

НЕРІВНОМІРНІСТЬ ЛОКАЛІЗАЦІЇ САЙТІВ CG І CNG У ТРАНСКРИПТАХ
ГЕНІВ *XIST* ТА *TSIX*

Хоча концентрація сайтів CG та CNG у *Xist* РНК у середньому становить 0,8 %, тобто навіть менше, ніж загальний рівень у геномі (1 %), локалізація цих сайтів вкрай нерівномірна. Концентрація сайтів CG та CNG досягає максимальної щільності, що перевищує середній рівень у декілька разів, на 5'-кінці *Xist*-транскрипту. Кількість сайтів CG і CNG у *Tsix* РНК навіть менша, ніж у *Xist*-транскрипті, та складає всього 0,6 %. Проте сайти CG та CNG у *Tsix* РНК концентруються поблизу ділянки, яка перекриває послідовність гена *Xist*, а також на 5'-кінці. Ці дані підтверджують гіпотезу, що послідовності 5'-кінця *Xist* РНК безпосередньо зв'язуються з комплементарною ДНК в інших ділянках X-хромосоми, викликаючи РНК-залежне метилювання ДНК та X-інактивацію.

КЛЮЧОВІ СЛОВА: X-хромосома, *Xist*, *Tsix*, CG, CNG, локалізація сайтів, РНК-залежне метилювання ДНК, X-інактивація.

ВСТУП. Число X-хромосом відрізняється у XX- та XY-індивідуумів, а у випадку X-анеуплоїдії може збільшуватись у декілька разів. Компенсація дози X-зчеплених генів досягається завдяки інактивації X-хромосом, що відбувається на ранніх стадіях розвитку в усіх клітинах, які містять більше однієї X-хромосоми. X-інактивація призводить до репресії більшості генів у всіх наявних у клітині X-хромосомах, окрім однієї, яка зберігає активність. Неактивний стан X-хромосом стабільно успадковується всіма подальшими поколіннями клітин (за винятком статевих клітин, що реактивують усі власні X-хромосоми).

X-інактивація залежить від двох генів, розташованих на X-хромосомі, – *Xist* та *Tsix*, що частково перекриваються. Дані гени зчитуються з утворенням довгих некодуючих РНК [6]. Варто відзначити, що *Xist*-транскрипт зазнає альтернативного сплайсингу з утворенням, за інформацією, отриманою нами з бази даних Ensembl, десяти різних варіантів зрілої РНК. *Tsix*-транскрипт зазнає сплайсингу не завжди [8]; очевидно, те розмаїття його послідовностей, яке вважали результатом сплайсингу, може виникати просто внаслідок розпаду первинного транскрипту.

Xist РНК зв'язується з тією ж X-хромосомою, з якої транскрибується, тобто діє у *cis*-положенні, й викликає гетерохроматинізацію даної

хромосоми та інактивацію її генів [6]. На противагу цьому, *Tsix* РНК зчитується з активної X-хромосоми і, будучи антисмисловою відносно послідовності гена *Xist*, протидіє активності останнього, теж у *cis*-положенні [6].

З метою пояснення молекулярного механізму репресії *Xist*-транскриптом своєї хромосоми ми запропонували гіпотезу, що 5'-кінець даного транскрипту безпосередньо гібридується з комплементарними ділянками у ДНК інших генів, витісняючи один із ланцюгів подвійної спіралі ДНК та формуючи натомість комплекси РНК–ДНК (рис. 1) [4]. Ми припускаємо, що такі комплекси приваблюють ДНК-метилтрансферази, й останні, як наслідок, метилують *de novo* залишки цитозину в сайтах метилювання CG та CNG (літерою N позначено будь-який із чотирьох нуклеотидів) у межах ДНК-мішеней, які виявились комплементарно зв'язаними з РНК [2]. Очевидно, взаємодія між *Xist*-транскриптом та комплементарною ДНК повинна починатись ще під час синтезу *Xist*-транскрипту, перед його відокремленням від РНК-полімерази, чим і пояснюється те, що X-інактивація має *cis*-характер (рис. 1) [3].

Імовірно, в X-хромосомі, що зберігає активність, транскрипт гена *Tsix* у подібний спосіб викликає інактивацію гена *Xist*, зв'язуючись із ДНК останнього, в результаті чого теж ініціюється РНК-залежне метилювання ДНК *de novo* (рис. 1) [3].

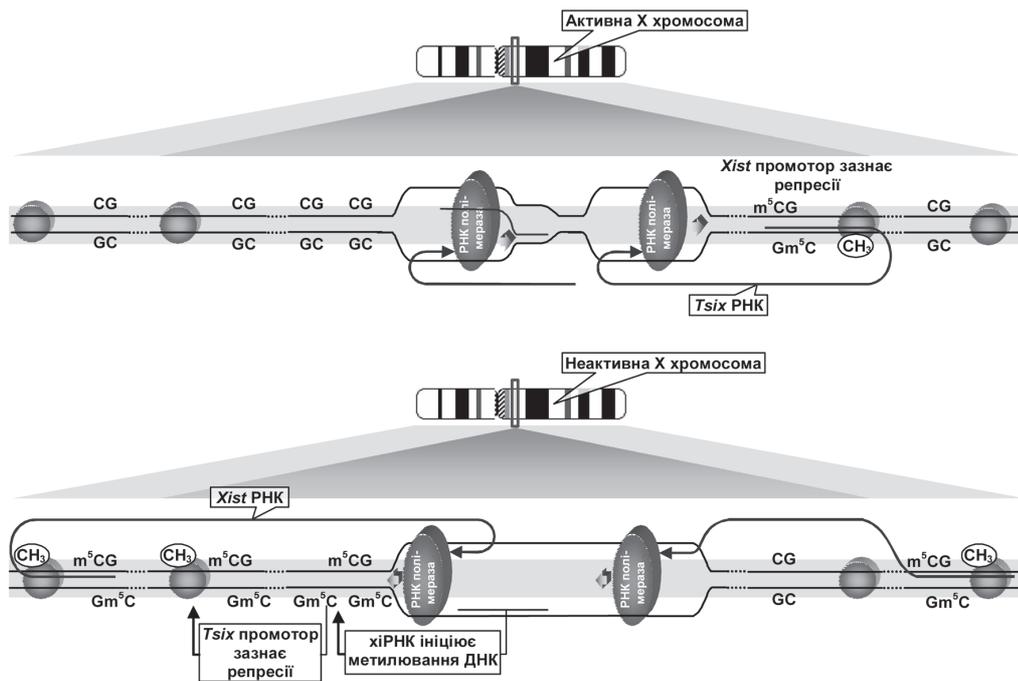


Рис. 1. Молекулярний механізм X-інактивації (m⁵C – 5-метилцитозин).

Оскільки лише одна з X-хромосом у клітині залишається активною, X-інактивація потребує, щоб ген *Xist* експресувався всіма алелями, за винятком одного. Таким чином, алельне виключення повинно стосуватись не гена *Xist*, а його репресора, тобто гена *Tsix* (або ж активаторів експресії останнього).

Нещодавно було показано, що транскрипти генів *Xist* та *Tsix* можуть розщеплюватися з утворенням т. зв. хіРНК – коротких некодуючих РНК довжиною 24–42 нуклеотиди [6]. Ми припускаємо, що хіРНК якраз і опосередковують алельне виключення гена *Tsix*, гібридизуючись із послідовностями його алелів [3]. Оскільки РНК-полімераза, як передбачається, у процесі транскрипції здатна відокремлювати РНК, що з'явилися на її шляху [2], хіРНК викликають РНК-залежне метилювання ДНК того алеля гена *Tsix*, який менш активно транскрибується (рис. 1) [3].

Запропонована гіпотеза передбачає, що в послідовностях *Xist* та *Tsix* повинні міститися сайти, комплементарні сайтам метилювання, тобто динуклеотиди CG та тринуклеотиди CNG. Більше того, ці сайти повинні бути зосереджені, головним чином, на 5'-кінці транскриптів.

Метою даної роботи було перевірити дані передбачення шляхом виявлення зазначених сайтів та встановлення особливостей їх розподілу у транскриптах генів *Xist* та *Tsix*.

МЕТОДИ ДОСЛІДЖЕННЯ. Нуклеотидні послідовності транскриптів генів довгих неко-

дуючих РНК людини *Xist* та *Tsix* було отримано з бази даних Ensembl (<http://www.ensembl.org>). Аналіз послідовностей проводили *in silico* за допомогою програмного пакета Vector NTI Advance® 11.5, виробництва Life Technologies Corp., а також Microsoft Excel.

Вміст кожного з нуклеотидів, очікувану концентрацію та насправді наявну концентрацію сайтів CG і CNG у досліджуваних послідовностях, розподіл сайтів CG та CNG у них визначали відповідно до методик, описаних нами раніше [1]. Зазначимо, що, хоча транскрипти генів *Xist* та *Tsix* зазнають процесингу і сплайсингу, вивчали їх повну первинну послідовність, оскільки передбачається, що взаємодія зазначених РНК з ДНК розпочинається ще до завершення їх синтезу і, відповідно, до вирізання інтронів. Проте отримана картина розподілу сайтів CG та CNG у *Xist*-транскрипті потім була зіставлена з його екзонами. Щодо транскрипту гена *Tsix*, то таке зіставлення не проводили з огляду на те, що послідовність цієї РНК зазнає процесингу не завжди і, за даними [7], свою активність проявляє у несплайсованій формі.

Інформацію щодо метилювання послідовностей ДНК генів *Xist* та *Tsix* у клітинах різного рівня диференціації було отримано з бази даних Epigenomics (<http://www.ncbi.nlm.nih.gov/epigenomics/>).

РЕЗУЛЬТАТИ Й ОБГОВОРЕННЯ. Тоді коли в геномі людини середня концентрація гуаніну,

а також цитозину складає 21 %, аденіну, а також тиміну – 29 %, послідовність первинного транскрипту гена *Xist* містить дещо нижчу кількість гуаніну (19,54 %), цитозину (20,27 %) та аденіну (27,35 %), а концентрація урацилу є вищою від середнього для геному рівня і становить 32,84 %. Відповідно, на більшості позицій нуклеотидів у даній послідовності з меншою імовірністю будуть зустрічатися гуанін, цитозин і аденін порівняно із середнім рівнем даних нуклеотидів у геномі. З огляду на ці дані, у *Xist* PHK варто очікувати середньої концентрації сайтів CNG на рівні $0,1954 \times (0,1954 + 0,2027 + 0,2735 + 0,3284) \times 0,2027 = 0,1954 \times 1 \times 0,2027 = 3,96$ %, що є нижчим значенням, ніж середній для геному рівень, який складає 4,41 %.

Відзначимо, що середня концентрація сайтів CG у геномі мала б так само дорівнювати 4,41 %, однак у ході еволюції більшість залишків цитозину в сайтах CG зазнала транзиції через заміну цитозину на тимін внаслідок дезамінування 5-метилцитозину. Тому в наш час у геномі людини динуклеотид CG зустрічається з частотою 1 % [5, 9].

Послідовність транскрипту гена *Tsix* містить нижчу кількість цитозину (19,30 %) та урацилу (28,50 %) і вищу – гуаніну (21,34 %) та аденіну (30,86 %) порівняно із середніми для геному рівнями концентрації даних нуклеотидів. Таким чином, розрахункова концентрація сайтів CNG у зазначеній послідовності повинна становити 4,12 %.

У результаті проведеного дослідження послідовностей ми, однак, виявили, що концентрація сайтів CG у послідовності первинного транскрипту гена *Xist* становить лише 0,8 %, а в послідовності *Tsix* PHK відносна кількість даних динуклеотидів ще менша і дорівнює всього 0,6 %.

Концентрація сайтів CNG у первинному транскрипті гена *Xist* складає 3,89 %, що також дещо менше, ніж рівень, середній для геному, та значення, очікуване для даної PHK. У транскрипті гена *Tsix* відносна кількість тринуклеотидів CNG становить 4,54 %, тобто перевищує як середню концентрацію в геномі, так і очікуване розрахункове значення для послідовності цієї PHK.

Проте локалізація сайтів CG і CNG у транскриптах генів *Xist* та *Tsix* характеризується винятковою нерівномірністю і наявністю декількох піків концентрації. Особливу увагу привертає ділянка, що включає перші 2000 нуклеотидів від 5'-кінця транскрипту гена *Xist* (тобто в межах його першого екзону), а також так звана А-ділянка, яка розташовується між

350-м та 770-м нуклеотидами від 5'-кінця. У даних ділянках концентрація сайту CG перевищує середній геномний рівень у декілька разів, досягаючи, відповідно, 3 та 5 %. Також концентрація тринуклеотиду CNG досягає свого максимуму в даних ділянках – 5,7 та 7,6 % відповідно (рис. 2). Оскільки ці піки концентрації сайтів CG та CNG перебувають у межах екзону, вони не втрачаються при сплайсингу *Xist* PHK.

У *Tsix* PHK найвищий пік концентрації сайтів CG та CNG (2,4 і 8,2 % відповідно) локалізується поблизу ділянки, послідовність якої перекривається з послідовністю гена *Xist* (рис. 3). Цікаво, що в первинному транскрипті гена *Xist* теж є декілька піків концентрації сайтів CG і CNG, розташованих впритул до ділянки, де послідовності генів *Xist* та *Tsix* перекриваються (рис. 2). Декілька значних піків концентрації сайтів CG та CNG було виявлено на 5'-кінці *Tsix* PHK (рис. 3).

Таким чином, сайти CG і CNG у транскриптах генів *Xist* та *Tsix* наявні (хоча й у нижчій кількості, ніж загалом у геномі). Більше того, на 5'-кінці даних транскриптів концентрація зазначених сайтів у декілька разів перевищує середній рівень. Це і передбачалось з огляду на запропоновану нами гіпотезу, що молекулярний механізм Х-інактивації включає в себе пряме зв'язування *Xist*-транскрипту з комплементарними послідовностями ДНК Х-хромосоми з наступною ініціацією репресії цих послідовностей шляхом PHK-залежного метилювання ДНК *de novo*.

За інформацією бази даних Epigenomics, у послідовності гена *Xist*, комплементарній 5'-кінцю *Xist* PHK (де, як ми встановили, локалізується найвищий пік концентрації сайтів CG та CNG), знаходяться метильовані CG-острівці, причому рівень метилювання даної ділянки зростає у диференційованих клітин порівняно з ембріональними стовбуровими. Значну концентрацію метильних міток у ДНК виявляють також з обох сторін ділянки, спільної для генів *Xist* та *Tsix*. А це свідчить про те, що дані острівці сайтів метилювання насправді функціонують.

ВИСНОВКИ. Послідовностям PHK-транскриптів генів *Xist* та *Tsix* людини властивий вкрай нерівномірний розподіл динуклеотидів CG та тринуклеотидів CNG, тобто сайтів, комплементарних сайтам метилювання у ДНК. На 5'-кінці *Xist*-транскрипту локалізуються найбільші, а на 5'-кінці *Tsix*-транскрипту – одні з найбільших піків концентрації сайтів CG та CNG для даних послідовностей. Ці дані підтверджують

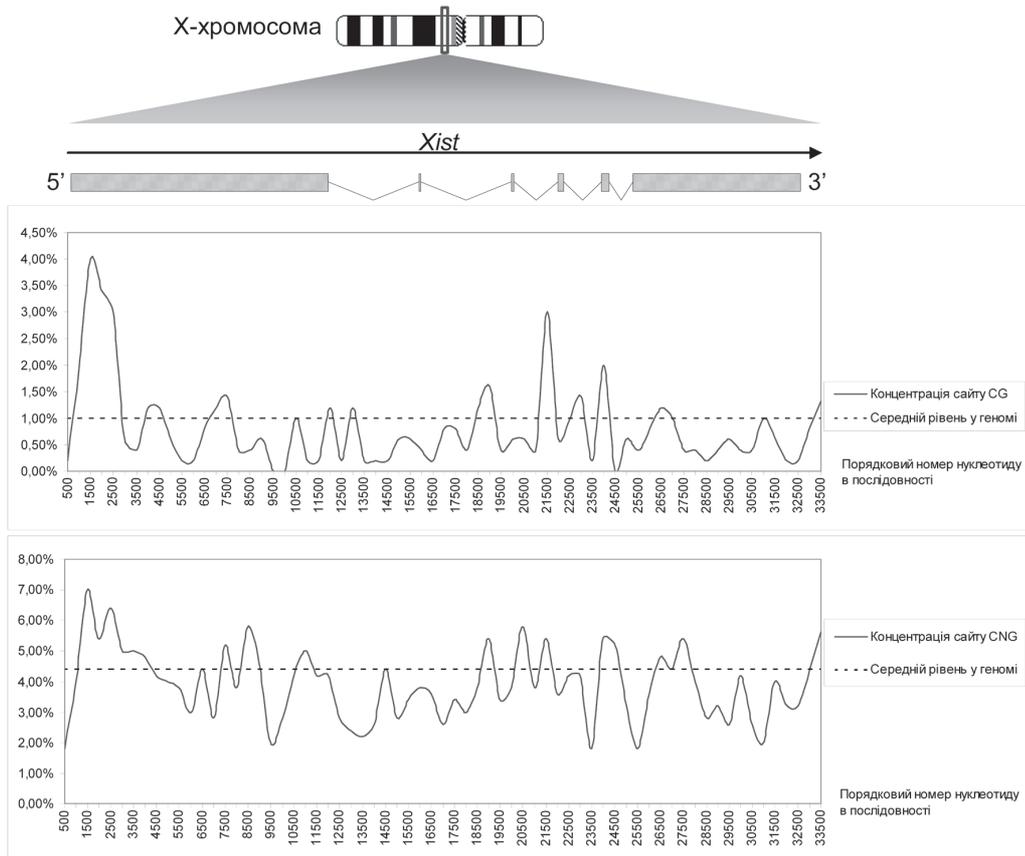


Рис. 2. Розподіл сайтів CG та CNG у транскрипті гена *Xist* (схему екзон-інтронної структури наведено у верхній частині рисунка, екзони позначено прямокутниками).

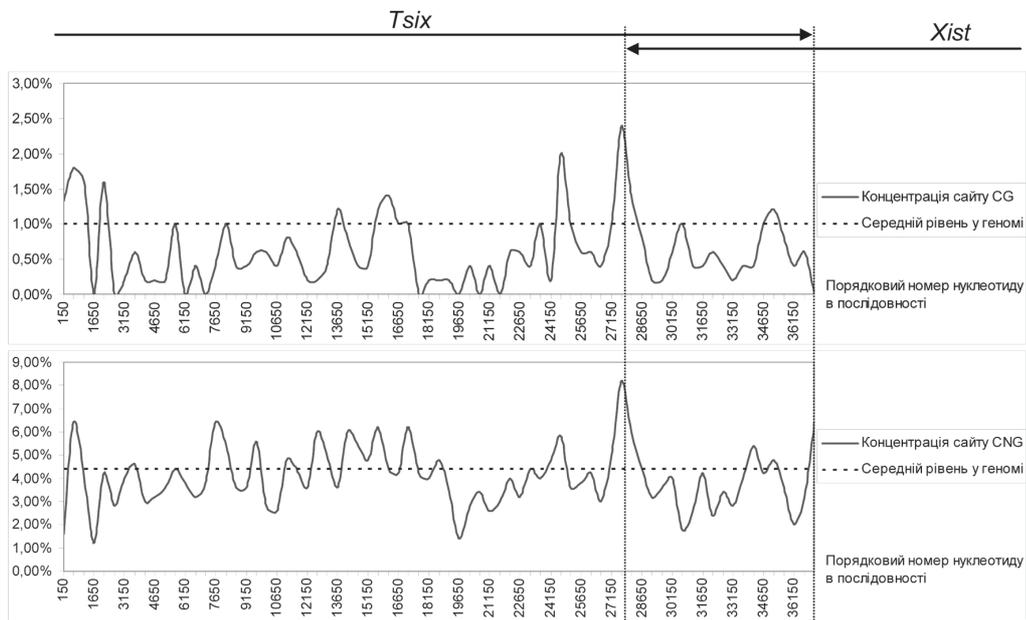


Рис. 3. Розподіл сайтів CG та CNG у транскрипті гена *Tsix* (пунктирними лініями обмежено ділянку перекриття послідовностей генів *Tsix* та *Xist*).

висунути нами гіпотезу про те, що Х-інактивація відбувається завдяки безпосередній гібридизації 5'-кінця транскрипту гена *Xist* з ДНК Х-хромосоми, що тягне за собою подальше метилювання *de novo* залишків цитозину в сай-

тах CG та CNG у тих послідовностях ДНК, які виявились комплементарно зв'язаними з *Xist*-транскриптом. Аналогічно репресія гена *Xist* може ініціюватись гібридизацією з його послідовністю 5'-кінця РНК-транскрипту гена *Tsix*.

СПИСОК ЛІТЕРАТУРИ

1. Галицький В. А. Локалізація сайтів 5'-CG-3', 5'-CNG-3', 5'-GC-3' та 5'-GNC-3' у послідовностях мікроРНК / В. А. Галицький, С. В. Комісаренко // *Biopolymers and Cell*. – 2011. – **27**, № 6. – С. 499–505.
2. Halytskyi V. A. Hypothesis of initiation of DNA methylation *de novo* and allelic exclusion by small RNAs / V. A. Halytskyi // *Cell Tissue Biol*. – 2008. – **2**, № 2. – P. 97–106.
3. Halytskyi V. A. Non-coding RNAs and epigenome: *de novo* DNA methylation, allelic exclusion and X-inactivation / V. A. Halytskyi, S. V. Komisarenko // *Ukrainian Biochem. J.* – 2013. – **85**, № 6. – P. 151–165.
4. Halytskyi V. RNA-dependent DNA methylation as a molecular basis of X-inactivation / V. Halytskyi // Joint Conference of HGM 2013 and 21st International Congress of Genetics “Genetics & Genomics of Global Health and Sustainability”, April 13–18, 2013. Abstract Book. – Singapore, 2013. – P. 282–283.
5. Jabbari K. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies / K. Jabbari, G. Bernardi // *Gene*. – 2004. – **333**. – P. 143–149.
6. Ogawa Y. Intersection of the RNA interference and X-inactivation pathways / Y. Ogawa, B. K. Sun, J. T. Lee // *Science*. – 2008. – **320**, № 5881. – P. 1336–1341.
7. Sado T. *Tsix* defective in splicing is competent to establish *Xist* silencing / T. Sado, Y. Hoki, H. Sasaki // *Development*. – 2006. – **133**, № 24. – P. 4925–4931.
8. Shibata S. Characterization and quantitation of differential *Tsix* transcripts: implications for *Tsix* function / S. Shibata, J. T. Lee // *Hum. Mol. Genet.* – 2003. – **12**, № 2. – P. 125–136.
9. The heterogeneity of thymine methyl group origin in DNA pyrimidine isostichs of developing sea urchin embryos / E. Scarano, M. Iaccarino, P. Grippo [et al.] // *Proc. Natl. Acad. Sci. USA*. – 1967. – **57**, № 5. – P. 1394–1400.

В. А. Галицький

ИНСТИТУТ БИОХИМИИ ИМЕНИ А. В. ПАЛЛАДИНА НАН УКРАИНЫ, КИЕВ

НЕРАВНОМЕРНОСТЬ ЛОКАЛИЗАЦИИ САЙТОВ CG И CNG В ТРАНСКРИПТАХ ГЕНОВ *XIST* И *TSIX*

Резюме

Хотя концентрация сайтов CG и CNG в *Xist* РНК в среднем составляет 0,8 %, то есть даже меньше, чем общий уровень в геноме (1 %), локализация этих сайтов крайне неравномерна. Концентрация сайтов CG и CNG достигает максимальной плотности, превышающей средний уровень в несколько раз, на 5'-конце *Xist*-транскрипта. Количество сайтов CG и CNG в *Tsix* РНК даже меньше, чем в *Xist*-транскрипте, и составляет всего 0,6 %. Тем не менее сайты CG и CNG в *Tsix* РНК концентрируются вблизи участка, перекрывающего последовательность гена *Xist*, а также на 5'-конце. Эти данные подтверждают гипотезу, что последовательности 5'-конца *Xist* РНК непосредственно связываются с комплементарной ДНК в других участках X-хромосомы, вызывая РНК-зависимое метилирование ДНК и X-инактивацию.

КЛЮЧЕВЫЕ СЛОВА: X-хромосома, *Xist*, *Tsix*, CG, CNG, локализация сайтов, РНК-зависимое метилирование ДНК, X-инактивация.

V. A. Halytskyi

O. V. PALLADIN INSTITUTE OF BIOCHEMISTRY OF NAS OF UKRAINE, KYIV

IRREGULARITY OF THE CG AND CNG SITE LOCALIZATION IN *XIST* AND *TSIX* GENE TRANSCRIPTS

Summary

Although the concentration of CG and CNG sites in *Xist* RNA averages 0.8 %, that is even less than the general genome level (1 %), the localization of these sites reveals extreme irregularity. CG and CNG concentration reaches maximum density, exceeding the average level in several times, at the 5' end of the *Xist* transcript. Frequency of CG and CNG sites in *Tsix* RNA is even less than in *Xist* transcript and amounts to 0.6 %. Nevertheless, CG and CNG in the *Tsix* RNA are abundant near the region, overlapping *Xist* sequence, as well as at the 5' end. These data support the hypothesis that the 5' end sequences of the *Xist* RNA directly bind with complementary DNA in remote regions of the X chromosome, triggering the RNA-dependent DNA methylation and X inactivation.

KEY WORDS: X chromosome, *Xist*, *Tsix*, CG, CNG, site localization, RNA-dependent DNA methylation, X inactivation.

Отримано 24.10.13

Адреса для листування: В. А. Галицький, Інститут біохімії імені О. В. Палладіна НАН України, вул. Леонтовича, 9, Київ, 01601, Україна, e-mail: volha@biochem.kiev.ua.