

Сравнительный анализ программ, предсказывающих трансмембранные сегменты в трансмембранных белках

Н.С.Садовская*, Р.А.Сутормин**, А.Б.Рахманинова***, М.С.Гельфанд****

*Институт проблем передачи информации РАН
101447, Москва, Большой Каретный переулок, 19, Россия
e-mail: natasha@imb.imb.ac.ru
тел. (095) 135-20-41

**Механико-математический факультет МГУ
Россия, 119899, Воробьевы горы

***Integrated Genomics, Россия, а/я 348, Москва 117337

****ГосНИИГенетика, Россия, Москва, 113545, 1-й Дорожный пр. 1

Поступила в редколлегию 20.05.2002

Аннотация—Предсказание трансмембранных (ТМ) спиралей в интегральных мембранных белках является важным аспектом биоинформатики. Большая часть методов пытается предсказать не только индивидуальные ТМ спирали, но и полную топологию белка, т.е. общее число ТМ спиралей и их ориентацию относительно мембраны. Предсказание топологии имеет важное применение в геномном анализе. Оно может использоваться для выделения общего направления в эволюции мембранных белков. При попытке получить пространственную структуру интегральных мембранных белков возникает ряд сложностей. Методы кристаллографии показали третичную структуру примерно для 20 ТМ белков. Была сделана предварительная оценка четырех наиболее часто используемых ТМ серверов.

1. ВВЕДЕНИЕ

Предсказание трансмембранных (ТМ) спиралей в интегральных мембранных белках является важным аспектом биоинформатики. Большая часть методов пытается предсказать не только индивидуальные ТМ спирали, но и полную топологию белка, т.е. общее число ТМ спиралей и их ориентацию относительно мембраны. Предсказание топологии имеет важное применение в геномном анализе. Оно может использоваться для выделения общего направления в эволюции мембранных белков.

При попытке получить пространственную структуру интегральных мембранных белков возникает ряд сложностей. Методы кристаллографии показали третичную структуру примерно для 20 ТМ белков.

Была сделана предварительная оценка четырех наиболее часто используемых ТМ серверов.

2. АНАЛИЗ НАИБОЛЕЕ ЧАСТО ИСПОЛЬЗУЕМЫХ СЕРВЕРОВ

Везде далее ТМ сегмент — это предсказанный ТМ сегмент. Анализ предсказания ТМ серверов основан на следующих предположениях:

1. В группе ортологичных белков сохраняется длина ТМ элементов;
2. При выравнивании относительные положения ТМ спиралей также сохраняются, т.е. система “ортологичных” ТМ спиралей при помощи выравнивания проецируется на один отрезок.

Следовательно, при сравнении ТМ сегментов, предсказанных для пары выровненных гомологичных белков, величина qq , равная отношению суммы пересечения всех ТМ сегментов к их суммарному

объединению, должна быть близка к единице (рис. 1). Величину kfs считали следующим образом: при анализе пары пересекающихся ТМ сегментов, имеющих длины l_i и l_j соответственно, считали величины kfs_{ij} и kfs_{ji} , равные отношению пересечения ТМ сегментов к длине соответствующего ТМ сегмента. При значении ≥ 0.5 им присваивали значение 1, при значении < 0.5 их считали равными 0. Далее считали отношение суммы kfs_{ij} и kfs_{ji} для всех ТМ спиралей к сумме ТМ спиралей в обоих белках. Соответственно, при совпадении предсказаний значение kfs тоже должно быть близко к единице.

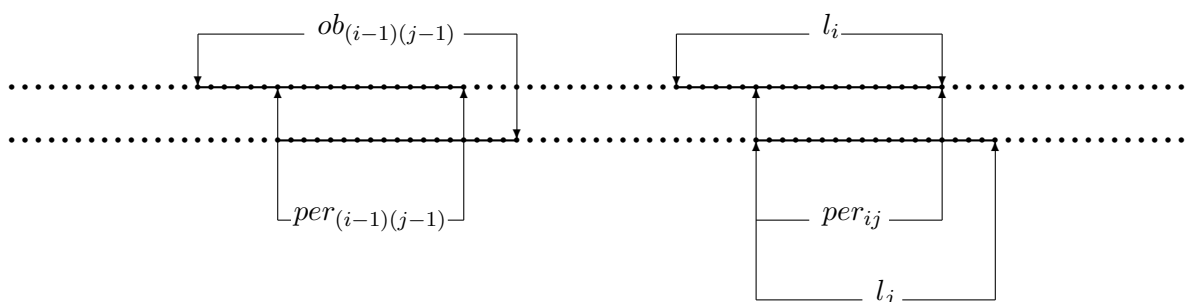


Рис. 1. Подсчет величин qq и kfs . Сплошной линией указаны ТМ сегменты, пунктиром — петли двух выровненных ортологических ТМ белков.

$$qq = \sum_{i,j} per_{ij} / \sum_{i,j} ob_{ij}, \quad kfs_{ij} = per/l_i, \quad kfs_{ji} = per/l_j, \quad kfs = \sum_{i,j} (kfs_{ij} + kfs_{ji}) / \sum_{i,j} (tm_i + tm_j),$$

где ob_{ij} — объединение, per_{ij} — пересечение ТМ сегментов первого и второго белков, l_i — длины ТМ сегментов первого ТМ белка, l_j — длины ТМ сегментов второго ТМ белка, tm_i — число ТМ спиралей в первом ТМ белке, tm_j — число ТМ спиралей во втором ТМ белке.

Для анализа взяли по одному случайному представителю бактериальных вторичных транспортеров (класс ТС.2А согласно классификации М. Сайера и соавт. (<http://www-biology.ucsd.edu/~msaier/transport/>) [1],[2]) из следующих семейств: GNTP (2А.8), MHS (2А.1.6), NCS2 (2А.40). Для каждого представителя подобрали три группы ортологических генов, согласно их проценту сходства, с использованием программы ERGO (<http://ergo.integratedgenomics.com/ERGO/>) [3]:

- 40%-60%;
- 60%-80%;
- 80%-100%.

Всего проанализировали 26 белков. Для семейства GNTP в таблицах 1, 2 приведены матрицы, показывающие процент сходства между анализируемыми белками.

		1	2	3	4
RSAX01242	1	100	99	88	62
RZW01701	2	99	100	88	62
RZY00850	3	88	88	100	65
RBS04001	4	62	62	65	100

Таблица 1. Процент сходства анализируемых белков для семейства GNTP.

Полученные аминокислотные последовательности выравнивали программой ClustalW [4]. Положение потенциальных ТМ элементов определили с использованием четырех наиболее часто используемых серверов:

DAS (<http://www.sbc.su.se/~miklos/DAS/>) [5],

TMAP (<http://www.mbb.ki.se/tmap/>) [6], ТМНММ1.0 (<http://www.cbs.dtu.dk/services>

/ТМНММ-2.0) [7], TMPred (http://www.ch.embnet.org/software/TMPRED_form.html).

		1	2	3	4	5	6	7	8	9	10	11	12	13
RBL00061	1	100	85	51	51	50	52	43	42	41	38	43	42	41
RBS04001	2	85	100	52	52	51	53	44	43	42	41	43	43	42
RSY03952	3	51	52	100	99	96	85	45	46	43	42	47	44	42
RTY03814	4	51	52	99	100	96	85	45	46	43	42	47	44	42
REC03333	5	50	51	96	96	100	84	45	45	43	42	46	44	42
RYP01565	6	52	53	85	85	84	100	46	45	44	43	46	45	42
RBPS05419	7	43	44	45	45	45	46	100	83	80	62	60	52	40
RBU11141	8	42	43	46	46	45	45	83	100	73	62	57	51	39
RAC01503	9	41	42	43	43	43	44	80	73	100	60	57	49	38
RPU05135	10	38	41	42	42	42	43	62	62	60	100	54	49	38
RBE02868	11	43	43	47	47	46	46	60	57	57	54	100	54	40
RPA04265	12	42	43	44	44	44	45	52	51	49	49	54	100	41
RCHU01060	13	41	42	42	42	42	42	40	39	38	38	40	41	100

Таблица 2. Процент сходства анализируемых белков для семейства GNTP.

3. РЕЗУЛЬТАТЫ И ВЫВОДЫ

Если для одного белка сервер предсказывал ТМ спирали, которые перекрывались или соприкасались, то их принимали за одну ТМ спираль. Для анализа ТМ серверов была написана программа ТМ.exe на языке СРР. При помощи нее для каждого сервера посчитали величины *qq* и *kfs* предсказанных ТМ спиралей. При этом представителей каждой группы ортологичных генов сравнивали попарно. Из всех четырех серверов самые плохие предсказания дает DAS сервер. Предсказания трех остальных серверов сравнимы между собой. В таблице 3 приведены среднеарифметические значения *qq* и *kfs* для семейства GNTP внутри каждой из трех групп.

	DAS	DAS	TMAP	TMAP	TMHMM	TMHMM	TMPRED	TMPRED
	qq	kfs	qq	kfs	qq	kfs	qq	kfs
40-60	0,365	0,499	0,564	0,784	0,551	0,735	0,482	0,671
60-80	0,469	0,692	0,619	0,866	0,579	0,842	0,567	0,855
80-100	0,827	0,945	0,769	0,916	0,778	0,906	0,828	0,935

Таблица 3. Результаты предсказания различных серверов для семейства GNTP внутри каждой из трех групп. Приведены среднеарифметические значения *qq* и *kfs*.

Следует отметить, что это только предварительные данные. В дальнейшем предполагается провести аналогичный анализ для всех представителей бактериальных вторичных транспортеров. Кроме того, мы надеемся определить, при каком проценте сходства исследуемые сервера дают наиболее точные предсказания.

Мы благодарим за ценные обсуждения А.А. Миронова, В.Ю. Макеева и О.В. Калинину.

Эта работа была частично поддержана грантами от РФФИ (00-15-99362), INTAS (99-1476), ННМИ (55000309) и LICR (CRDF RBO-1268).

СПИСОК ЛИТЕРАТУРЫ

1. Milton H., Saier Jr. A Functional-Phylogenetic System for the Classification of Transport Proteins. *J. Cell Biochem.*, 1999, vol. 75, no. S32, pp. 84–94.
2. Milton H., Saier Jr. A Functional-Phylogenetic Classification System for Transmembrane Solute Transporters. *Microbiol. Mol. Biol. Rev.*, 2000, vol. 64, no. 2, pp. 354–411.

3. Overbeek R., Larsen N., Pusch G.D., D'Souza M., Selkov E.Jr, Kyrpides N., Fonstein M., Maltsev N., Selkov E. WIT: Integrated System for High-Throughput Genome Sequence Analysis and Metabolic Reconstruction. *Nucleic Acids Res.*, 2000, vol. 28, pp. 123–125.
4. Thompson J.D., Higgins D.G., Gibson T.J. CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nucleic Acids Res.*, 1994, vol. 22, pp. 4673–4680.
5. Cserzo M., Wallin E., Simon I., von Heijne G., Elofsson A. Prediction of Transmembrane Alpha-Helices in Prokaryotic Membrane Proteins: the Dense Alignment Surface Method. *Protein Eng.*, 1997, vol. 10, pp. 673–676.
6. Persson B., Argos P. Topology Prediction of Membrane Proteins. *Protein Sci.*, 1996, vol. 5, pp. 363–371.
7. Sonnhammer E.L., von Heijne G., Krogh A. A Hidden Markov Model for Predicting Transmembrane Helices in Protein Sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.*, 1998, vol. 6, pp. 175–182.

Статью представил к публикации член редколлегии В.А. Любецкий