

УДК 577.218

МЕТОД EMSA–SELEX–SEQ ДЛЯ АНАЛИЗА ПОСЛЕДОВАТЕЛЬНОСТЕЙ САЙТОВ СВЯЗЫВАНИЯ В КОМПЛЕКСАХ ДНК–БЕЛОК

© 2024 г. И. Б. Назаров^{1,*}, М. Н. Гордеев¹, А. А. Кузьмин¹, Д. С. Зилов¹, Е. В. Потапенко²,
А. Н. Томилин^{1,**}

¹Институт цитологии РАН, Санкт-Петербург, 194064, Россия

²Институт эволюционного развития, отдел эволюции и биологии окружающей среды Университета Хайфы, Хайфа,
3498838, Израиль

*E-mail: i.nazarov@incras.ru

**E-mail: a.tomilin@incras.ru

Поступила в редакцию 06.12.2023

После доработки 08.12.2023

Принята к публикации 08.12.2023

Белок BOB1 (OBF1, OCA-B) представляет собой транскрипционный коактиватор двух POU-доменных белков: экспрессирующегося во всех клетках OBT1 и лимфоид-специфического OBT2. Взаимодействие BOB1 с OBT1/2 играет важную роль в регуляции иммунных ответов как в физиологическом, так и в патологическом контексте. Известно, что BOB1 образует тройной комплекс с OBT1/2 связанных с ДНК в мономерных и определенных димерных конфигурациях, меняя сиквенс-специфичность такого связывания. Для анализа последовательностей ДНК из этих комплексов в настоящей работе нами предложен метод EMSA-SELEX-seq, основанный на разделении OBT/BOB1 комплексов различного состава в неденатурирующем полиакриламидном геле (EMSA) с последующим выделением и амплификацией содержащихся в них олигонуклотидов (SELEX). На основании нескольких раундов такого обогащения с последующим NGS-секвенированием и биоинформатическим анализом были определены ДНК-последовательности и подтверждена релевантность данного подхода. Таким образом, предложенный метод EMSA-SELEX-seq позволяет проводить анализ последовательностей ДНК в сложных ДНК-белковых комплексах с различающейся мерностью его белковых компонентов.

Ключевые слова: EMSA, SELEX, NGS секвенирование, OBT1, OBT2, BOB1, OBF1, OCA-B

Принятые сокращения: ТФ – транскрипционные факторы; EMSA – метод сдвига электрофоретической подвижности (Electrophoretic Mobility Shift Assay); CAP – последовательная аффинная очистка (Consecutive Affinity Purification); SELEX – систематическая эволюция лигандов путем экспоненциального обогащения (Systematic Evolution of Ligands by Exponential Enrichment).

DOI: 10.31857/S0041377124030052, **EDN:** PEKAZN

Определение специфических последовательностей ДНК, связывающих одиночные транскрипционные факторы (ТФ) или комплексы таковых, является актуальной задачей многочисленных современных исследований. Большинство ТФ эукариот обладает довольно невысокой специфичностью при связывании с ДНК, допуская вариативность нуклеотидов во многих позициях сайтов узнавания (Weirauch et al., 2013; Kibet and Machanick, 2015; Meers et al., 2019). Более того, модуляция узнавания последовательностей при взаимодействии двух и более ТФ с ДНК может быть одним из важнейших комбинаторных механизмов регуляции транскрипции. Экспериментально было продемонстрировано, что специфичность узнавания последовательностей в ДНК в составе

димерных комплексов ТФ заметно отличается от таковой в случае отдельных ТФ (Jolma et al., 2015). Значительная часть таких данных была получена с использованием метода CAP-SELEX (consecutive affinity-purification systematic evolution of ligands by exponential enrichment) и последующего NGS-секвенирования. Однако не все ТФ, коактиваторы и другие ДНК-связывающие белки могут быть проанализированы с помощью данного подхода вследствие недостаточной аффинности одного из компонентов. Кроме того, исследование ДНК-последовательностей, специфически связываемых белками в мономерной, димерной, или более сложных конфигурациях (Luo et al., 1992; Gstaiger et al., 1995, 1996; Luo, Roeder, 1995; Strubin et al., 1995; Tomilin et al., 2000; Remenyi et

al., 2001), часто затруднено невозможностью разделения этих комплексов с помощью методов простой аффинной очистки, таких как CAP-SELEX. В настоящей работе мы показали, что указанные препятствия могут быть преодолены с помощью сочетания метода EMSA (electrophoretic mobility shift assay) — основного метода идентификации ДНК-белок комплексов *in vitro* — с процедурой SELEX и высокопроизводительным секвенированием. Мы использовали данный метод, названный EMSA-SELEX-seq, для оценки способности коактиватора BOB1 влиять на специфичность распознавания ДНК POU-доменными белками OCT1 и OCT2 в мономерной и димерной конфигурациях.

МАТЕРИАЛ И МЕТОДИКА

Плазмидные конструкторы, экспрессия и очистка белков. Для клонирования были использованы последовательности, кодирующие POU-домены OCT1 (POU1) и OCT2 (POU2), а также полноразмерного BOB1. Для получения необходимой кДНК методом обратной транскрипции была выделена мРНК из лейкоцитов человека. Для экспрессии исследуемых белков были созданы следующие плазмидные конструкторы на основе вектора pET28a (Novagen, США): 1) 1t-POU1: 6xHis tag—thrombin site—POU1, 2) 2t-POU1: 6xHis tag—thrombin site—(20 aa + POU1 + 20 aa)—TEV site—SBP tag, 3) 2t-POU2: 6xHis tag—thrombin site—(20 aa + POU2 + 30 aa)—TEV site—SBP tag, 4) 1t-BOB1: 6xHis tag—thrombin site—BOB1, 5) 2t-BOB1: 6xHis tag—thrombin site—BOB1—TEV site—3x Flag tag. Цифрами перед аббревиатурой aa (amino acids) обозначено число примыкающих к POU-доменам аминокислот, включенных в указанные конструкторы. Корректность клонирования верифицировали секвенированием. Экспрессию белка индуцировали 1 mM IPTG в клетках *E. coli* BL21 (BL21(DE3) pLysS (Novagen, США) в течение 12—18 ч при температуре 25 °C. Белки очищали в денатурирующих условиях и ренатурировали с помощью медленного пошагового диализа. Для этого клетки собирали центрифугированием и лизировали в буфере A1 (8 M мочевины, 0.1 M натрий-фосфатный буфер, 0.01 M трис-Cl, pH 8.0, 5 mM β-меркаптоэтанол). Лизат клеток наносили на Ni-NTA His Bind Resin (Novagen, США), промывали буфером A2 (8 M мочевины, 0.1 M натрий-фосфатный буфер, 0.01 M трис-Cl pH 6.3, 5 mM β-меркаптоэтанол) и элюировали буфером A3 (8 M мочевины, 0.1 M натрий-фосфатный буфер, 0.01 M трис-Cl pH 4.5, 5 mM β-меркаптоэтанол).

Образцы белков затем подвергали четырехэтапному диализу против буферов с двукратным уменьшением концентрации мочевины и постепенным увеличением pH на каждом этапе. Конечную стадию диализа проводили против PBS, содержащего 5 mM β-меркаптоэтанол, в течение 12 ч. Для получения гомогенных и полноразмерных форм белков, POU-пробы дополнительно очищали с помощью Pierce Streptavidin Plus UltraLink Resin, а BOB1 — на AN-TI-FLAG M2 Affinity Gel. Условия очистки соответствовали рекомендациям изготовителей аффинных сорбентов. POU1(2) элюировали буфером, содержащим 10 mM биотина, а BOB1 — кислым буфером.

EMSA. 25 нг ДНК смешивали с 30 нг белка POU-белков и, где необходимо, 28 нг белка BOB1 в 10 мкл связывающего буфера: 20 mM Трис-НСl, pH 7.6, содержащий 10 % глицерина, 0.14 M NaCl, 0.5 mM EDTA, 0.05 % Triton X-100, 0.1 мг/мл бычьего сывороточного альбумина, 0.1 мг/мл поли(dI-dC) (Amersham, Великобритания), 5 mM ДТТ, и инкубировали в течение 20 мин при комнатной температуре. Затем, смесь наносили на 5 %-ный полиакриламидный гель для электрофоретического разделения. Для визуализации ДНК и ДНК-белковых комплексов использовали флуоресцентно меченые ДНК пробы.

CAP-SELEX. Метод заключается в поэтапном обогащении случайных ДНК, последовательностями, которые аффинно взаимодействуют с исследуемыми белками. В первую очередь, нами была получена ДНК, содержащая случайную 35-нуклеотидную последовательность, фланкированную олигонуклеотидами с известной последовательностью, и предназначенных, как амплификации, так и для пробоподготовки в процессе NGS секвенирования. Для этого синтезирован олигонуклеотид 5'-ACGACGCTCTTCCGATCT(35N)GATCGGAAGAGCACACGTC-3' (Евроген, Москва), общей длиной 72 нуклеотида. Одноцепочечную ДНК переводили в двуцепочечную форму и амплифицировали с помощью ПЦП реакции используя Phusion High-Fidelity DNA Polymerase (Thermo Scientific). Реакцию связывания полученной ДНК для SELEX с исследуемыми белками проводили аналогично процедуре EMSA, за исключением того, что количество ДНК составляло 300 нг. После связывания комплексы выделяли с помощью 5 мкл Pierce Streptavidin Plus UltraLink Resin, используя взаимодействие с SBP-тегом рекомбинантных POU-белков. Затем, осадок тщательно отмывали, а элюцию комплексов проводили в связывающем буфере с 10 mM биотином. Следующий шаг представлял

собой повторное выделение комплексов, используя взаимодействие Flag-тега белка BOB1 с ANTI-FLAG M2 Affinity Gel. Для этого 5 мкл геля инкубировали с элюатом после первого выделения, гель отмывали, а связанную ДНК получали разрушением комплексов при нагревании проб до 95 °С в течение 20 мин. ДНК амплифицировали с помощью ПЦР.

EMSA-SELEX. ДНК (300 нг) для SELEX инкубировали для связывания с исследуемыми белковыми факторами в объеме 20 мкл при условиях, аналогичных в методе EMSA. После этого образцы разделяли с помощью электрофореза в 6 %-ном нативном полиакриламидном геле, приготовленном на 0.5-кратном буфере TBE, в течение 50 мин при комнатной температуре. Чтобы избежать перекрестного загрязнения образцов, их наносили через одну пустую лунку. Затем гель разделяли на фрагменты, выделяли из них ДНК, амплифицировали её с помощью ПЦР, осаждали этанолом и использовали в следующем цикле обогащения. После четырех раундов EMSA-SELEX образцы секвенировали с помощью Illumina Solexa HiSeq1500 в компании «Геноаналитика» (Москва). Выявление мотивов было выполнено с помощью онлайн-инструмента MEME (Bailey, Elkan, 1994).

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В результате экспрессии и очистки белков с использованием аффинных сорбентов были получены белки, содержащие только N-концевой His-тег (1t-POU1 и 1t-BOB1) и белки с двумя тегам:

N-концевой His-тег и C-концевой SBP-тег (2t-POU1 и 2t-POU2), N-концевой His-тег и C-концевой FLAG-тег (2t-BOB1). Взаимодействие полученных белков с ДНК было верифицировано с помощью EMSA. Типичные результаты EMSA представлены на рис. 1.

Первоначально для оценки способности коактиватора BOB1 влиять на специфичность распознавания ДНК POU-доменными белками OCT1 и OCT2 в мономерной и димерной конфигурациях нами была осуществлена попытка аффинного выделения ДНК-белковых комплексов с помощью метода CAP-SELEX (Jolma et al., 2013, 2015). Для этого были использованы дважды тегированные POU1, POU2 и BOB1. При выделении ДНК из комплексов ДНК/POU1(2)/POU(2) или ДНК/POU1(2)/POU1(2)/BOB1 позитивный результат наблюдается только в случае использования очистки за SBP-тега белков POU1 и POU2 (рис. 2а). Однако, при выделении ДНК из ДНК/POU1(2)/POU1(2)/BOB1 комплекса наблюдался только неспецифический сигнал (рис. 2б). По-видимому, аффинность взаимодействия белка BOB1 не является достаточной для выделения такого комплекса. Следует отметить, что при этом аффинная очистка Flag-тегированного белка BOB1 с использованием ANTI-FLAG M2 Affinity Gel прошла вполне успешно. Попытки изменения условий выделения ДНК из ДНК/POU1(2)/POU1(2)/BOB1 комплекса позитивного результата не принесли. В итоге нами был предложен альтернативный способ получения ДНК из комплексов ДНК/POU1(2)/POU1(2) и ДНК/POU1(2)/POU1(2)/BOB1, который дал положительный результат. Этот способ заключал-

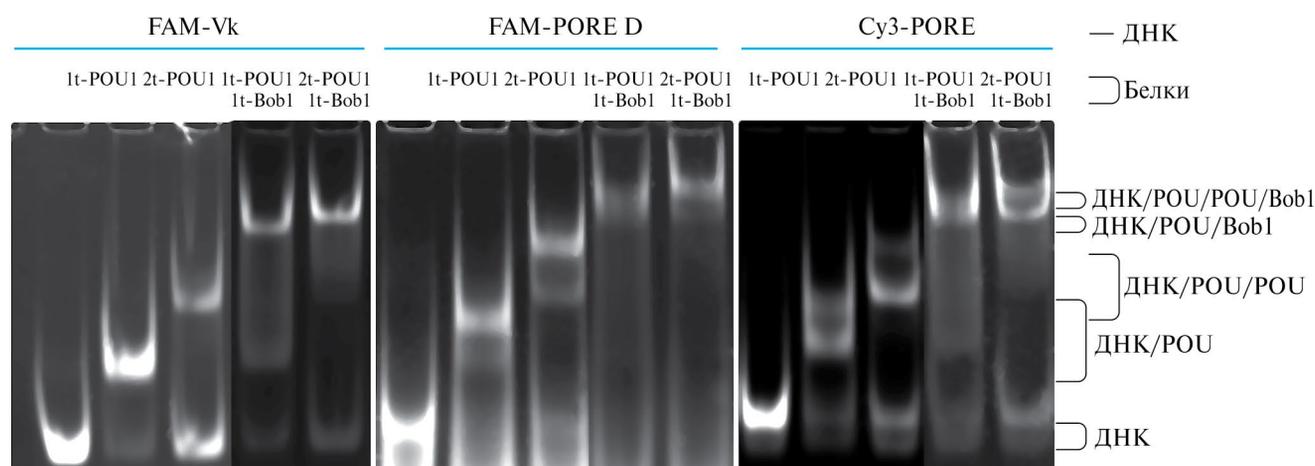


Рис. 1. Гель-ретардация (EMSA) препаратов указанных тегированных POU-белков и BOB1. В качестве ДНК-проб были использованы двуцепочечные флуоресцирующие олигонуклеотиды, которые связывают OCT-факторы в мономерной (Vk) и димерной (PORE и PORE^D) конфигурациях (Botquin et al., 1998; Tomilin et al., 2000). Для визуализации, олигонуклеотиды были мечены флуоресцеином (FAM) или Cy3.

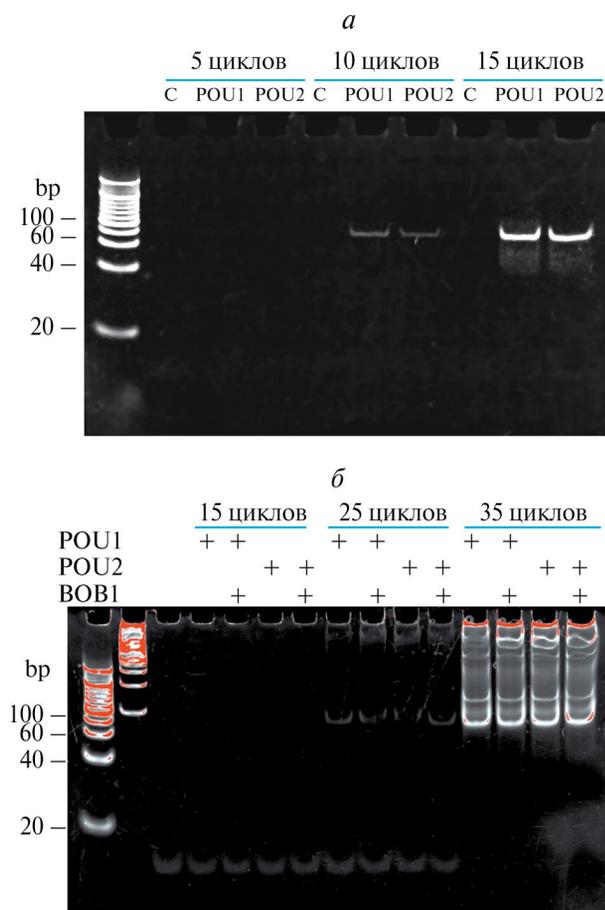


Рис. 2. ПЦР амплификация ДНК полученная после аффинной очистки ДНК-белковых комплексов. *а* — Очистка белков POU1 и POU2 с использованием SBP-тега на Pierce Streptavidin Plus UltraLink Resin. Сигналы специфичны. С — проба без добавления белков POU1 и POU2. *б* — Очистка белков POU1 и POU2 с использованием SBP-тега и последующей очистки белка BOB1 с использованием Flag-тега на ANTI-FLAG M2 Affinity Gel. Выявленные сигналы являются неспецифичными.

ся в использовании метода EMSA в качестве метода разделения комплексов ДНК-белок, вырезании зон геля, где мигрируют соответствующие комплексы, и последующем выделении и амплификации ДНК (рис. 3).

Используя такой подход, нами было проведено 4 цикла обогащения (SELEX) и последующей амплификации ДНК из комплексов ДНК/POU1(2)/POU1(2), свободных от BOB1 или включающих этот коактиватор. Полученные образцы ДНК были секвенированы с помощью NGS технологии. После соответствующей обработки данных секвенирования был проведен биоинформатический анализ последовательностей ДНК на наличие мотивов. Основные мотивы, выявленные с помощью онлайн-инстру-

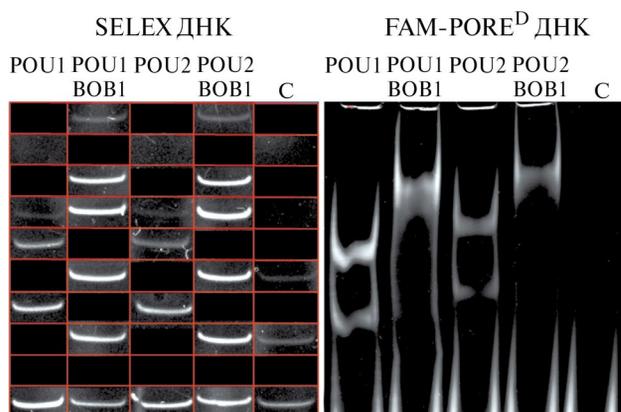


Рис. 3. Интенсивность сигналов при амплификации SELEX ДНК (10 циклов ПЦР), выделенной из фрагментов геля после EMSA (слева). Гель, в котором SELEX ДНК была заменена на FAM-PORE^D ДНК (справа). С — дорожка для контрольной пробы с соответствующей ДНК без белков.

мента MEME (Bailey, Elkan, 1994) представлены на рис. 4.

Полученные в результате процедуры EMSA-SELEX-seq мотивы вполне соответствовали ожидаемым и имели значительное сходство с мотивами, полученными в других лабораториях, например, с мотивами, полученными с помощью CAP-SELEX-seq для POU1 и POU2 (Jolma et al., 2013), а также при анализе данных ChIP-seq для OCT1 (Song et al., 2021). Полученные результаты подтверждают релевантность использованного нами подхода. Например, данный метод позволяет выявлять повышение частоты встречаемости аденина в позиции 5 октамерного мотива, что характерно именно для тройных комплексов, включающих белок BOB1 (рис. 4). Этот факт был установлен уже более двух десятилетий назад и хорошо известен (Gstaiger et al., 1995, 1996). Также следует отметить, что присутствие BOB1 в комплексах ведет к направленному изменению частот нуклеотидов в позициях 2, 3 и 4 октамерного мотива, что предполагает смещение специфичности связывания с ДНК сайтами, и может являться предметом углубленного биоинформатического анализа.

Предлагаемый вариант SELEX процедуры превосходит возможности подхода CAP-SELEX, так как позволяет разделять мономерные и димерные комплексы, а также более сложные комплексы с разным количеством отдельных факторов. Разделение таких сложных комплексов является важной и актуальной задачей, поскольку известно, что при связывании двух и более ТФ, происходит значительное изменение специфичности узнавания последовательностей в ДНК (Jolma et al., 2015).

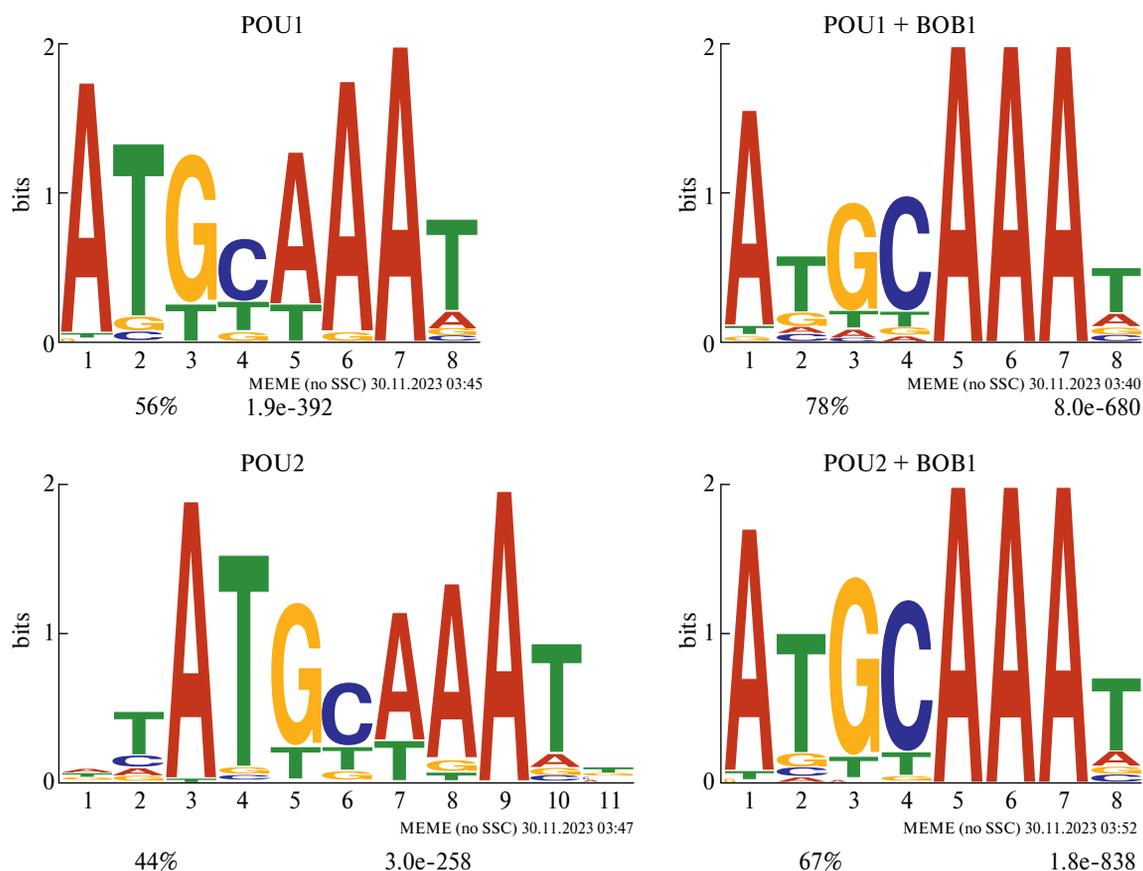


Рис. 4. Основные мотивы, выявленные с помощью онлайн-инструмента MEME (Bailey, Elkan, 1994) после обогащения с помощью EMSA-SELEX и NGS секвенирования. Под изображениями мотивов указана частота (%), слева) встречаемости данных мотивов в последовательностях и статистическая значимость мотива (E-value, справа).

ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена при поддержке Российского научного фонда (проект № 21-14-00369).

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая статья не содержит исследований с использованием животных или людей в качестве объектов.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

СПИСОК ЛИТЕРАТУРЫ

- Bailey T. L., Elkan C. 1994. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. Menlo Park, California: AAAI Press. P. 28.
- Botquin V., Hess H., Fuhrmann G., Anastassiadis C., Gross M. K., Vriend G., Schöler H. R. 1998. New POU di-

mer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2. *Genes Dev.* V. 12. P. 2073.

Gstaiger M., Georgiev O., van Leeuwen H., van der Vliet P., Schaffner W. 1996. The B cell coactivator Bob1 shows DNA sequence-dependent complex formation with Oct-1/Oct-2 factors, leading to differential promoter activation. *EMBO J.* V. 15. P. 2781.

Gstaiger M., Knoepfel L., Georgiev O., Schaffner W., Hovens C. M. 1995. A B-cell coactivator of octamer-binding transcription factors. *Nature.* V. 373. P. 360.

Jolma A., Yan J., Whittington T., Toivonen J., Nitta K. R., Rastas P., Morgunova E., Enge M., Taipale M., Wei G., Palin K., Vaquerizas J. M., Vincentelli R., Luscombe N. M., Hughes T. R., et al. 2013. DNA-binding specificities of human transcription factors. *Cell.* V. 152. P. 327.

Jolma A., Yin Y., Nitta K. R., Dave K., Popov A., Taipale M., Enge M., Kivioja T., Morgunova E., Taipale J. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature.* V. 527. P. 384.

Kibet C. K., Machanick P. 2015. Transcription factor motif quality assessment requires systematic comparative analysis. *F1000Res.* V. 4: ISCB Comm J-1429.

Luo Y., Fujii H., Gerster T., Roeder R. G. 1992. A novel B cell-derived coactivator potentiates the activation of im-

- munoglobulin promoters by octamer-binding transcription factors. *Cell*. V. 71. P. 231.
- Luo Y., Roeder R. G. 1995. Cloning, functional characterization, and mechanism of action of the B-cell-specific transcriptional coactivator OCA-B. *Mol. Cell Biol.* V. 15. P. 4115.
- Meers M. P., Janssens D. H., Henikoff S. 2019. Pioneer factor-nucleosome binding events during differentiation are motif encoded. *Mol. Cell*. V. 75. P. 562.
- Reményi A., Tomilin A., Pohl E., Lins K., Philippsen A., Reinbold R., Schöler H. R., Wilmanns M. 2001. Differential dimer activities of the transcription factor Oct-1 by DNA-induced interface swapping. *Mol. Cell*. V. 8. P. 569.
- Song S., Cao C., Choukrallah M. A., Tang F., Christofori G., Kohler H., Wu F., Fodor B. D., Frederiksen M., Willis S. N., Jackson J. T., Nutt S. L., Dirnhofer S., Stadler M. B., Matthias P. 2021. OBF1 and Oct factors control the germinal center transcriptional program. *Blood*. V. 137. P. 2920.
- Strubin M., Newell J. W., Matthias P. 1995. OBF-1, a novel B cell-specific coactivator that stimulates immunoglobulin promoter activity through association with octamer-binding proteins. *Cell*. V. 80 P. 497.
- Tomilin A., Reményi A., Lins K., Bak H., Leidel S., Vriend G., Wilmanns M., Schöler H. R. 2000. Synergism with the coactivator OBF-1 (OCA-B, BOB-1) is mediated by a specific POU dimer configuration. *Cell*. V. 103. P. 853.
- Weirauch M. T., Cote A., Norel R., Annala M., Zhao Y., Riley T. R., Saez-Rodriguez J., Cokelaer T., Vedenko A., Talukder S., DREAM5 Consortium, Bussemaker H. J., Morris Q. D., Bulyk M. L., Stolovitzky G. et al. 2013. Evaluation of methods for modeling transcription factor sequence specificity. *Nat. Biotechnol.* V. 31. P. 126.

EMSA-SELEX-SEQ METHOD FOR ANALYSIS OF BINDING SITE SEQUENCES IN DNA-PROTEIN COMPLEXES

I. B. Nazarov^{a, *}, M. N. Gordeev^a, A. A. Kuzmin^a, D. S. Zilov^a, E. V. Potapenko^b, A. N. Tomilin^{a, **}

^aInstitute of Cytology RAS, St. Petersburg, 194064, Russia

^bInstitute of Evolution, Department of Evolutionary and Environmental Biology, University of Haifa, Haifa, 3498838, Israel

*e-mail: i.nazarov@incras.ru

**e-mail: a.tomilin@incras.ru

The BOB1 protein (OBF1, OCA-B) is a transcriptional coactivator of two POU domain proteins — OCT1, expressed in all cells, and lymphoid-specific OCT2. The interaction of BOB1 with OCT1/2 plays an important role in the regulation of immune responses in both physiological and pathological contexts. BOB1 is known to form a ternary complex with OCT1/2 bound to DNA in monomeric and certain dimeric configurations, changing the sequence specificity of the binding. To analyze DNA sequences from these complexes, in this work we proposed the EMSA-SELEX-seq method, based on the separation of OCT/BOB1 complexes of various compositions in a non-denaturing polyacrylamide gel (EMSA) followed by the isolation and amplification of the oligonucleotides that they contain (SELEX). Based on several rounds of the enrichment followed by the NGS sequencing and bioinformatics analysis, the DNA sequences were determined and the relevance of this approach was confirmed. Thus, the proposed EMSA-SELEX-seq method allows the analysis of DNA sequences in DNA-protein complexes with varying dimensions of its protein components.

Keywords: EMSA, SELEX, NGS sequencing, OCT1, OCT2, BOB1, OBF1, OCA-B