

ВЫЯВЛЕНИЕ НОВЫХ САЙТОВ СВЯЗЫВАНИЯ ТРАНСКРИПЦИОННЫХ ФАКТОРОВ SREBP В ПРОМОТОРНЫХ РАЙОНАХ ГЕНОВ ПОЗВОНОЧНЫХ НА ОСНОВЕ КОМБИНАЦИИ БИОИНФОРМАТИЧЕСКОГО И ЭКСПЕРИМЕНТАЛЬНОГО ПОДХОДОВ

Е.В. Игнатьева^{1,2}, Т.И. Меркулова^{1,2}, Д.Ю. Ощепков^{1,2}, Н.В. Климова¹,
Г.В. Васильев¹, И.И. Турнаев¹, В.Ф. Кобзев¹, Н.А. Колчанов^{1,2}

¹ Учреждение Российской академии наук Институт цитологии и генетики Сибирского
отделения РАН, Новосибирск, Россия, e-mail: eignat@bionet.nsc.ru;

² Новосибирский государственный университет, Новосибирск, Россия

Активность транскрипционных факторов семейства SREBP (Sterol Regulatory Element Binding Protein) зависит от внутриклеточного уровня холестерина. Для формирования целостной картины влияния повышенного уровня холестерина на клеточные функции необходима идентификация возможно большего числа генов, регулируемых факторами SREBP. На основе метода SITECON нами выявлены консервативные конформационные и физико-химические особенности сайтов связывания SREBP (SRE типа) и разработан компьютерный метод распознавания SRE. Выявлены новые потенциальные SRE в регуляторных районах генов липидного метаболизма и клеточного цикла позвоночных. Проведенная затем экспериментальная проверка *in vitro* способности предсказанных сайтов специфически связываться с белком SREBP1-а, который эффективно взаимодействует с SRE, продемонстрировала высокую точность метода. Таким образом, выявлен ряд новых сайтов связывания SREBP, которые с высокой степенью вероятности окажутся функциональными.

Ключевые слова: SREBP, сайты связывания транскрипционных факторов, методы распознавания сайтов, экспериментальная проверка.

Введение

Транскрипция генов эукариот осуществляется при участии транскрипционных факторов, специфически взаимодействующих с короткими (5–20 п.о.) участками ДНК в регуляторных районах генов – сайтами связывания транскрипционных факторов (ССТФ). Каждый фактор обеспечивает координированную активацию или подавление транскрипции большого числа генов, определяя адекватный уровень их экспрессии в ответ на различные сигналы внешней и внутренней среды, а также ее тканеспецифичность.

Транскрипционные факторы семейства SREBP функционируют в составе холестеринного сенсора, обеспечивая регуляцию экспрессии генов в зависимости от уровня холес-

терина в клетке (Gimpl *et al.*, 2002). Семейство SREBP включает белки SREBP-1a, SREBP-1c, SREBP-2. Они принадлежат к классу спираль–петля–спираль/лейциновая застежка (basic Helix–loop–helix/leucine zipper (bHLH-ZIP)) повсеместно экспрессирующихся факторов. N-терминальный домен белков SREBP (~ 480 аминокислотных остатков) соответствует зрелой активной форме факторов и включает ДНК-связывающий основной район (basic region) и димеризационный HLH-Zip район (Wang *et al.*, 1993). Факторы семейства SREBP синтезируются в форме неактивных предшественников preSREBP с молекулярной массой ~ 125 kd и присутствуют в клетке, будучи прикрепленными к мембранам эндоплазматического ретикулума (Yokoyama *et al.*, 1993). Образование активных транскрипционных факторов SREBP

(с мол. массой ~ 68 kd) осуществляется путем расщепления предшественников preSREBP в аппарате Гольджи. Транспорт preSREBP в аппарат Гольджи усиливается при низком уровне холестерина и замедляется при его высоком содержании в клетке.

Факторы SREBP влияют на транскрипцию генов, взаимодействуя с сайтами связывания двух типов: SRE (Sterol Regulatory Element) и E-боксы. Сайты E-боксы типа (ATCACGTGA) соответствуют инвертированному повтору тетрауклеотида TCAC. SRE (ATCACCCCAC) включают несовершенный прямой повтор 5'-ATCAC-3' и 5'-CCCAC-3' (Kim *et al.*, 1995). Показано, что способность к взаимодействию с сайтами SRE типа является уникальной только для SREBP. Напротив, последовательности типа E-боксы являются сайтами связывания как для SREBP, так и для ряда других, родственных им, факторов семейства bHLH, экспрессирующихся в широком круге тканей (USF, TFE3, TFEB и др.) (Kim *et al.*, 1995; Hu *et al.*, 2005). По этой причине только сайты связывания SRE типа способны эффективно опосредовать регуляцию транскрипции генов в ответ на изменение уровня холестерина (Athanikar *et al.*, 1998).

Известно, что факторы SREBP регулируют транскрипцию многих генов системы липидного метаболизма: ферментов пути биосинтеза холестерина и жирных кислот, белков, обеспечивающих транспорт липидов и гидрофобных молекул (ApoA2, CETP, STAR), транскрипционных факторов (SREBP2, PPAR γ). В числе SREBP регулируемых генов имеются также гены из систем клеточного цикла и апоптоза: *CDKN1A*, кодирующий циклин-зависимый ингибитор киназы человека p21 WAF1/CIP1, задерживающий клеточный цикл на стадии G1 (Inoue *et al.*, 2005), и *CASP2*, кодирующий апоптоз-зависимую цистеиновую пептидазу человека (Logette *et al.*, 2005). Последние два примера указывают на то, что регуляторная роль SREBP может выходить за рамки системы липидного метаболизма. Необходимость регуляции экспрессии генов клеточного цикла в зависимости от внутри-клеточного уровня холестерина определяется тем, что рост и деление клеток сопровождается построением клеточной мембраны, в состав которой входит холестерин. Очевидно, что при недостатке холестерина пос-

троение клеточной мембраны будет затруднено, а следовательно, и клеточный цикл должен осуществляться медленнее.

Целью настоящей работы является создание эффективного компьютерного метода распознавания сайтов связывания транскрипционных факторов SREBP и поиск новых сайтов связывания SREBP в генах позвоночных. Для ее достижения на первом этапе на основе данных базы TRRD (Kolchanov *et al.*, 2002) была сформирована выборка экспериментально подтвержденных сайтов связывания SREBP. С использованием компьютерного метода SITECON (Oshchepkov, 2004a) выявлены консервативные конформационные и физико-химические особенности сайтов SRE типа и разработан компьютерный метод распознавания SRE. На следующем этапе с использованием метода SITECON были выявлены новые потенциальные SRE в регуляторных районах генов липидного метаболизма и клеточного цикла. Проведенная затем экспериментальная проверка способности предсказанных сайтов специфически связываться с изоформой SREBP1-а, которая эффективно взаимодействует с SRE (Amemiya-Kudo, 2002), продемонстрировала высокую эффективность метода. Результаты экспериментов *in vitro*, а также наличие сайтов SREBP1 в регуляторных районах генов-ортологов указывают на то, что ряд выявленных нами сайтов с высокой долей вероятности окажутся функциональными.

Материалы и методы

Выборка сайтов связывания SREBP SRE типа (38 последовательностей) была сформирована с использованием информации из базы TRRD (<http://www.bionet.nsc.ru/trrd/>), в рамках которой имеется тематический раздел LM-TRRD, посвященный регуляции транскрипции генов липидного метаболизма (Kolchanov *et al.*, 2002). В состав выборки были включены последовательности сайтов связывания SREBP длиной 38 нуклеотидов, для которых имелись строгие экспериментальные подтверждения связывания с рассматриваемым белком. Поиск потенциальных SRE осуществляли в 5'-фланкирующих областях 46 генов из системы липидного метаболизма из базы LM-TRRD, для которых

на момент составления выборки регуляция факторами SREBP еще не была известна. Для 38 генов в анализ были взяты участки промоторных районов -1000 до $+100$ нуклеотидов относительно старта транскрипции, а для 8 генов – районы $-600/+100$ ¹. Нами также были проанализированы промоторные ($-1000/+100$) районы 20 генов человека из базы данных TRRD, продукты которых участвуют в регуляции клеточного цикла.

Для распознавания потенциальных сайтов использовался метод SITECON, основанный на анализе консервативных конформационных и физико-химических свойств в позициях выборки ССТФ. Подробно метод описан ранее (Oshchepkov, 2004a, b).

Экспериментальное исследование связывающей способности потенциальных SRE осуществлялось методом задержки комплекса ДНК/белок в геле (EMSA) с использованием рекомбинантного SREBP-1а человека (Kolchanov *et al.*, 2007).

Результаты

Консервативные конформационные и физико-химические особенности сайтов SRE. В результате анализа выборки сайтов связывания SRE типа был обнаружен ряд особенностей конформационных и физико-химических свойств ДНК, три из которых представляются наиболее интересными (рис. 1). Оказалось, что в районе сайта ширина большой и ширина малой бороздки значимо консервативны, при этом ширина большой бороздки (рис. 1, а) уменьшена, а ширина малой бороздки (рис. 1, б) увеличена по сравнению со случайными последовательностями. Кроме того, во фланкирующих районах обоих тринуклеотидов САС консенсусной последовательности наблюдаются пониженные значения жесткости к изгибу в сторону большой бороздки (рис. 1, в) по сравнению со случайными последовательностями. Выявленные особенности находятся в согласии с данными рентгеноструктурного анализа (РСА) комплекса гомодимера SREBP1-а с сайтом SRE

¹ Длина 5'-фланкирующих участков генов, взятых в анализ, определялась наличием последовательностей необходимой длины в карточках базы нуклеотидных последовательностей EMBL.

типа (Párraga *et al.*, 1998). Согласно этим данным, структура ДНК сайта в комплексе с белком мало отличается от В-формы, за исключением двух параметров: средняя величина ширины большой бороздки в районе сайта увеличена, а ширина малой бороздки уменьшена относительно значений этих свойств для В-формы ДНК. Кроме того, в центральных 4 парах ДНК (соответствующих нуклеотидам СССС консенсусной последовательности) не наблюдалось изгиба, в то время как по обоим концам сайта наблюдался изгиб в сторону большой бороздки. Как видно, эти данные хорошо совпадают с особенностями свойств ДНК для сайта SRE, предсказанными при использовании технологии SITECON.

Характеристика точности метода распознавания SRE на основе технологии SITECON.

Нами было проведено сравнение точности предсказания SRE для метода SITECON и метода весовых матриц. Метод весовых матриц был запрограммирован на основе работы Gershenson *et al.* (2005). Расчет ошибок производился идентично для обоих подходов на основе выборки SRE, описанной выше. Ошибка I рода рассчитывалась с помощью подхода Jack-knife (Efron, Gong, 1983). Эта процедура включала в себя распознавание одного из сайтов из выборки SRE на основе обучения по неполной выборке, не включающей распознаваемый сайт, такая процедура проводилась последовательно для всех сайтов выборки. Для расчета ошибки II рода обоих методов использовалась одна и та же случайная последовательность длиной 1×10^6 п.о., сгенерированная путем многократного случайного перемешивания нуклеотидов обучающей выборки. Из графика рис. 2 видно, что метод SITECON и метод весовых матриц показывают сходное качество распознавания. При этом для значений ошибки II рода, пригодных для геномных исследований (менее чем 10^{-3}), до критических значений ошибки I рода (более 90 %) качество распознавания методом SITECON даже незначительно превосходит качество метода весовых матриц. Для выбранного нами порога уровня конформационного сходства ошибка I рода метода SITECON составляет 0,53. При заданном уровне ошибки I рода (0,53) метод весовых матриц дает ошибку II рода $9,90 \times 10^{-4}$. В то же время метод SITECON позволяет получить меньшую ошибку: $7,55 \times 10^{-4}$.

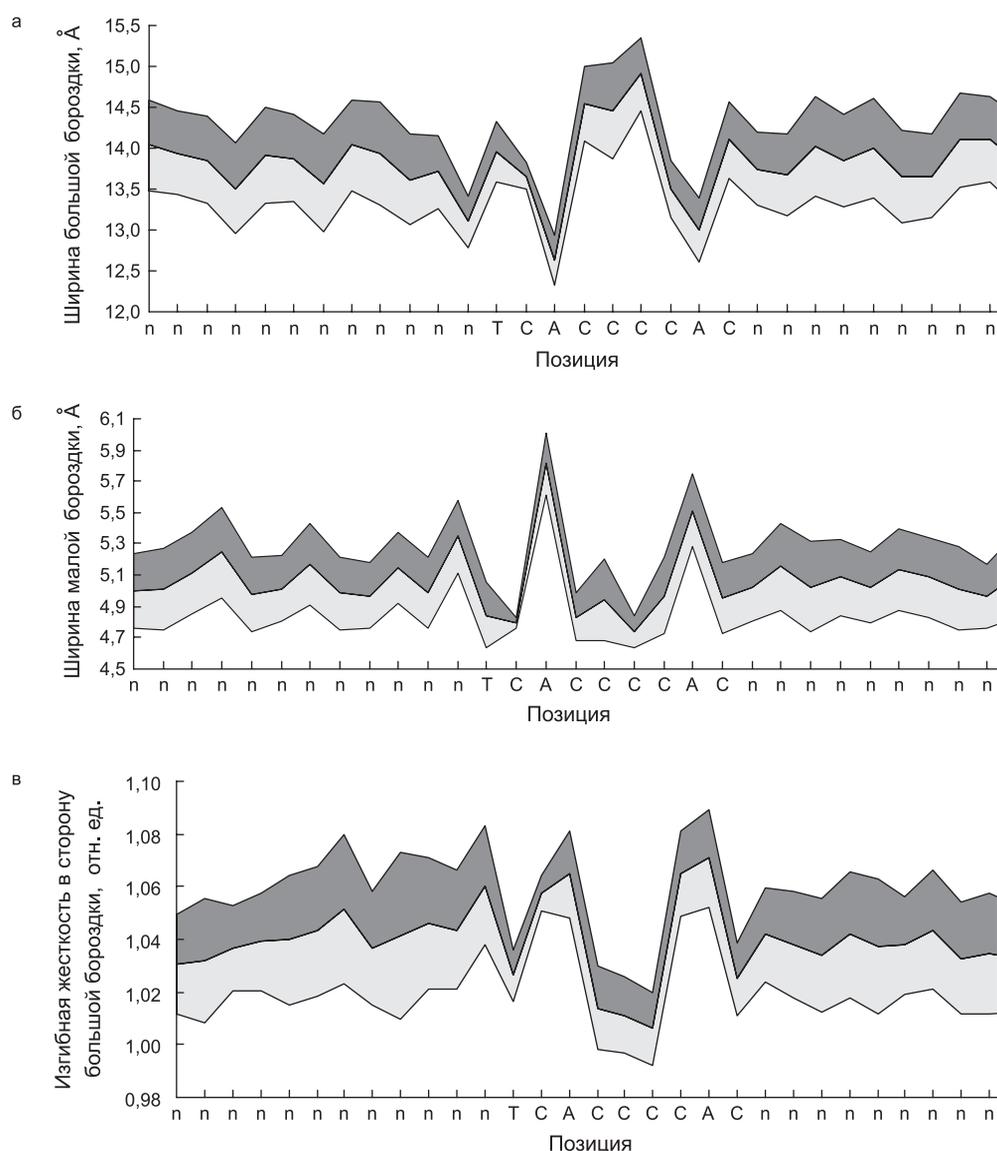


Рис. 1. Значения конформационных и физико-химических свойств ДНК и их консервативность, рассчитанные для выборки выровненных сайтов SRE типа.

а – ширина большой бороздки; б – ширина малой бороздки; в – изгибная жесткость ДНК в сторону большой бороздки. Центральная кривая соответствует среднему значению свойства для каждой из позиций выборки; верхняя кривая соответствует среднему значению плюс значение стандартного отклонения для позиции; нижняя кривая соответствует среднему значению минус значение стандартного отклонения для позиции. Участок, соответствующий сайту SRE, обозначен по оси OX его консенсусной последовательностью TCACCCAC.

Выявление новых SRE в регуляторных районах генов позвоночных и экспериментальное исследование связывающей способности новых SRE. При анализе 5'-фланкирующих областей генов позвоночных методом SITECON выявлены потенциальные SRE. Сайты выявлены как на прямой, так и на обратной цепи ДНК и располагаются на разном удалении

от старта транскрипции. Часть потенциальных SRE, выявленных в генах липидного метаболизма, представлена в табл. 1. Выявлены также SRE в 15 генах из системы регуляции клеточного цикла человека: *CDKN2A*, *UNG*, *RBL2*, *GADD45G*, *CCNA1*, *TGFB1*, *BIRC5*, *CDKN1A*, *CCND1*, *MYB*, *JUN*, *RB1*, *CCNE1*, *E2F1*, *GADD45A*. Гены, содержащие потенциальные

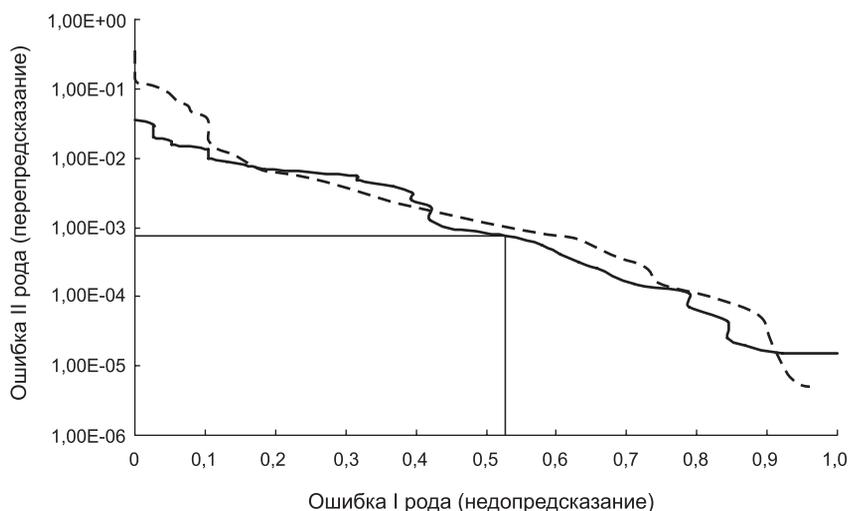


Рис. 2. Сравнение ошибок I и II рода для методов SITECON (сплошная линия) и метода весовых матриц (пунктир).

Линиями обозначены значения ошибок I и II рода ($0,53$ и $7,55 \times 10^{-4}$ соответственно) для порога, на котором проводилось предсказание методом SITECON.

SRE, кодируют белки с различными функциями: транскрипционные факторы (LXRA, E2F, JUN), ферменты (FAS, CPT2, LPL), транспортные белки (MSR, IBABP) и др.

Выборочной экспериментальной проверке *in vitro* подверглись 30 сайтов, распознанных в генах липидного метаболизма (табл. 1). Типичные результаты, полученные методом задержки

в геле, представлены на рис. 3. На дорожке 2, соответствующей ДНК-белковому комплексу, включавшему консенсусную последовательность SRE, выявлено 2 полосы задержки, что согласуется с данными других авторов (Moop, 2000). ДНК-белковые комплексы с такой же подвижностью выявлены и на дорожках 4–14 и 17–26.

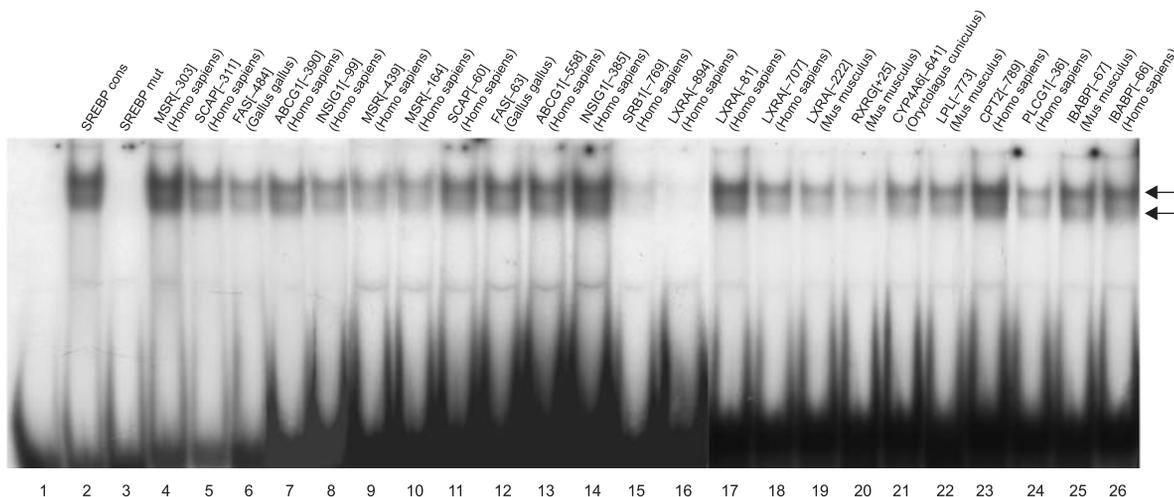


Рис. 3. Экспериментальное подтверждение SRE, предсказанных методом SITECON.

Инкубация олигонуклеотидной пробы (1 нг), меченной ^{32}P , в отсутствие белка (1-я дорожка) либо с белком SREBP-1a (0,5 мкг) (дорожки 2–26). SREBP cons – олигонуклеотид, включавший консенсусную последовательность, SREBPmut – олигонуклеотид, включавший консенсусную последовательность с мутацией. Сверху указаны названия генов, виды организмов и позиции сайтов в генах относительно старта транскрипции. Стрелками показаны комплексы ДНК с белком SREBP-1a, доказывающие связывающую способность предсказанных сайтов.

Таблица 1

Потенциальные SRE, предсказанные методом SITECON в промоторах генов системы липидного метаболизма, и результаты их экспериментальной проверки

№	Название гена (вид)	Позиция*	Ориентация	R**	Последовательность ССТФ	Экспериментальное подтверждение
а) потенциальные SRE, имеющие уровень конформационного сходства более 73 %						
1	<i>MSR</i> (человек)	-439	+	0,750	GGACCATGAGGTCAGGGGATCAAGACCATC	+
2	<i>MSR</i> (человек)	-303	+	0,778	GAGATCACACCACTGCACTCCACCCTGGTG	+
3	<i>MSR</i> (человек)	-164	-	0,796	AGATAACTCACTCTTACCCCCATTTAGGCC	+
4	<i>SCAP</i> (человек)	-311	+	0,777	TCTCCCCCGTACACACAGGGGCCCGCGGA	+
5	<i>FAS</i> (цыпленок)	-484	+	0,756	GCCTAGCAACACCCTCACGGGCTCCGCATT	+
6	<i>FAS</i> (цыпленок)	-63	+	0,749	GGAGTAAACCGTCAGCCCATGTGGTGGCCG	+
7	<i>ABCG1</i> (человек)	-558	-	0,732	CCCTTCCCGCCCCACCCACCCCGACGGTGT	+
8	<i>INSIG1</i> (человек)	-385	+	0,736	GGCCCGAAAAAATCACCCAAGGATGCAAGC	+
9	<i>INSIG1</i> (человек)	-99	+	0,769	GCGGGGCAAGCTCAGGCCACGCCCTGGGC	+
10	<i>LXRA</i> (человек)	-81	-	0,780	CTGCCGAGGGTTCAGGCCACCCCTCCTTCC	+
11	<i>CYP4A6</i> (кролик)	-641	-	0,783	ACAAGCATTTCACCCCACTGCCCTCAACT	+
12	<i>LPL</i> (мышь)	-773	+	0,805	CCACACTCTAGCACCCATATCCCCTCGC	+
13	<i>CPT2</i> (человек)	-789	+	0,731	CAATCCGAGGTCACCTAAGATAACAAA	+
14	<i>IBABP</i> (мышь)	-67	-	0,863	AGCCCCAAAGTCACCCCACTTCTCAGTCCT	+
15	<i>IBABP</i> (человек)	-66	-	0,881	CAGCCCCAAGTCACCCCACTTCTTCTCCC	+
б) потенциальные SRE, имеющие уровень конформационного сходства менее 73 %						
16	<i>LXRA</i> (мышь)	-222	+	0,729	TCTCTGTCCCCTCAGGGCATCCACACACAA	+
17	<i>PLCG1</i> (человек)	-36	+	0,728	GGGGTGGGGCCTCAGGGCAGCCCCGCCCG	+
18	<i>RXRG</i> (мышь)	+25	+	0,727	GGAAGAGTCGCTCACAGCAGCCCAACCCAG	+
19	<i>SRBI</i> (человек)	-769	+	0,725	GGGGGCCCGCCTCACGGGTTGGCATCCCAG	-
20	<i>MSR</i> (человек)	-323	-	0,704	AGTGGTGTGATCTCAGCTCACTGCAACCTC	+
21	<i>LXRA</i> (человек)	-707	-	0,717	CAACTCAGGTGGCAGCCCAGGATGGACCTG	+
22	<i>LXRA</i> (человек)	-894	-	0,717	CCAGCCCTGTTTCAGGGGTGTTGGGGGAGA	-
23	<i>FAS</i> (цыпленок)	-429	-	0,716	AGTGCCGGGCCGGCACCCAACGCGGGGGCG	-
24	<i>CAVI</i> (мышь)	-239	-	0,715	GACCCACAACCTCACCTGCTCCGTGGTCC	-
25	<i>SIP</i> (человек)	-101	+	0,713	GAACACGCCTGGGCACTCCATTCGGGGCTG	-
26	<i>ABCG1</i> (человек)	-390	+	0,712	CCCCGCAGCCTCACCCGTGCTGGCCCAGCC	+
27	<i>INSIG1</i> (человек)	-47	-	0,712	GCGGGGCCCGCATCACGTGAGCCTGCCCGG	-
28	<i>FAS</i> (цыпленок)	-129	+	0,705	CAGGGGCTGGCATCACCCGGGGCCGGCTC	+
29	<i>CAVI</i> (мышь)	-501	-	0,699	TGCCAAGACCTGTCACCCCCACCCCATGC	-
30	<i>SCAP</i> (человек)	-60	+	0,694	CAGGCGCCCTGAGCACCCGCGTGACGGTAC	+

* Позиция указана относительно старта транскрипции; ** уровень конформационного сходства с известными сайтами связывания SREBP, оцененный методом SITECON.

Уровень конформационного сходства с ССТФ из обучающей выборки варьировал от 69 до 88,1 %. Способность взаимодействовать с белком SREBP-1a подтверждена для 23 из 30 сайтов. При этом все 15 сайтов, обладающих уровнем конформационного сходства более чем 73 %, оказались способными к связыванию с белком SREBP-1a *in vitro* (табл. 1, а). Только 8 из 15 потенциальных сайтов с уровнем конформационного сходства менее чем 73 % оказались способными к связыванию с белком SREBP-1a *in vitro* (табл. 1, б). Следует отметить тот факт, что 4 из 5 сайтов с уровнем конформационного сходства от 72 до 73 % оказались способными к связыванию с белком SREBP-1a *in vitro*. Этот результат подтверждает тот факт, что показатель «уровень конформационного сходства» адекватно отражает потенциальную способность предсказанных сайтов к взаимодействию с факторами SREBPs.

Обсуждение

Во всех известных к настоящему времени работах выявление потенциальных сайтов связывания SREBP осуществляется на основе консенсуса или методом весовых матриц с использованием данных из базы TRANSFAC. Однако достоверных данных об эффективности этих подходов до настоящего времени нет, поскольку для их получения необходима массовая экспериментальная проверка связывающей способности потенциальных сайтов. Настоящая работа и была направлена на решение подобной задачи.

На первом этапе сформирована выборка геномных последовательностей SRE, превышающая по объему представленные в литературе компиляции. Для распознавания SRE был разработан метод SITECON, в основе которого лежат анализ и выявление консервативных контекстно-зависимых конформационных и физико-химических свойств спирали ДНК в районе сайтов связывания (Oshcherkov *et al.*, 2004a). Данный подход с успехом применялся ранее для анализа и распознавания сайтов связывания E2F и SF1 (Oshcherkov *et al.*, 2004b; Игнатъева и др., 2007). Настоящее исследование вновь подтвердило эффективность технологии SITECON. С помощью метода SITECON для сайтов SRE был выявлен ряд свойств ДНК, которые со-

гласуются с данными рентгеноструктурного анализа ДНК-белкового комплекса этого сайта с фактором SREBP-1a (Párraga *et al.*, 1998). Экспериментальная проверка потенциальных сайтов, предсказанных методом SITECON, продемонстрировала высокую эффективность метода. Все 15 сайтов с уровнем конформационного сходства выше 73 % обладали высокой связывающей активностью по отношению к белку SREBP-1a в условиях *in vitro*.

В исследовании связывающей способности потенциальных сайтов с уровнем конформационного сходства ниже выбранного нами порога 73 % (от 69 до 72,7 %) выявлено (табл. 1, б), что только 53 % потенциальных сайтов этой группы оказались способными к связыванию с белком SREBP-1a *in vitro*. Таким образом, преимуществом метода SITECON является не только новое знание о биологической модели взаимодействия ДНК–белок, но и то, что показатель «уровень конформационного сходства» адекватно отражает потенциальную способность предсказанных сайтов к взаимодействию с факторами SREBP. Наличие связывающей активности всех 15 протестированных сайтов с уровнем конформационного сходства более 73 % подтверждает правильность выбранного порога распознавания для выявления высокодостоверных SRE.

Для подтверждения функциональности выявленных нами новых сайтов связывания SREBP безусловно необходимы дальнейшие эксперименты. Однако уже имеющиеся данные свидетельствуют о высокой вероятности того, что они являются функциональными. Например, роль SREBP в регуляции транскрипции генов *ABCG1* человека и *FAS* цыпленка была подтверждена экспериментально другими авторами (Gondret *et al.*, 2001; Ecker *et al.*, 2007). Кроме того, поскольку показано, что SREBP участвует в регуляции транскрипции *LPL* человека (Schoonjans *et al.*, 2000), выявленный нами сайт (в позиции –773) в одноименном гене мыши, по всей вероятности, также является функциональным. Для остальных генов, в которых предсказанные SRE продемонстрировали способность взаимодействовать с SREBP-1a *in vitro*, регуляция факторами SREBP также достаточно вероятна вследствие того, что кодируемые этими генами белки являются важными элементами системы липидного метаболизма.

Нами также выявлены сайты связывания SREBP в 15 генах клеточного цикла человека, для ряда из которых (*E2F*, *JUN*, *GADD45A* и др.) при анализе ортологичных генов у мыши и крысы также были выявлены сайты SRE. Мы считаем выявленные SRE в генах клеточного цикла чрезвычайно перспективными для дальнейшего изучения, поскольку участие SREBP в регуляции генов клеточного цикла имеет достаточно ясный биологический смысл.

Проведенное нами исследование позволило разработать метод распознавания SRE на базе технологии SITECON, идентифицировать новые потенциальные SRE в генах позвоночных, оценить эффективность метода и адекватность используемого уровня конформационного сходства на основе выборочной проверки предсказанных сайтов.

Благодарности

Работа поддержана грантом РФФИ № 08-04-01214, проектом № 8 программы Президиума РАН № 22 «Молекулярная и клеточная биология», проектом № 29 программы Президиума РАН № 22 «Биологическое разнообразие», подпрограммой «Эволюция гео-биологических систем» программы Президиума РАН «Происхождение и эволюция биосферы», интеграционными проектами СО РАН (№ 26, 113, 119, 115, 18.13), а также НШ-2447.2008.4. Авторы выносят благодарность Т.Ф. Осборне (University of California, Irvine, CA) за предоставление плазмиды pCMV5, И.В. Лоховой за библиографическую поддержку, А.Л. Проскура за участие в аннотировании данных в базе TRRD.

Литература

- Игнатъева Е.В., Климова Н.В., Ощепков Д.Ю. и др. Поиск новых сайтов связывания транскрипционного фактора SF1 методом SITECON: экспериментальная проверка и анализ регуляторных районов генов-ортологов // Докл. РАН. 2007. Т. 415. № 1. С. 120–124.
- Amemiya-Kudo M., Shimano H., Nasty A.H. *et al.* Transcriptional activities of nuclear SREBP-1a, -1c, and -2 to different target promoters of lipogenic and cholesterologenic genes // *J. Lipid Res.* 2002. V. 43. № 8. P. 1220–1235.
- Athanikar J.N., Osborne T.F. Specificity in cholesterol regulation of gene expression by coevolution of sterol regulatory DNA element and its binding protein // *Proc. Natl Acad. Sci. USA.* 1998. V. 95. № 9. P. 4935–4940.
- Ecker J., Langmann T., Moehle C., Schmitz G. Isomer specific effects of conjugated linoleic acid on macrophage ABCG1 transcription by a SREBP-1c dependent mechanism // *Biochem. Biophys. Res. Commun.* 2007. V. 352. № 3. P. 805–811.
- Efron B., Gong G. A leisure look at the bootstrap, the jackknife and cross-validation // *Am. Stat.* 1983. V. 37. P. 36–48.
- Gershenzon N.I., Stormo G.D., Ioshikhes I.P. Computational technique for improvement of the position-weight matrices for the DNA/protein binding sites // *Nucl. Acids Res.* 2005. V. 33. № 7. P. 2290–2301.
- Gimpl G., Burger K., Fahrenholz F. A closer look at the cholesterol sensor // *Trends Biochem. Sci.* 2002. V. 27. № 12. P. 596–569.
- Gondret F., Ferré P., Dugail I. ADD-1/SREBP-1 is a major determinant of tissue differential lipogenic capacity in mammalian and avian species // *J. Lipid Res.* 2001. V. 42. № 1. P. 106–113.
- Hu J., Banerjee A., Goss D.J. Assembly of b/HLH/z proteins c-Myc, Max, and Mad1 with cognate DNA: importance of protein-protein and protein-DNA interactions // *Biochemistry.* 2005. V. 44. № 35. P. 11855–11863.
- Inoue N., Shimano H., Nakakuki M. *et al.* Lipid synthetic transcription factor SREBP-1a activates p21WAF1/CIP1, a universal cyclin-dependent kinase inhibitor // *Mol. Cell. Biol.* 2005. V. 25. № 20. P. 8938–8947.
- Kim J.B., Spotts G.D., Halvorsen Y.D. *et al.* Dual DNA binding specificity of ADD1/SREBP1 controlled by a single amino acid in the basic helix-loop-helix domain // *Mol. Cell. Biol.* 1995. V. 15. P. 2582–2588.
- Kolchanov N.A., Ignatieva E.V., Ananko E.A. *et al.* Transcription Regulatory Regions Database (TRRD): its status in 2002 // *Nucl. Acids Res.* 2002. V. 30. № 1. P. 312–317.
- Kolchanov N.A., Merkulova T.I., Ignatieva E.V. *et al.* Combined experimental and computational approaches to study the regulatory elements in eukaryotic genes // *Briefings in Bioinformatics.* 2007. V. 8. № 4. P. 266–274.
- Logette E., Solary E., Corcos L. Identification of a functional DNA binding site for the SREBP-1c transcription factor in the first intron of the human caspase-2 gene // *Biochim. Biophys. Acta.* 2005. V. 1738. № 1/3. P. 1–5.
- Moon Y.A., Lee J.J., Park S.W. *et al.* The roles of sterol regulatory element-binding proteins in the transactivation of the rat ATP citrate-lyase promoter // *J. Biol. Chem.* 2000. V. 275. № 39.

- P. 30280–30286.
- Oshchepkov D.Y., Vityaev E.E., Grigorovich D.A. *et al.* SITECON: a tool for detecting conservative conformational and physicochemical properties in transcription factor binding site alignments and for site recognition // Nucl. Acids Res. 2004a. V. 32. P. W208–W212.
- Oshchepkov D. Yu., Turnaev I.I., Pozdnyakov M.A. *et al.* SITECON—a tool for analysis of DNA physicochemical and conformational properties: E2F/DP transcription factor binding site analysis and recognition // Bioinformatics of Genome Regulation and Structure / Ed. N. Kolchanov, R. Hofstaedt. Boston/Dordrecht/London: Kluwer Academic Publ., 2004b. P. 93–102.
- Párraga A., Bellolell L., Ferré-D’Amaré A.R., Burley S.K. Co-crystal structure of sterol regulatory element binding protein 1a at 2.3 Å resolution // Structure. 1998. V. 6. № 5. P. 661–672.
- Schoonjans K., Gelman L., Haby C. *et al.* Induction of LPL gene expression by sterols is mediated by a sterol regulatory element and is independent of the presence of multiple E boxes // J. Mol. Biol. 2000. V. 304. № 3. P. 323–334.
- Wang X., Briggs M.R., Hua X. *et al.* Nuclear protein that binds sterol regulatory element of low density lipoprotein receptor promoter // J. Biol. Chem. 1993. V. 268. P. 14497–14504.
- Yokoyama C., Wang X., Briggs M.R. *et al.* SREBP-1, a basic-helix-loop-helix-leucine zipper protein that controls transcription of the low density lipoprotein receptor gene // Cell. 1993. V. 75. № 1. P. 187–197.

**PREDICTION OF NEW SREBP BINDING SITES
IN THE PROMOTER REGIONS OF VERTEBRATE GENES
USING A COMBINATION OF A BIOINFORMATIC
AND AN EXPERIMENTAL APPROACH**

**E.V. Ignatieva^{1,2}, T.I. Merkulova^{1,2}, D.Y. Oshchepkov^{1,2}, N.V. Klimova¹, G.V. Vasiliev¹,
I.I. Turnaev¹, V.F. Kobzev¹, N.A. Kolchanov^{1,2}**

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

²Novosibirsk State University, Novosibirsk, Russia

Summary

The activity of transcription factors in the family of sterol regulatory element binding proteins (SREBPs) is regulated by cellular cholesterol content. To have better insight into the implications of elevated cholesterol for cellular functions, it is critical to identify as many genes that SREBPs regulate, as possible. Using SITECON method, we have determined the conserved conformational, physical and chemical properties of SREs, which are SREBP binding sites, and developed a computer method for SRE prediction. New potential SREs in the regulatory regions of the vertebrate genes involved in control of lipid metabolism and cell cycle have been predicted. An *in vitro* validation of the ability of potential SREs to specifically bind to SREBP1-a – an isoform, which binds efficiently to SREs – showed a high accuracy of SITECON-based recognition. In summary, new SREBP binding sites have been found, which will most likely be functional.