.L. Occupational seafood allergy: a review // Occup. Environ. Med. 2001. V. 58. No. 2. P. 553–562.
- Jenkins J.A., Breiteneder H., Mills E.N. Evolutionary distance from human homologs reflects allergenicity of animal food proteins // J. Allergy Clin. Immunol. 2007. V. 120. No. 6. P. 1399–1405.
- Kong W., Tan T.S., Tham L., Choo K.W. Improved prediction of allergenicity by combination of multiple sequence motifs // In Silico Biol. 2007. V. 7. No. 1. P. 77–86.
- Kuroda M., Yamashita A., Hirakawa H. *et al.* Whole genome sequence of *Staphylococcus saprophyticus* reveals the pathogenesis of uncomplicated urinary tract infection // Proc. Natl Acad. Sci. USA. 2005. V. 102. No. 37. P. 13272–13277.
- Lapidus A., Goltsman E., Auger S. *et al.* Extending the *Bacillus cereus* group genomics to putative food-borne pathogens of different toxicity // Chem. Biol. Interact. 2008. V. 171. No. 2. P. 236–249.
- Li K.B., Issac P., Krishnan A. Predicting allergenic proteins using wavelet transform // Bioinformatics. 2004. V. 20. No. 16. P. 2572–2578.
- Locksley R.M. Asthma and allergic inflammation // Cell. 2010. V. 140. No. 6. P. 777–783.
- MacDonald A.S., Araujo M.I., Pearce E.J. Immunology of parasitic helminth infections // Infect. Immun. 2002. V. 70. No. 2. P. 427–433.
- Muh H.C., Tong J.C., Tammi M.T. AllerHunter: a SVM-pairwise system for assessment of allergenicity and allergic cross-reactivity in proteins // PLoS One. 2009. V. 4. No. 6. e5861.
- Pennisi E. In industry, extremophiles begin to make their mark // Science. 1997. V. 276. No. 5313. P. 705–706.
- Platts-Mills T.A. Allergy in evolution // Chem. Immunol. Allergy. 2012. V. 96. P. 1–6.
- Platts-Mills T.A., Vaughan J.W., Carter M.C., Woodfolk J.A. The role of intervention in established allergy: avoidance of indoor allergens in the treatment of chronic allergic disease // J. Allergy Clin. Immunol. 2000. V. 106. No. 5. P. 787–804.
- Puc M. Characterisation of pollen allergens // Ann. Agric. Environ. Med. 2003. V. 10. No. 2. P. 143–149.
- R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical

- Computing. Vienna, Austria. 2011. URL <http://www.R-project.org/>.
- Reginald K., Westritschnig K., Werfel T. *et al.* Immunoglobulin E antibody reactivity to bacterial antigens in atopic dermatitis patients // *Clin. Exp. Allergy*. 2011. V. 41. No. 3. P. 357–369.
- Saha S., Raghava G.P. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes // *Nucl. Acids Res.* 2006. 34(Web Server issue):W202–209.
- Stadler M.B., Stadler B.M. Allergenicity prediction by protein sequence // *FASEB J.* 2003. V. 17. No. 9. P. 1141–1143.
- Sudha V.T., Arora N., Singh B.P. Serine protease activity of Per a 10 augments allergeninduced airway inflammation in a mouse model // *Eur. J. Clin. Invest.* 2009. V. 39. No. 6. P. 507–516.
- Takeuchi F., Watanabe S., Baba T. *et al.* Whole-genome sequencing of staphylococcus haemolyticus uncovers the extreme plasticity of its genome and the evolution of human-colonizing staphylococcal species // *J. Bacteriol.* 2005. V. 187. No. 21. P. 7292–7308.
- Van den Burg B. Extremophiles as a source for novel enzymes // *Curr. Opin. Microbiol.* 2003. V. 6. No. 3. P. 213–218.
- WAO White Book on Allergy / Eds R. Pawankar, G.W. Canonica, S.T. Holgate, R.F. Lockey. Milwaukee, Wisconsin: World Allergy Organization, 2011. P. 12.
- Zorzet A., Gustafsson M., Hammerling U. Prediction of food protein allergenicity: a bioinformatic learning systems approach // *In Silico Biol.* 2002. V. 2. No. 4. P. 525–534.

COMPUTERIZED ANALYSIS OF THE RELATIONSHIP BETWEEN ALLERGENICITY OF MICROORGANISMS AND THEIR HABITATS

A.O. Bragin¹, P.S. Demenkov¹, E.S. Tiys¹, R. Hofestädt², V.A. Ivanisenko¹, N.A. Kolchanov^{1,3,4}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: ibragim@bionet.nsc.ru;

² Bio-Medical Informatics Department, Bielefeld University, Bielefeld, Germany;

³ Novosibirsk National Research State University, Novosibirsk, Russia;

⁴ National Research Centre «Kurchatov Institute», Moscow, Russia

Summary

The prevalence of allergic diseases was rapidly increasing in the 20th century. Currently, many people suffer from allergy in industrial countries. Therefore, analysis of allergenic properties of proteins is an urgent task. The following factors were formerly hypothesized to determine the allergenicity of a protein: size, enzymatic properties, and similarity to human proteins. However, no analysis of the relationship between allergenicity of proteins and the habitat of the organisms producing them has been conducted hitherto. We predict allergenicity of proteins from proteomes of more than 500 species of microorganisms. It is shown that the number of allergenic proteins in the proteomes of microorganisms is significantly associated with their pathogenicity, habitat, temperature conditions of the habitat, and oxygen demand.

Key words: protein allergenicity prediction, allergenicity of microbial proteomes, habitat of microorganisms, extremophiles.