

СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ СТРУКТУР N-КОНЦЕВОГО МУЛЬТИМЕРИЗУЮЩЕГО ДОМЕНА БЕЛКОВ CTCF ВЫСШИХ МНОГОКЛЕТОЧНЫХ МЕТОДОМ МАЛОУГЛОВОГО РАССЕЯНИЯ РЕНТГЕНОВСКИХ ЛУЧЕЙ (SAXS).

Бончук А.Н.¹, Качалова Г.С.², Бойко К.М.², Максименко О.Г.¹, Георгиев П.Г.¹

¹Институт биологии гена РАН

ул. Вавилова, д. 34/5, г. Москва, 119334, РФ

²Институт биохимии им. А.Н. Баха, ФИЦ Биотехнологии РАН

Ленинский проспект, д. 33, г. Москва, 119071, РФ

e-mail: errinaceus@rambler.ru

Аннотация. При помощи малоуглового рассеяния рентгеновских лучей (SAXS) получены данные о пространственной структуре N-концевых доменов инсуляторных белков CTCF представителей основных филогенетических групп высших многоклеточных. В отличие от N-концевого домена белка CTCF *Drosophila*, являющегося мультимером, аналогичные домены других животных не способны формировать олигомеры, большинство из них являются развернутыми. Таким образом, функция мультимеризации при помощи этого домена является уникальной для насекомых.

Ключевые слова: мультимеризация, малоугловое рассеяние рентгеновских лучей, хроматин.

STRUCTURAL STUDIES OF OF N-TERMINAL DOMAINS OF CTCF PROTEINS FROM HIGHER METAZOANS USING SMALL-ANGLE X-RAY SCATTERING (SAXS)

Bonchuk A.N.¹, Kachalova G.S.², Boyko K.M.², Maksimenko O.G.¹, Georgiev P.G.¹

¹Institute of gene biology RAS

Vavilova street, 34/5, Moscow, 119334, Russia

²A.N. Bach Institute of biochemistry, Research Center of Biotechnology RAS,

Leninsky av., 33, Moscow, 119071, Russia

e-mail: errinaceus@rambler.ru

Abstract. Spatial structure of N-terminal domains of CTCF insulator proteins from higher metazoans was studied using small-angle X-ray scattering (SAXS). In contrast to N-terminal domains of *Drosophila* CTCF protein, similar domains from other animals were unable to form oligomers and mainly are unfolded. Multimerization ability probably is unique for insects.

Keywords: multimerization, small-angle X-ray scattering, chromatin.

Инсуляторы – геномные элементы, обладающие способностью блокировать энхансер-промоторные взаимодействия и препятствовать распространению гетерохроматина. При этом инсуляторы способны взаимодействовать между собой на больших расстояниях в геноме (более 1000 п.н.) и таким образом принимать участие в организации пространственной структуры хроматина [1]. CTCF – консервативный среди высших многоклеточных белок, связывающийся с ДНК инсуляторов и единственный известный ДНК-связывающий инсуляторный белок у позвоночных [2, 3]. Изолированные сайты связывания белка CTCF способны обеспечивать взаимодействие между удаленными участками ДНК [1, 4]. По всей видимости, в процессе установления и поддержания таких контактов в геноме принимают участие белковые домены, обеспечивающие специфичное белок-белковое взаимодействие и формирование мультимеров. Белок CTCF имеет в своем составе 11 ДНК-связывающих доменов типа «цинковые пальцы», а также N- и C- концевые домены с неизвестными функциями. Была показана способность белка CTCF мыши к димеризации, но домен, необходимый для этого не был идентифицирован [5]. Ранее нами была разработана схема экспрессии и очистки N-концевого домена белка CTCF *Drosophila melanogaster* и была показана его способность к формированию мультимеров *in vitro* и *in vivo*, мультимеризация необходима для правильного функционирования белка [6]. N-концевой домен белка CTCF *Drosophila* не имеет гомологов среди известных белковых доменов, биоинформатическими методами предсказывается наличие короткого участка coiled-coil, который, однако, не является необходимым для мультимеризации. При помощи методов кругового дихроизма и малоуглового рассеяния рентгеновского излучения было показано, что N-концевой домен белка CTCF *Drosophila* является глобулярным тетрамером, но при этом практически не имеет элементов вторичной структуры. Аналогичные домены белков CTCF представителей основных филогенетических групп высших многоклеточных не имеют выраженной гомологии с изученным мультимеризующимся доменом CTCF *Drosophila*, однако, выполняют схожие функции, с целью получения данных об их структуре мы провели исследование данных доменов при помощи метода малоуглового рассеяния рентгеновских лучей (SAXS).

Нами были клонированы, экспрессированы и очищены N-концевые домены белков CTCF следующих организмов: тихоокеанская устрица (*Crassostrea gigas*) 1-177 ак; асцидия (*Ciona intestinalis*) 1-288 ак; дафния обыкновенная (*Daphnia pulex*) 1-220 ак; калифорнийский морской еж (*Strongylocentrotus purpuratus*) 1-325 ак; человек (*Homo sapiens*) 1-264 ак. Ранее при помощи метода кругового дихроизма было показано, что N-концевой домен белка CTCF человека неструктурирован [7]. Эксперименты по химической шивке показали, что ни один из перечисленных полипептидов не способен к формированию мультимеров. Дальнейшее исследование N-концевого домена белка CTCF проводилось при помощи метода малоуглового рассеяния рентгеновского излучения (SAXS). Эксперимент проводился на источнике рентгеновского излучения BM29 BioSAXS в ESRF (Гренобль, Франция). Измерения проводились при длине волны 0.099 нм, концентрация белка варьировалась от 1 до 7.5 мг/мл. Радиус гирации (R_g)

белковой молекулы в растворе был рассчитан при помощи аппроксимации Гинье при малых углах ($s < 1.3/Rg$, $s = 4\pi \sin\theta/\lambda$, где 2θ – угол рассеяния), интенсивность $I(s)$ принималась равной $I_0 \exp(-(sRg)^2/3)$. Для оценки максимального размера частиц D_{max} , функция парных расстояний $P(r)$ была рассчитана при помощи алгоритма GNOM из программного пакета ATSAS. Молекулярный вес молекул был рассчитан исходя из исключенного объема гидратированных белковых молекул по закону Поро для гомогенных частиц. Данные, полученные для N-концевых доменов белков CTCF при разных концентрациях образцов, приведены в таблице 1. Видно, что только для белка CTCF *D. melanogaster* молекулярная масса по результатам измерений в несколько раз превосходит молекулярную массу мономера. Почти двукратное превышение наблюдаемой молекулярной массы наблюдается также для CTCF *D. pulex* и *S. purpuratus*, однако мультимеризация этих доменов не подтверждается другими методами и полученные результаты могут быть следствием того, что для данных доменов измерения проводились при более низких концентрациях, что стало причиной большой погрешности.

Таблица 1 – Рассчитанные по данным малоуглового рассеяния рентгеновских лучей характеристики частиц N-концевых доменов белков CTCF в растворе

Полипептид	Концентрация, мг/мл	R g, нм	D _{max} , нм	V _p , нм ³	Молекулярная масса, кДа	Молекулярная масса мономера, кДа
Crassostrea gigas 1-177	1.35	4.36	14.94	42.85	25	19.9
Crassostrea gigas 1-177	5.4	4.18	14.63	41.8	24	19.9
Ciona intestinalis 1-288	1.65	4.3	15.04	61.73	36	31.6
Ciona intestinalis 1-288	6.5	3.83	13.43	45.9	27	31.6
Daphnia pulex 1-220	0.87	5.4	20.18	85.2	50	25.1
Daphnia pulex 1-220	1.75	5.2	18.28	77.9	45	25.1
Strongylocentrotus purpuratus 1-325	0.85	5.9	20.86	112.5	66	35.5
Strongylocentrotus purpuratus 1-325	1.7	5.7	17.75	93.25	54	35.5
Homo sapiens 1-264	1.0	4.2	16.55	77.2	45	31.4
Homo sapiens 1-264	4.2	3.9	14.64	68.5	40	31.4
Drosophila melanogaster 1-163	1.0	4.65	15.2	110.5	56	18.7
Drosophila melanogaster 1-163	7.5	5.0	16.0	125.5	85	18.7

Для анализа структурированности исследуемых доменов проводился анализ графиков Кратки ($I \times s^2$ относительно s), форма которых сильно зависит от степени свернутости исследуемого белка. Данные представлены на рисунке 1. Как видно на примере N-концевого домена CTCF *D. melanogaster*, для структурированных полипептидов характерна куполообразная форма кривой при малых значениях s , в то же время для развернутых полипептидов кривая имеет логарифмическую форму. Помимо *D. melanogaster* куполообразный изгиб наблюдается для N-концевого домена *C. intestinalis*.

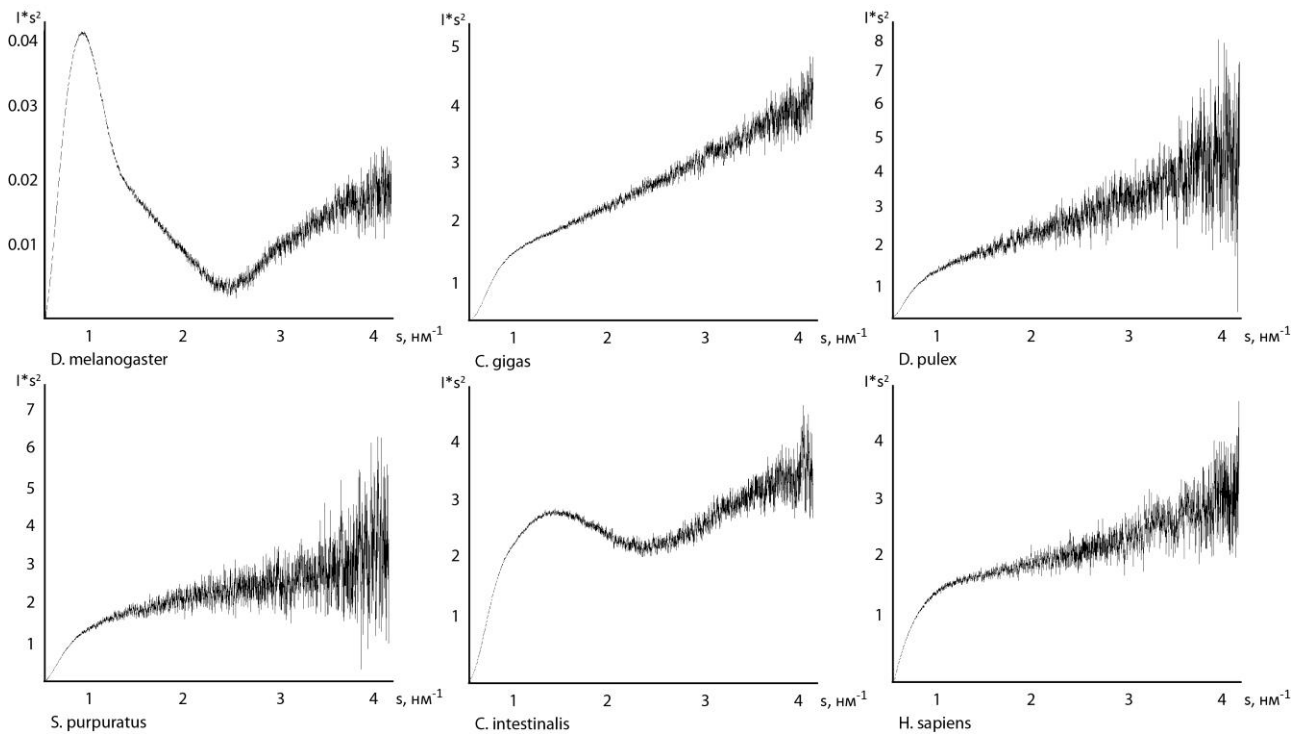


Рисунок 1 – Данные малоуглового рассеяния рентгеновских лучей (SAXS), полученные для N-концевых доменов белков CTCF высших многоклеточных, представленные в виде графиков Кратки

Таким образом, полученные нами данные демонстрируют, что N-концевые домены инсуляторных белков CTCF являются мономерами, за исключением CTCF *Drosophila*, и являются развернутыми, за исключением белков *D. melanogaster* и *C. intestinalis*. Вероятно, функцию мультимеризации у этих белков выполняют другие домены.

Работа выполнена при поддержке Российского научного фонда (проект 14–24–00166).

Список литературы / References:

1. Kyrchanova O., Chetverina D., Maksimenko O., Kullyev A., Georgiev P. Orientation-dependent interaction between *Drosophila* insulators is a property of this class of regulatory elements. *Nucleic Acids Res.*, 2008, vol. 36, iss. 22, pp. 7019-7028.
2. Moon H., Filippova G., Loukinov D., Pugacheva E., Chen Q., Smith S., Munhall A., Grewe B., Bartkuhn M., Arnold R. CTCF is conserved from *Drosophila* to humans and confers enhancer blocking of the Fab-8 insulator. *EMBO Rep.*, 2005, vol. 6, iss. 2, pp. 165-170.
3. Bell A., West A., Felsenfeld G. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell*, 1999, vol. 98, iss. 3, pp. 387-396.
4. Kyrchanova O., Ivlieva T., Toshchakov S., Parshikov A., Maksimenko O., Georgiev P. Selective interactions of boundaries with upstream region of Abd-B promoter in *Drosophila* bithorax complex and role of dCTCF in this process. *Nucleic Acids Res.*, 2011, vol. 39, pp. 3042-3052.
5. Pant V., Kurukuti S., Pugacheva E., Shamsuddin S., Mariano P., Renkawitz R., Klenova E., Lobanekov V., Ohlsson R. Mutation of a single CTCF target site within the H19 imprinting control region leads to loss of Igf2 imprinting and complex patterns of de novo methylation upon maternal inheritance. *Mol. Cell Biol.*, 2004, vol. 24, iss. 8, pp. 3497-3504.
6. Bonchuk A., Maksimenko O., Kyrchanova O., Ivlieva T., Mogila V., Deshpande G., Wolle D., Schedl P., Georgiev P. Functional role of dimerization and CP190 interacting domains of CTCF protein in *Drosophila melanogaster*. *BMC Biol.*, 2015, vol. 13, no. 63.
7. Martinez S., Miranda J. CTCF terminal segments are unstructured. *Protein Sci.*, 2010, vol. 19, iss. 5, pp. 1110-1116.