

В поисках N α -ацетилтрансфераз бактерии *Thermus thermophilus*: идентификация, структурный анализ и предсказание субстратной специфичности

Лаптева Ю.*, Кудряшов Т., Локтюшов Е., Трунилина М., Дерюшева Е.,
Быков В., Соколов А.

ФИЦ «Пуцинский научный центр биологических исследований» РАН, Институт биологического приборостроения, Пуцино, Россия

* yulia.s.lapteva@gmail.com

Ключевые слова: N-ацетилтрансфераза; GNAT-домен; N-концевое ацетилирование; субстратная специфичность; *Thermus thermophilus*

Мотивация и цель: N α -ацетилтрансферазы (N α -АТ) катализируют перенос ацетильной группы от ацетилкоэнзима А на α -аминогруппу широкого спектра субстратов (аминокислоты, белки, серотонин и др.) [1]. Это большой класс ферментов, характеризующихся консервативной трехмерной структурой, но низким процентом идентичности последовательностей, что затрудняет их корректную аннотацию в геномах и предсказание их субстратной специфичности. N α -АТ входят в состав суперсемейства GNAT-ацетилтрансфераз, которое характеризуется специфическим чередованием элементов вторичных структур, а именно $\beta 0$ - $\beta 1$ - $\alpha 1$ - $\alpha 2$ - $\beta 2$ - $\beta 3$ - $\beta 4$ - $\alpha 3$ - $\beta 5$ - $\alpha 4$ - $\beta 6$ - $\beta 7$, и их уникальной трехмерной укладкой – так называемый GNAT-домен [2]. Ферменты семейства N α -АТ (как и лизин-АТ (КАТ)) содержат единичный GNAT-домен, в то время как ферменты других семейств АТ, могут содержать несколько GNAT, а также другие домены. Центральная часть GNAT-домена содержит петлю, связывающую ацетил-Ко А, которая имеет консенсус аминокислот Q/R-X-X-G-G/A. Учитывая важность N α -ацетилирования в контроле активности ферментов [3] и функционирования белков (в том числе и факторов патогенности) в клетках [4]; участие в биосинтезе и деградации антибиотиков [5]; применение N α -АТ в биотехнологии ацетилирования рекомбинантных белков [6], актуальным представляется их корректное предсказание и аннотация в геномах микроорганизмов. Цель данной работы заключается в использовании алгоритмов и методов биоинформатики для поиска N α -АТ в геноме бактерии *T. thermophilus* HB8, их структурно-функционального анализа с целью предсказания их субстратной специфичности.

Методы и алгоритмы: Поиск аминокислотных последовательностей множества предполагаемых ацетилтрансфераз в геноме бактерии *T. thermophilus* проводили в NCBI по запросу “acetyltransferase”. Выравнивание аминокислотных последовательностей белков выполняли с использованием программ Clustal Omega [7] и/или T-coffee [8]. Предсказание вторичной структуры белков проводили при помощи онлайн-сервера PsiPred [9]. Предсказание третичной структуры белков осуществляли в программе AlphaFold2 [10]. Для выравнивания третичных структур белков использовали программу PyMol. Поиск белков-ортологов проводили при помощи сервера BLAST и DALI [11]. С целью анализа предполагаемой ферментативной активности гены некоторых ферментов были клонированы с использованием

методов генной инженерии. Получены очищенные препараты ферментов. $N\alpha$ -ацетилтрансферазная активность была исследована в реакции *in vitro* ацетилирования модельного белка [12].

Результаты: Анализ генома бактерии *T. thermophilus* HB8 на предмет наличия аннотированных “ $N\alpha$ -acetyltransferase” или “N-terminal acetyltransferase” не дал результата. Однако это свидетельствует не об отсутствии $N\alpha$ -АТ в бактерии, а о сложности адекватной аннотации ферментов данного многочисленного семейства. Менее специфичный поиск по запросу “acetyltransferase” позволил выявить свыше десяти аннотированных «предполагаемых» АТ. Однако это значительно меньше, чем в бактерии *E. coli* (29 АТ) или *Streptomyces* (72 АТ), что опять свидетельствует о сложной идентификации и корректной аннотации ферментов данного семейства. Для отбора среди множества АТ именно $N\alpha$ -АТ мы провели множественное выравнивание аминокислотных последовательностей исследуемых ферментов с детально изученными белками-ортологами из других бактерий. Выравнивание позволило отобрать шесть ферментов, несущих только один GNAT-домен, в основном это сравнительно небольшие белки, размером от 150–205 аминокислот. Выравнивание трехмерных структур отобранных шести предполагаемых $N\alpha$ -АТ выявило характерную для ферментов данного семейства топологию GNAT-домена, в особенности расположение двух β -листов, формирующих активный центр. Однако для трех из шести предполагаемых $N\alpha$ -АТ чередование элементов вторичных структур не соответствует каноническому за счет дополнительных вставок/делеций α -спиралей и/или β -структур. Множественное выравнивание последовательностей отобранных шести $N\alpha$ -АТ позволило идентифицировать консенсус аминокислот, связывающих ацетил-Ко А. У *T. thermophilus* он имеет вид $Q_{/R}-G_{/R}-X-G-L_{/V}-G_{/I}$, где нижним индексом отмечены остатки, встречающиеся только в одной последовательности $N\alpha$ -АТ из шести. Нами были клонированы, получены очищенные препараты ряда ферментов и исследована их $N\alpha$ -ацетилтрансферазная активность в условия *in vitro* на модельном белке.

Выводы: Результаты работы позволили выделить специфические критерии для поиска, предсказания и отбора $N\alpha$ -АТ среди множеств GNAT-АТ. Полученные данные в дальнейшем могут лечь в основу создания алгоритма идентификации $N\alpha$ -АТ в геномах других организмов, что в настоящее время, несомненно, очень актуально для корректной аннотации ферментов семейства $N\alpha$ -АТ и предсказания их субстратной специфичности.

Финансирование: Исследование выполнено при финансовой поддержке Российского научного фонда, грант № 23-24-00478, <https://rscf.ru/project/23-24-00478/>.

In search for $N\alpha$ -acetyltransferases from the bacterium *Thermus thermophilus*: identification, structural analysis and prediction of substrate specificity

Lapteva Yu.*, Kudryashov T., Loktyushov E., Trunilina M., Deryusheva E., Bykov V., Sokolov A.

Federal Research Center “Pushchino Scientific Center for Biological Research”, RAS,
Institute of Biological Instrumentation, Pushchino, Russia

* yulia.s.lapteva@gmail.com

Key words: N-acetyltransferase; GNAT domain; N-terminal acetylation; substrate specificity; *Thermus thermophilus*

Motivation and Aim: N α -acetyltransferases (N α -AT) catalyze the transfer of an acetyl group from acetyl coenzyme A to the α -amino group of a wide range of substrates (amino acids, proteins, serotonin, etc.) [1]. This is a large class of enzymes characterized by a conserved three-dimensional structure but a low percentage of sequence identity, which makes it difficult to correctly annotate them in genomes and predict their substrate specificity. N α -ATs are part of the GNAT acetyltransferase superfamily, which is characterized by a specific alternation of secondary structure elements, namely “ β 0- β 1- α 1- α 2- β 2- β 3- β 4- α 3- β 5- α 4- β 6- β 7”, and their unique three-dimensional folding, the so-called “GNAT domain” [2]. Enzymes of the N α -AT family (like lysine AT (KAT)) contain a single GNAT domain, while enzymes of other AT families may contain multiple GNATs as well as other domains. The central part of the GNAT domain contains an acetyl-Co A binding loop, which has the amino acid consensus “Q/R-X-X-G-G/A”. Considering the importance of N α -acetylation in the control of enzyme activity [3] and the functioning of proteins (including pathogenicity factors) in cells [4]; participation in the biosynthesis and degradation of antibiotics [5]; the use of N α -AT in the biotechnology of acetylation of recombinant proteins [6], their correct prediction and annotation in the genomes of microorganisms seems relevant. The purpose of this work is to use algorithms and bioinformatics methods to search for N α -AT in the genome of the bacterium *T. thermophilus* HB8, their structural and functional analysis in order to predict their substrate specificity.

Methods and Algorithms: A search for amino acid sequences of multiple putative acetyltransferases in the genome of the bacterium *Thermus thermophilus* was carried out in NCBI using the query “acetyltransferase”. Alignment of amino acid sequences of proteins was carried out using the Clustal Omega [7] and/or T-coffee [8] programs. Prediction of the secondary structure of proteins was carried out using the PsiPred online server [9]. Prediction of the tertiary structure of proteins was carried out using the Alphafold2 program [10]. The PyMol program was used to align protein tertiary structures. The search for orthologous proteins was carried out using the BLAST and DALI servers [11]. In order to analyze putative enzymatic activities, the genes for some enzymes have been cloned using genetic engineering techniques. Purified enzyme preparations were obtained. N α -acetyltransferase activity was studied in the in vitro reaction of acetylation of a model protein [12].

Results: Analysis of the genome of the bacterium *T. thermophilus* HB8 for the presence of an annotated “N α -acetyltransferase” or “N-terminal acetyltransferase” was inconclusive. However, this does not indicate the absence of N α -AT in the bacterium, but rather the difficulty of adequately annotating the enzymes of this large family. A less specific search for “acetyltransferase” identified over ten annotated “putative” Abs. However, this is significantly less than in the bacteria *E. coli* (29 AT) or *Streptomyces* (72 AT), which again indicates the complex identification and correct annotation of enzymes of this family. To select N α -AT among many ATs, we carried out multiple alignment of the amino acid sequences of the enzymes under study with detailed orthologous proteins from other bacteria. The alignment allowed us to select six enzymes that carry only one GNAT domain, mostly relatively small proteins ranging in size from 150-205 amino acids. Alignment of the three-dimensional structures of the selected six putative N α -ATs revealed the GNAT domain topology characteristic of enzymes of this

family, in particular the arrangement of the two β -sheets forming the active site. However, for three of the six putative N α -ATs, the alternation of secondary structure elements does not correspond to the canonical one due to additional insertions/deletions of α -helices and/or β -structures. Multiple sequence alignment of selected six N α -ATs made it possible to identify a consensus of amino acids binding acetyl-Co A. In *T. thermophilus*, it has the form “Q/R-G/R-X-G-L/V-G/I”, where the subscript denotes residues found only in one N α -AT sequence out of six. We cloned and obtained purified preparations of a number of enzymes and studied their N α -acetyltransferase activity under in vitro conditions on a model protein.

Conclusion: The results of the work made it possible to identify specific criteria for searching, predicting and selecting N α -ATs among the sets of GNAT-ATs. The data obtained can subsequently form the basis for creating an algorithm for identifying N α -ATs in the genomes of other organisms, which is currently, undoubtedly, very important for the correct annotation of enzymes of the N α -AT family and prediction of their substrate specificity.

Funding: Research was funded by a grant from the Russian Science Foundation, No. 23-24-00478, <https://rscf.ru/project/23-24-00478/>.

Список литературы/References

1. Deng S., Marmorstein R. Protein N-Terminal Acetylation: Structural Basis, Mechanism, Versatility, and Regulation. *Trends Biochem Sci.* 2021;46:15-27
2. Abboud A., Bedoucha P., Byska J., Arnesen T., Reuter N. Dynamics-function relationship in the catalytic domains of N-terminal acetyltransferases. *Comput Struct Biotechnol J.* 2020;18:532-547
3. Parks A.R., Escalante-Semerena J.C. Modulation of the bacterial CobB sirtuin deacylase activity by N-terminal acetylation. *Proc Natl Acad Sci USA.* 2020;117:15895-15901
4. Permyakov S.E., Vologzhannikova A.A., Emelyanenko V.I., Knyazeva E.L., Kazakov A.S., Lapteva Y.S., Permyakova M.E., Zhadan A.P., Permyakov E.A. The impact of alpha-N-acetylation on structural and functional status of parvalbumin. *Cell Calcium.* 2012;52:366-376
5. Favrot L., Blanchard J.S., Vergnolle O. Bacterial GCN5-Related N-Acetyltransferases: From Resistance to Regulation. *Biochemistry.* 2016;23:989-1002
6. Esipov R.S., Makarov D.A., Stepanenko V.N., Miroshnikov A.I. Development of the intein-mediated method for production of recombinant thymosin beta4 from the acetylated in vivo fusion protein. *J Biotechnol.* 2016;228:73-81
7. Sievers F., Higgins D.G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* 2018;27:135-145
8. Notredame C., Higgins D.G., Heringa J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol.* 2000;302:205-217
9. McGuffin L.J., Bryson K., Jones D.T. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000;16:404-405
10. Jumper J., Evans R., Pritzel A. et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596:583-589
11. Holm L. DALI and the persistence of protein shape. *Protein Sci.* 2020;29:128-140
12. Lapteva Y.S., Vologzhannikova A.A., Sokolov A.S., Ismailov R.G., Uversky V.N., Permyakov S.E. In Vitro N-Terminal Acetylation of Bacterially Expressed Parvalbumins by N-Terminal Acetyltransferases from Escherichia coli. *Appl Biochem Biotechnol.* 2021;193:1365-1378