



6. *Обоснование* использования специалистами средств индивидуальной защиты при воздействии авиационного шума / В.Н.Зинкин, С.К.Солдатов, А.В.Богомолов, А.П.Шведов // Информатика и системы управления. – 2009. – № 4(22).
7. *Особенности* условий труда и заболеваемости инженерно-технического состава авиации / В.И.Свидовый, В.Н.Зинкин, С.К.Солдатов, Ю.А.Кукушкин // Вестник Санкт-Петербургской государственной медицинской академии им. И.И.Мечникова. – 2006. – № 2. – С. 46 – 49.
8. *Результаты* исследований акустической обстановки на рабочих местах инженерно-технического состава авиации / С.А.Щербаков, Ю.А.Кукушкин, А.В.Богомолов и др. // Проблемы безопасности полетов. – 2007. – №3. – С.27–32.
9. *Ушаков И.Б., Кукушкин Ю.А., Богомолов А.В.* Физиология труда и надежность деятельности оператора. – М.: Наука, 2008.
10. *Passchier-Vermeer W., Passchier W.F.* Noise exposure and public health // Environmental health perspect. – Vol. 108, supplement 1. 2000. – P.123-131.
11. *McLean E.K., Tarnopolsky A.* Noise, discomfort and mental health: a review of the socio-medical implications of disturbance by noise // Psychological Medicine, – 1977, № 7. – P. 19-62.

Статья представлена к публикации членом редколлегии Ю.М. Перельманом.

E-mail:

Зинкин В.Н. – zinkin-vn@yandex.ru

УДК 51-76:57.087.1

© 2011 г. **И.Э. Памирский**

(Амурская государственная медицинская академия, Благовещенск),

К.С. Голохваст, канд. биол. наук

(Дальневосточный федеральный университет, Владивосток)

БИОИНФОРМАЦИОННОЕ ИССЛЕДОВАНИЕ ГОМОЛОГИИ И ЭВОЛЮЦИИ БЕЛКОВ, УЧАСТВУЮЩИХ В БИОМИНЕРАЛИЗАЦИИ

В статье приводятся примеры собственного биоинформационного исследования пептидов и белков (силикатеинов, силиказы, силаффинов, транспортеров кремния, магнетосомных белков), участвующих в биоминерализации, с использованием современных компьютерных методов.

Ключевые слова: биоинформатика, выравнивание последовательностей, гомология белков.

Введение

К актуальным задачам биологии, химии и смежных научных отраслей относится исследование физико-химических свойств молекул, взаимодействия веществ между собой, влияния веществ на живой организм, создание молекул с заданными свойствами. Решение подобных задач носит не только фундаментальный, но и прикладной характер. Тестирование свойств, в том числе биологиче-

ских эффектов, любого соединения классическим опытным путем является довольно сложным процессом. Как правило, подобные исследования состоят из долгих, трудоемких и дорогостоящих этапов, и многие эксперименты доступны не всем исследователям. Одним из подходов к решению этой проблемы является применение компьютерных методов, представляющих собой электронные базы данных веществ различной природы, вычислительные интернет-сервера и компьютерные программы молекулярного моделирования. В совокупности данные методы легли в основу сравнительно молодой науки – *биоинформатики*, относящейся к области вычислительной биологии, а сам биоинформационный подход получил название *in silico* (дословно «в кремнии»). В последнее время компьютерные методы широко применяются в фармакологических разработках и испытаниях, предсказании и открытии новых биологических целей веществ [1 – 4]. Примером результатов использования биоинформатики является открытие новых противораковых препаратов [5]. За рубежом вычислительное моделирование рекомендуется внедрить как средство наблюдения за метаболическими изменениями организма [6]. Вычислительная биология не только ускоряет процесс изучения веществ, делая этот процесс экономически выгодным, но и представляет совершенно новый путь исследований [7].

Многие биоинформационные приложения размещены в сети Интернета в свободном доступе и понятны практически любому пользователю, поскольку не требуют глубоких знаний в области информационных технологий. Однако работа с данными приложениями требует освоения определенной терминологии и языка структурированных запросов (например, SQL). Биоинформатика активно развивается и уже ориентирована не только на низкомолекулярные вещества, но и на высокомолекулярные соединения и комплексы – такие как пептиды, белки, а также гены и целые геномы. В частности, с помощью интернет-серверов и компьютерных программ можно определить функции белка, исходя из его последовательности. Сегодня стало возможным аннотировать белки, основываясь лишь на последовательности аминокислотных остатков в полипептидной цепи.

Пример алгоритма исследования белков *in silico*

Применение методов биоинформатики мы рассмотрим на примере собственного исследования гомологии ряда белков. Объектами данного исследования стали силикатеины, силиказа, силаффины, транспортеры кремния, магнетосомные белки, участвующие в формировании/разрушении неорганических структур на основе кремния (кремнезем) и железа (магнетит).

В Интернете нами было найдено около 50 международных, в основном англоязычных, электронных баз данных, хранящих информацию о веществах пептидной природы: [AARSDB](#), [ASTRAL](#), [BRENDA](#), [COG](#), [GTOP](#), [HISTONES](#), [LIGAND](#) и др.

В данных базах содержится информация следующего рода: идентификационный номер молекулы в конкретной базе; описание молекул (уровни организации, предшественники, реакционные и связывающие центры, механизм действия, молекулы-мишени, спектр биологической активности, белки-гомологи); класси-

фикация и таксономические сведения (кланы, подкланы, семейства, подсемейства); ссылки на литературу.

Вся информация в виде электронных файлов, различного текстового, графического и других специальных форматов представлена на рис. 1, где каждый аминокислотный остаток обозначен прописной английской буквой, а числа указывают номер аминокислоты в последовательности белка.

```

      10      20      30      40      50      60
MFTSVLLLCI TSQVLAVNLP QYEFVEEWKM WKNRHGMSYE SDLHDLDKHL VRLSNKKFIE

      70      80      90     100     110     120
LHNANSHIFG YSLAMNHFGD LTDLEWNEKY GTYSSSSAGN YTKVFKADPY QSYPEVDWR

     130     140     150     160     170     180
TKGAVTSVKD QSQCASAYAF SAMAALEGAN ALATDTLVNL SEQNLIDCSV PYGNHGCKGG

     190     200     210     220     230     240
NMLYAFKYVI ANEGVDTANS YPFYGKQSSC VYNEKYAAVK ISGMVRISQG SESDLLGAVA

     250     260     270     280     290     300
NVGPPVAVAID GSSNAFRFYS SGVYDSSRCS SSKLNHAIWV TGYGSYSGKK YWLVKNSWGK

     310     320
NWGNYGYIMM ARGKYNQCGI ASDASYPTL

```

Рис.1. Последовательность силикатеина А1 губки в формате FASTA.

Для дальнейшей работы использовали базу данных белков *UniProt/Swiss-Prot* (www.uniprot.org), в которой на момент исследования содержалась информация о более чем 10 млн. последовательностях. Было обнаружено 5 силаффинов, более 100 транспортеров кремния, около 100 магнетосомных белков и около 50 силикатеинов. Последовательность силиказы брали из работы группы авторов под руководством Шредера [8].

При помощи инструментов «*Blast*» (Basic Local Alignment Search Tool) и «*Align*» («Выравнивание»), доступных на этой базе данных, был проведен анализ исследуемых белков внутри семейств, а также поиск гомологов из числа других белков. «*Align*» и «*Blast*» работают с форматом FASTA, но также поддерживают и другие идентификаторы последовательности, в том числе из других баз. Подобные инструменты реализованы и в виде отдельного программного обеспечения (*Blast*, *ClustalW*, *GenDoc*, *BioEdit* и др.). Через «*Blast*» (использовалась матрица BLOSUM62), предназначенный для поиска гомологичных последовательностей методом попарного выравнивания, была установлена степень гомологии между исследуемыми протеинами. Также был осуществлен поиск гомологов этих белков.

Принцип подсчета гомологии последовательностей проводится согласно матрице стохастической оценки (рис. 2) или оценивающей матрице (PAM40, PAM80, BLOSUM, BLOSUM62, DAYHOFF и др.), выбор которой определяется пользователем. На рис. 2 каждый остаток известной аминокислоты обозначен прописной английской буквой, неизвестной аминокислоты – звездочкой. Числа указывают количество очков согласно алгоритму выравнивания.

BLOSUM62 [scoring matrix](#)

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X	*
A	4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-1	-2	-1	1	0	-3	-2	0	-2	-1	0	-4
R	-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-1	-3	-2	-3	-1	0	-1	-4
N	-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-2	-3	3	0	-1	-4
D	-2	-2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-3	-1	0	-1	-4	-3	-3	4	1	-1	-4
C	0	-3	-3	-3	9	-3	-4	-3	-3	-1	-1	-3	-1	-2	-3	-1	-1	-2	-2	-1	-3	-3	-2	-4
Q	-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-3	-1	0	-1	-2	-1	-2	0	3	-1	-4
E	-1	0	0	2	-4	2	5	-2	0	-3	-3	1	-2	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
G	0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3	-1	-2	-1	-4
H	-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	2	-3	0	0	-1	-4
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	3	-3	-3	-1	-4
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	-2	2	0	-3	-2	-1	-2	-1	1	-4	-3	-1	-4
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	-2	0	1	-1	-4
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5	0	-2	-1	-1	-1	-1	1	-3	-1	-1	-4
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1	-3	-3	-1	-4
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2	-2	-1	-2	-4
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4	1	-3	-2	-2	0	0	0	-4
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5	-2	-2	0	-1	-1	0	-4
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11	2	-3	-4	-3	-2	-4
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7	-1	-3	-2	-1	-4
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	-3	-2	-1	-4
B	-2	-1	3	4	-3	0	1	-1	0	-3	-4	0	-3	-3	-2	0	-1	-4	-3	-3	4	1	-1	-4
Z	-1	0	0	1	-3	3	4	-2	0	-3	-3	1	-1	-3	-1	0	-1	-3	-2	-2	1	4	-1	-4
X	0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	-2	0	0	-2	-1	-1	-1	-1	-1	-1	-4
*	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	1

Рис. 2. Пример матрицы оценки BLOSUM62.

Матрица выдает оценку по каждой позиции выравнивания на основе частоты совпадений в пределах сравниваемых последовательностей, а результат выражает как пропорцию идентичных аминокислотных остатков. Пример основного алгоритма выравнивания:

$$S_{ij} = \max \begin{cases} P_{ij}, \\ S_{i-Ij-I} + sub(a_i, b_j), \\ Q_{ij}; \end{cases} \quad (1)$$

$$P_{ij} = \max \begin{cases} S_{i-Ij} + vw, \\ P_{i-Ij} + v; \end{cases} \quad Q_{ij} = \max \begin{cases} S_{ij-I} + wl, \\ q_{i-Ij} + v, \end{cases}$$

где i и j – колонки и столбцы матрицы, в которых записаны последовательности a и b ; $sub(a_i, b_j)$ – оценка (согласно оценивающей матрице) спаривания остатка i последовательности $a(a_i)$ с остатком j последовательности $b(b_j)$; wl – штрафной балл за создание пробела; v – штрафной балл за расширение пробела; $S_{i,j}$ – общая оценка выравнивания позиций i,j в матрице S при каждом совпадении. $P_{i,j}$ и $Q_{i,j}$ – оценки, показывающие величину для каждой возможной позиции выравнивания, учитывающие штрафную систему.

Инструмент «Align» разработан с целью одновременного выравнивания множества аминокислотных и нуклеотидных последовательностей (multiple sequence alignment) и на базе данных белков UniProt/Swiss-Prot реализован в приложении «ClustalW». Оценка для каждой пары последовательностей может генерироваться следующим образом: 1) все позиции сравниваются непосредственно для каждой пары последовательностей, поочередно; 2) все пустоты, а также сим-

отрицательно заряженные остатки. Параллельно с оценкой гомологии возможно получение и данных по количественному и процентному соотношению аминокислотных остатков в последовательности, графиков энтропии выравнивания (мера отсутствия «информационного содержимого» по каждой позиции выравнивания на основе информационной теории Клода Шеннона), профиля гидрофобности и ряда других показателей.

Следующим этапом было выявление родственных связей между исследуемыми белками и их гомологами. Результаты могут быть представлены в различном виде (диаграммы, таблицы и др.). Пример результата таксономического анализа группы силикатеинов приведен на рис. 5, где вместо названий белков приведен ID (идентификационный номер) по UniProt/Swiss-Prot.

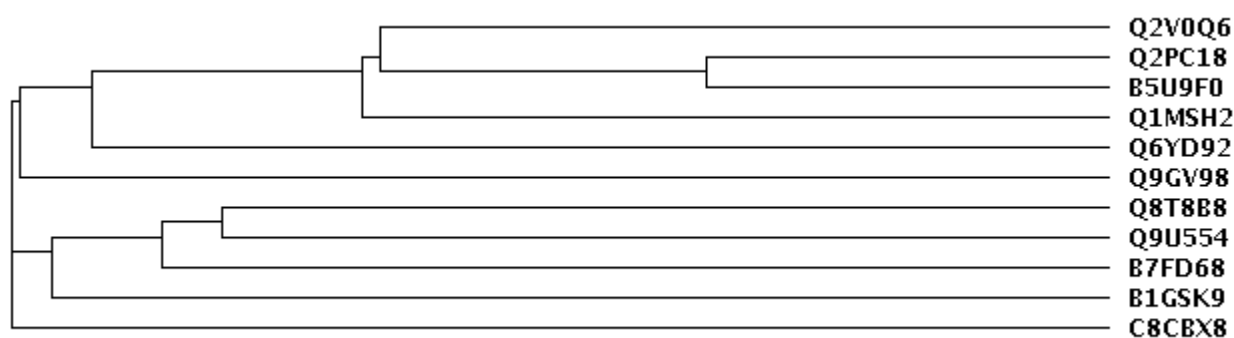


Рис. 5. Таксономическое распределение некоторых силикатеинов губок на основе выравнивания последовательностей.

In silico показано близкое родство исследуемых белков с ферментами различных семейств, широко распространенных по филогенетическому дереву. Спектр биологических функций большинства гомологов еще не установлен, но некоторые из них участвуют в процессах биоминерализации, что позволяет предположить происхождение данных белков от одной эволюционной группы белков (или белка), имеющих тесную связь с минералообразованием.

Заключение

В ходе исследования установлено, что степень гомологии между силикатеинами лежит в пределах 40-72%, силаффинами – 90-92%, транспортерами кремния – 45-99%, магнетосомными белками – до 76%. Найденные гомологи транспортеров кремния (гомология до 40%), силикатеинов (гомология более 40%), силиказы губки (гомология 29-40%), силаффинов (гомология 30-40%) и некоторых магнетосомных белков (гомология 20-45%) содержатся во многих видах растений и животных, от одноклеточных до высших организмов. Таксономическое исследование показало близкие родственные связи между исследуемыми белками и их гомологами. Полученные нами данные согласуются с результатами более ранних исследований ряда авторов [13 – 15]. Исходя из вышеизложенного, можно судить об адекватности методов биоинформатики (описанных в данной работе) в исследовании гомологии и эволюции белков. В то же время, полученные таким методом данные не следует рассматривать как истину до подтверждения вруч-

ную. Применение подобных методов возможно в исследовании патологических процессов, протекающих с образованием минералов, формирование которых сопряжено с работой белков.

ЛИТЕРАТУРА

1. Protein structure prediction in structure based drug design / *Takeda-Shitaka M., Takaya D., Chiba C. et al.* // *Curr. Med. Chem.* – 2004. – Vol. 11, № 5. – P.551-558.
2. Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery / *Blundell T.L., Bancinyane L. Sibanda K. et al.* // *Phil. Trans. R. Soc. B.* – 2006. – Vol. 361, № 1467. – P. 413–423.
3. *Ekins S., Mestres J., Testa B.* In silico pharmacology for drug discovery: applications to targets and beyond // *British Journal of Pharmacology.* – 2007. – Vol. 152. – P.21–37.
4. *Li Q., Lai L.* Prediction of potential drug targets based on simple sequence Properties // *BMC Bioinformatics.* – 2007. – Vol. 8. – P.353-364.
5. *Geromichalos G.D.* Importance of molecular computer modeling in anticancer drug development // *J. Buon.* – 2007. – Vol. 12, Suppl. 1. – P.101-118.
6. *Hall K.D., Baracos V.E.* Computational modeling of cancer cachexia // *Curr. Opin. Clin. Nutr. Metab. Care.* – 2008. – Vol. 11, № 3. – P.214-221.
7. *Ambesi-Impiombato A., Bernardo D.* Computational Biology and Drug Discovery: From Single-Target to Network Drugs // *Curr. Bioinformatics.* – 2006. – Vol. 1, № 4. – P.3-13.
8. Silicateins, silicase and spicule-associated proteins: synthesis of demosponge silica skeleton and nanobiotechnological applications / *Schroder H.C., Krasko A., Brandt D. et al.* // *Biodiversity, innovation and Sustainability.* – 2007. – P.581-592.
9. *Smith T.F., Waterman M.S.* Identification of common molecular subsequences // *J. Mol. Biol.* – 1981. – Vol. 147, № 1. – P.195-197.
10. *Myers E.W., Miller W.* Optimal alignments in linear space // *Comput. Appl. Biosci.* – 1988. – Vol. 4, № 1. – P.11-17.
11. *Gotoh O.* An improved algorithm for matching biological sequences // *J. Mol. Biol.* – 1982. – Vol. 162, № 3. – P.705-708.
12. *Needleman S.B., Wunsch C.D.* A general method applicable to the search for similarities in the amino acid sequence of two proteins // *J. Mol. Biol.* – 1970. – Vol. 48, № 3. – P.443-453.
13. Silicatein alpha: cathepsin L-like protein in sponge biosilica / *Shimizu K., Cha J., Stucky G.D., Morse D.E.* // *PNAS.* – 1998. – № 95. – P.6234-6238.
14. The unique skeleton of siliceous sponges (Porifera; Hexactinellida and Demospongiae) that evolved first from the Urmetazoa during the Proterozoic: a Review / *Müller W.E.G., Li J., Schroder H.C. et al.* // *Biogeosciences Discuss.* – 2007. – № 4. – P.385–416.
15. *Bduerlein E.* Biomineralization of unicellular organisms: an unusual membrane biochemistry for the production of inorganic nano- and microstructures // *Angew. Chem. Int. Ed. Engl.* – 2003. – Vol. 10, № 42. - P . 614-641.

Статья представлена к публикации членом редколлегии Ю.М. Перельманом.

E-mail:

Памирский И.Э. – parimski@mail.ru;

Голохваст К.С. – droopy@mail.ru.