

БІОІНФОРМАТИВНИЙ АНАЛІЗ ПОТЕНЦІЙНИХ САЙТІВ ПОСТТРАНСЛЯЦІЙНИХ МОДИФІКАЦІЙ O⁶-МЕТИЛГУАНІН-ДНК МЕТИЛТРАНСФЕРАЗИ (MGMT) ЛЮДИНИ

А. П. ЯЦИШИНА¹, З. М. НІДОЄВА², О. В. ПІДПАЛА¹, Л. Л. ЛУКАШ¹

¹Інститут молекулярної біології та генетики НАН України, Київ;
e-mail: a.p.iatsyshyna@imbg.org.ua;

²Національний педагогічний університет імені М. П. Драгоманова, Київ, Україна

В амінокислотній послідовності протеїну O⁶-метилгуанін-ДНК метилтрансферази людини (MGMT) за допомогою аналізу *in silico* виявлено низку потенційних сайтів посттрансляційних модифікацій, зокрема ацетилювання залишку Gly3 N-кінця протеїну та внутрішніх залишків Lys132 і Lys135; метилування Arg166; сумоїлювання Lys63 та убіквітинування Lys31, Lys39, Lys49, Lys63, Lys67, Lys135, Lys156, Lys196, Lys209. Також передбачили 16 нових потенційних сайтів фосфорилування залишків серину (позиції 13, 124, 144, 182, 183, 190, 215, 216 і 230), тирозину (позиції 100 і 189), треоніну (позиції 23, 69, 94, 126 і 229) і п'ять сайтів зв'язування з кіназами та іншими протеїнами (Ser13 із 14-3-3, Val21 й Ile172 із D-доменом, Pro78 і Pro111 з SH₃-доменом, Pro111 із MAPK3). Деякі передбачені нами кінази виявлено раніше як партнери MGMT людини, що підтверджує ймовірність модифікації цих сайтів. Потенційні нові сайти потребують експериментального підтвердження модифікації та вивчення їх впливу на стабільність та репаративну активність цього ензиму.

Ключові слова: O⁶-метилгуанін-ДНК метилтрансфераза (MGMT), посттрансляційна модифікація протеїну, ацетилювання протеїну, метилування протеїну, сумоїлювання протеїну, убіквітинування протеїну, фосфорилування протеїну.

Ензим O⁶-метилгуанін-ДНК метилтрансфераза (MGMT) репарує пошкодження ДНК, при цьому він каталізує реакцію нуклеофільного приєднання алкільної групи O⁶-позиції гуаніну та O⁴-позиції тиміну до власного залишку цистеїну в активному центрі, інактивує та не спричиняючи розривів у молекулі ДНК [1, 2]. Незважаючи на мінорність таких пошкоджень ДНК, вони виявляють значний цитотоксичний, мутагенний та канцерогенний вплив на клітини [3]. Таким чином, MGMT забезпечує захист клітин від негативного впливу алкілування ДНК, однак, експресуючись у злоякісних клітинах, MGMT обмежує ефективність хіміотерапії із використанням алкілувальних сполук.

Показано, що експресія гену та активність самого ензиму MGMT мають широкі межі як між-, так і внутрішньоіндивідуальної варіації [1, 3–5], що вказує на складну регуляцію цього гену. Молекулярні основи варіації експресії MGMT, серед яких є генетичні та епігенетичні фактори, визначено недостатньо [5]. До епігенетичних факторів також належить регуляція активності та стабільності протеїнів за рахунок посттрансляційних модифікацій, наприклад, алкілування, ацетилювання, ме-

тилування, убіквітинування, фосфорилування, сумоїлювання та ін. Алкілування протеїну MGMT людини є найдослідженішою модифікацією, яка пов'язана з його репаративною функцією. Раніше вважали, що це єдина функція цього ензиму й після приєднання алкільної групи MGMT убіквітується та деградує за рахунок протеолізу [2]. Однак на сьогодні показано, що метилований протеїн MGMT людини може змінювати свою ензиматичну функцію на регуляторну транскрипційного супресора, беручи участь у контролі клітинної проліферації [6].

Експериментально показано, що протеїн MGMT людини фосфорилується за певними амінокислотними залишками [7–9] й убіквітується [10], проте сайти останньої модифікації не визначено. Фосфорилування протеїнів – важливий зворотній механізм регуляції активності протеїнів, зміни їхніх функцій, а також один з основних молекулярних механізмів передачі позаклітинних стимулів від рецепторів клітинної мембрани через компоненти сигнальних шляхів до відповідних компартментів клітини. Тобто передача сигналу від факторів росту, гормонів тощо опосередковується каскадом послідовних

фосфорилювань/дефосфорилювань сигнальних молекул клітини [11]. Показано, що фосфорилований рекомбінантний MGMT людини втрачає до 30–65% своєї репаративної активності, тоді як дефосфорилювання лужними фосфатазами відновлює його активність [7]. Припускають, що втрата ензиматичної активності модифікованого протеїну пов'язана з тим, що він убіквітинується та деградує [7]. Однак на сьогодні показано, що протеасома убіквітинозалежна деградація MGMT відбувається після репарації ДНК. Зокрема, приєднання алкільної групи до цистеїну в активному центрі ензиму призводить до конформаційних змін, які розпізнаються убіквітиною лігазою та завершуються деградацією протеїну [12]. На відміну від нього, фосфорилований MGMT шура виявляє підвищену стійкість до протеїназ [13]. Проте спостерігається брак даних щодо функціонального значення фосфорилювання MGMT.

Беручи до уваги те, що посттрансляційні модифікації протеїнів здатні не лише змінювати їхню активність, надавати інших функцій в клітині, а й регулювати протеїн-протеїнові взаємодії, ідентифікація нових сайтів модифікації та протеїнів-партнерів MGMT людини є актуальною. Це дасть змогу виявити нові неканонічні функції цього протеїну та нові функціональні зв'язки в різних процесах клітини. Тому метою цієї роботи був пошук потенційних сайтів посттрансляційних модифікацій протеїну MGMT людини з використанням методу аналізу *in silico*.

Матеріали і методи

Амінокислотну послідовність протеїну MGMT *Homo sapiens* (2.1.1.63) взяли з бази даних GenBank серверу NCBI (NP_002403.2, 238 амінокислотних залишків) [14]. Відомих партнерів цього протеїну шукали в базах даних APID [15], BIND (BOND) [16], BioGRID [17], 12D [18], Gene/NCBI [19], HPRD [20], MINT [21], PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>). Пошук ізоформ MGMT здійснювали за допомогою бази даних ASAP [22].

Для пошуку потенційних сайтів посттрансляційних модифікацій цього протеїну використали такі програми та бази даних: NetAcet 1.0 [23], PAIL [24] і Phosida [25] для передбачення сайтів ацетилювання, SUMOsp [26] і Phosida – сумоїлювання, UbPred [27] і BDM-PUB (<http://bdmpub.biocuckoo.org/prediction.php>) – убіквітинування, MeMo 2.0 [28] – метилювання; DISPHOS 1.3 [29], GPS 2.1

[30], KinasePhos [31], NetPhos 2.0 [32], NetPhosK 1.0 [33], Phosida, SCANSITE [34] – фосфорилювання протеїну. Для визначення доменної організації використали програму SMART (Simple Modular Architecture Research Tool) [35], а для пошуку неструктурованих ділянок як потенційних сайтів фосфорилювання та ділянок зв'язування з іншими протеїнами – DISPHOS, IUPred [36] і Anchor [37].

Результати та обговорення

У джерелах літератури і базах даних виявили понад 70 описаних партнерів протеїну MGMT людини [6, 7, 12, 38]. Серед них є протеїни, які беруть участь у формуванні хроматину (наприклад, гістони H2A, H2B, H3 і H4) в реплікації ДНК (топоізомераза I, протеїн-репарації помилкових спарювань нуклеотидів MSH-2 та ін.), в метаболізмі РНК (поліаденілатзв'язувальний протеїн 3, ядерний фосфопропротеїн нуклеолін та ін.) та в трансляції (фактори елонгації EF1 α 1, EF1 α 2, EF1 β та ін.) [6, 38]. Серед виявлених партнерів методами афінної очистки та мас-спектрометричного аналізу є значна кількість серин-треонінових і тирозінових кіназ та інших протеїнів, як наприклад 14-3-3, залучених у регуляції клітинного циклу та сигналіну [38]. Залучення протеїн кінази А, казеїн кінази II і протеїн кінази С у модифікації MGMT також показано в експериментах фосфорилювання рекомбінантного протеїну *in vitro*, а ендogenous – в клітинних екстрактах *in vivo* [7]. Низка виявлених партнерів функціонує в убіквітинзалежному протеолізі, наприклад, убіквітин, убіквітин лігаза, субодинаця 1 протеасоми 26S та інші [38]. Не менш цікавою є взаємодія MGMT із протеїнами клітинного метаболізму (наприклад, носій ADP/АТФ) та цитоскелета (актин β , тубулін α і β та ін.) [38].

Зважаючи на все різноманіття функцій виявлених партнерів, постає питання про можливість існування невідомих ще функцій ензиму репарації алкілювання ДНК MGMT.

Проведений біоінформативний аналіз бази даних ASAP на наявність ізоформ MGMT людини не виявив відомих альтернативних ізоформ цього протеїну.

Доменна організація протеїну MGMT людини та пошук потенційних ділянок зв'язування з іншими протеїнами. В амінокислотній послідовності протеїну MGMT людини довжиною 238 амінокислот за допомогою програми SMART визначили локалізацію двох доменів, зокрема метилтрансферазного (позиції залишків амінокислот 37-121)

і ДНК-зв'язувального (позиції залишків амінокислот 123-208) (рис. 1, *A, B*). Виявлені домени пов'язані з виконанням цього ензиму своєї функції репарації ДНК. Оскільки неструктуровані протеїни не мають чіткої енергетично стабільної третинної структури в нативному стані [36] та досягають стабілізації за рахунок взаємодії зі структурованим, глобулярним партнером [37], а власне в межах неструктурованих ділянок протеїнів переважно відбувається фосфорилювання [29], ми проаналізували амінокислотну послідовність MGMT на наявність таких ділянок із використанням IUPred і DISPHOS (рис. 1, *B*). Чим вище значення тенденції до формування неструктурованої ділянки, тим ймовірніше вона буде фосфорилювана за певними амінокислотними залишками (червона крива на рис. 1, *B*, тенденція до формування неструктурованої ділянки – на вісі ординат). За зниження пропускового порога до показника 0,3 збільшуються кількість та розміри потенційних до цієї модифікації ділянок (червона крива на рис. 1, *B*, поріг дорівнює 0,5). Для розпізнавання неструктурованих ділянок сервер IUPred використовує метод попарного оцінювання енергії [36].

Anchog також використовує цей підхід для передбачення неструктурованих ділянок і ділянок зв'язування з іншими протеїнами [37]. Як показано на рис. 1, *B* і *B*, MGMT людини є структурованим протеїном й поза виявленими доменами сервер Anchog передбачив дві ділянки зв'язування з іншими протеїнами, зокрема у положеннях 22–29 та 221–223 (синя крива на рис. 1, *B, B*).

Відомі посттрансляційні модифікації та їхні наслідки для протеїну MGMT. На сьогодні відомо вісім сайтів модифікації протеїну MGMT людини (табл. 1), серед яких сім сайтів фосфорилювання [7–9] та один – алкілювання [10, 12, 39, 40]. Серед цих сайтів фосфорилювання виявлено та підтверджено експериментально локалізацію лише трьох, зокрема Thr41, Thr42 і Ser232 [8, 9]. Решта сайтів (Ser45, Ser53, Ser70 і Tyr145) передбачено біоінформативним аналізом, тоді як експериментально показано лише саму можливість фосфорилювання рекомбінантного та ендогенного MGMT за амінокислотними залишками серину, тирозину і треоніну [7]. На жаль, їх локалізацію не було підтверджено в цьому дослідженні експериментально [7], хоча фосфорилювання амінокислотного залишку Thr42 також виявлено в іншій роботі [9].

На прикладі рекомбінантного MGMT показано, що фосфорилювання призводить до

зворотної інактивації репаративної функції цього ензиму, яка відновлюється *in vitro* після додавання лужної фосфатази [7]. Автори зазначеної роботи припустили, що фосфорилований MGMT втрачає ензиматичну активність, оскільки убіквітинується *in vivo* та деградує.

Функціональна інактивація цього ензиму настає після нуклеофільного приєднання алкільної групи до залишку Cys176 в активному центрі [10, 12, 40]. Це призводить до зміни структури протеїну, що краще розпізнається убіквітиною лігазою, яка мітить його для подальшої протеасомної деградації [12]. Однак сайти убіквітинування на сьогодні експериментально не визначено.

Інтригуючими є експериментальні дані, що алкілований MGMT людини необов'язково убіквітинується та деградує, а й може виконувати роль в регуляції транскрипції генів у відповідь на пошкодження ДНК [6]. Зокрема, показано, що модифікований протеїн може зв'язуватися з рецептором естрогену і ефективно пригнічувати його функцію як транскрипційного активатора генів, які сприяють клітинному поділу [6]. Виявлено, що він змінює свою конформацію та перемикає функцію ензиму репарації ДНК на транскрипційний регулятор, вказуючи на існування іншої неканонічної, тобто безпосередньо не пов'язаної з репарацією ДНК, функції. Алкілований MGMT людини виконує роль індукованого алкілюванням ДНК транскрипційного супресора, даючи можливість клітині відреагувати на алкілювальні сполуки [6].

Таким чином, посттрансляційні модифікації протеїну MGMT людини впливають на цей ензим та потребують детального вивчення.

Пошук потенційних сайтів ацетилювання. У клітинах еукаріот ацетилювання протеїнів може відбуватися як під час трансляції (ацетилювання N-кінця, або N-ацетилювання), так і посттрансляційно (ацетилювання залишків лізину). У першому випадку приєднання ацетильної групи є незворотною модифікацією та каталізується N-кінцевими ацетилтрансферазами [41].

Відомо, що N-ацетилювання є еволюційно консервативною та поширеною модифікацією протеїнів еукаріот (80–90 % протеїнів людини), проте біологічна роль такого типу ацетилювання мало вивчена [42]. Обговорюється його можливий вплив на функціональну активність протеїнів, а також роль у захисті протеїну від деградації або, навіть навпа-

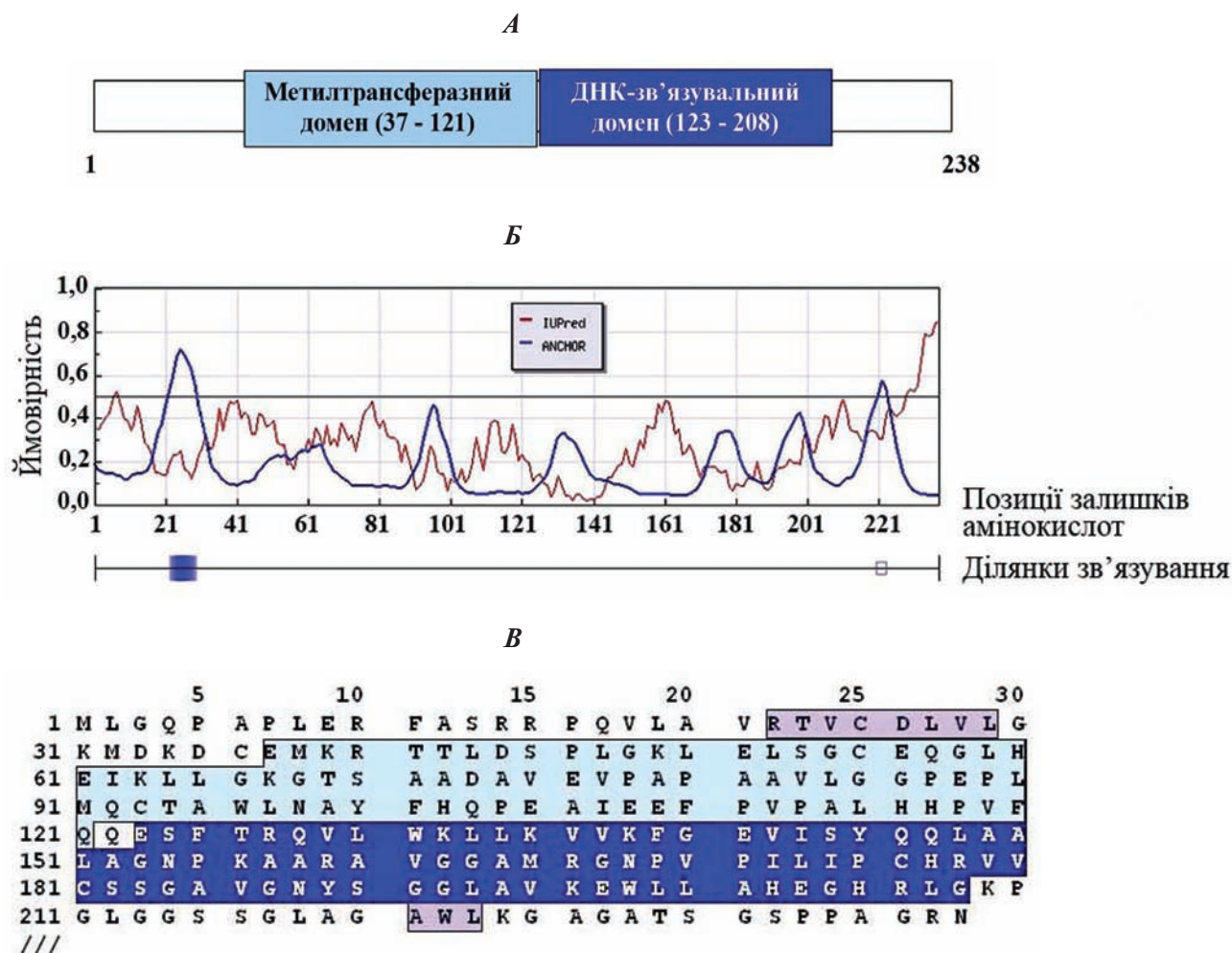


Рис. 1. Схематичне зображення доменної організації та потенційних ділянок взаємодії протеїну MGMT людини (NP_002403.2, 238 амінокислотних залишків). А – метилтрансферазний (37–121) та ДНК-зв’язувальний (123–208) домени визначені за допомогою програми SMART [35]. Б – крива червоного кольору ілюструє тенденцію до формування неструктурованих ділянок, передбачених сервером IUPred методом парного оцінювання енергії [36]. На вісі ординат – ймовірність, або тенденція, формування таких ділянок. Пропускний поріг дорівнює 0,5. Крива синього кольору ілюструє тенденцію до зв’язування MGMT з іншими протеїнами, визначену сервером Anchor [37]. Потенційні ділянки зв’язування (22–29; 221–223) ідентифіковано за пропускового порога 0,5. В – амінокислотна послідовність MGMT. Блакитним прямокутником виділено метилтрансферазний, синім – ДНК-зв’язувальний домен. Поза доменами світлобузковим кольором позначено дві ділянки взаємодії з іншими протеїнами

ки, мітки для специфічної деградації [41, 43]. Оскільки N-ацетилювання відбувається під час трансляції в переважній кількості протеїнів клітини та є незворотним процесом, зрозуміло, що ця модифікація відіграє певну специфічну регуляторну роль.

За допомогою сервера NetAcet, що використовує метод штучної нейронної мережі для віртуального скринінгу субстратів N-ацетилтрансферази, детектували можливість ацетилювання залишку Gly3 N-кінця MGMT людини (табл. 2), а пропуск-

ний поріг аналізу становив 0,3. Оскільки немає чіткого консенсусного мотиву для передбачення N-ацетилювання, можливості сервера обмежені скринінгом модифікацій залишків Ala, Gly, Ser і Thr у позиціях 1–3 [23].

Для передбачення ацетилювання внутрішніх лізину використали програму PAIL і базу даних Phosida (табл. 2). PAIL використовує алгоритм на основі Байєсівського дискримінантного правила [24], поріг для ймовірності ацетилювання залишків лізину в положеннях 31, 39, 49, 63,

Таблиця 1. Відомі сайти посттрансляційних модифікацій протеїну MGMT людини

Локалізація*	Тип модифікації	Роль модифікації
Thr41	Фосфорилювання	Не показано [9]
Thr42	Фосфорилювання	Пригнічення або інактивація активності [7]. Не показано [9]
Ser45	Фосфорилювання	Пригнічення або інактивація активності [7]
Ser53	Фосфорилювання	Пригнічення або інактивація активності [7]
Ser70	Фосфорилювання	Пригнічення або інактивація активності [7]
Tyr145	Фосфорилювання	Пригнічення або інактивація активності [7]
Ser232	Фосфорилювання	Пригнічення або інактивація активності [7]. Не показано [8]
Cys176	Алкідування	Інактивація репаративної функції та протеоліз [10, 12, 39, 40]
		Транскрипційний регулятор [6]

* Локалізація перерахована відповідно до протеїну MGMT довжиною 238 амінокислот (NP_002403.2). Модифікація залишку цистеїну в положенні 176 в активному центрі цього ензиму позначена блакитним кольором

Таблиця 2. Передбачені сайти ацетилювання в амінокислотній послідовності протеїну MGMT людини (NP_002403.2)

Сайт	База даних, програма		
	NetAcet 1.0	PAIL	Phosida
Gly3	+	–	–
Lys31	–	+	–
Lys39	–	+	–
Lys49	–	+	–
Lys63	–	+	–
Lys67	–	+	–
Lys132	–	+	+
Lys135	–	+	+
Lys138	–	+	–
Lys156	–	+	–
Lys196	–	+	–
Lys209	–	+	–

Наявність або відсутність потенційного сайту ацетилювання відповідно позначено «+» і «–»

67, 132, 135, 138, 156, 196 і 209 становив 0.5. База даних Phosida використовує віртуальний скринінг амінокислотної послідовності на наявність специфічних мотивів, тобто сайтів модифікації, у тому числі й ацетилювання [25]. За результатами цього аналізу виявили

такі потенційні сайти ацетилювання: Lys132 (ймовірність 91%) і Lys135 (імовірність 92%). Ці сайти також передбачено програмою PAIL (табл. 2).

Ацетилювання ϵ -аміногруп залишків лізину не є такою поширеною посттрансляційною модифікацією, як N-ацетилювання, але є зворотною модифікацією багатьох протеїнів за участю гістонової ацетилтрансферази та гістонової деацетилази [44]. Відомо, що зворотне ацетилювання залишку лізину може змінювати активність протеїну, найімовірніше, за рахунок зміни заряду самого залишку та відповідно електростатичних властивостей протеїну [44]. Цілком можливо, що залучаються й конформаційні зміни. Останнім часом також припускають, що ця модифікація може перехресно взаємодіяти з метилюванням, сумоїлюванням, убіквітинуванням, фосфорилюванням та іншим динамічним контролем передачі сигналів у клітині [41].

Ацетилювання лізину є важливим посттрансляційним механізмом регуляції протеїнів цитоскелета, молекулярних шаперонів, транскрипційних факторів, ефекторних протеїнів та ін., тому можна припустити, що ацетилювання лізину MGMT може надавати цьому ензиму інших, репаративних функцій в клітині. На користь підтвердження цього припущення є той факт, що алкілований протеїн MGMT людини може зв'язуватися з рецептором естрогену та репресувати його активність, функціонуючи як транскрипційний регулятор [6].

Передбачення сайтів сумоїлування та убіквітинування протеїну MGMT людини. Посттрансляційна модифікація сумоїлування полягає в зворотному ковалентному приєднанні SUMO (Small Ubiquitin-like Modifier) протеїнів до залишку лізину протеїну за участю низки ензимів [45]. SUMO – протеїни подібні до убіквітину, та й ензиматичний каскад сумоїлування аналогічний такому за убіквітинування, проте основною відмінністю є те, що сумоїлування не є міткою для деградації [45].

Молекулярні наслідки сумоїлування важко передбачити, але головний принцип його дії полягає у зміні внутрішньо- і/або міжмолекулярних взаємодій внаслідок появи чи маскуванню сайту взаємодії після конформаційних змін, механічного перешкодження взаємодії із протеїном-партнером або утворення додаткового сайту взаємодії [46]. Таким чином, зазначена вище модифікація впливає на стабільність, активність, функції та субклітинну локалізацію мішені, а також регулює протеїн–протеїнові взаємодії.

Із використанням бази даних Phosida і програми SUMOsp детектували потенційний спільний сайт сумоїлування MGMT людини – лізин у позиції 63 (табл. 3). Оскільки така посттрансляційна модифікація протеїнів регулює численні клітинні процеси, наприклад, ядерно-цитоплазматичний транспорт протеїнів, транскрипцію генів, сегрегацію хромосом, репарацію ДНК і відповідь на стрес, апоптоз і клітинний цикл, передачу сигналів, впливаючи на стабільність та активність протеїнів [45], подальше експериментальне дослідження цієї модифікації MGMT і вплив на його функції також має важливе значення для з'ясування неканонічних функцій алкілтрансферази.

Не менш важливим для протеїну є його убіквітинування, тобто зворотне ковалентне приєднання за допомогою ензимів одного або декількох мономерів і/або полімеру убіквітину до залишків лізину протеїну-мішені, змінюючи його активність, функції, локалізацію тощо [47]. Раніше вважали, що убіквітинування мітить протеїн для подальшого протеолізу, однак з'являється все більше даних про регуляторну роль цієї посттрансляційної модифікації в процесах проліферації, розвитку та диференціації клітини, репарації ДНК, реакції клітини на стрес та ін. [47, 48].

За результатами програм BDM-PUB і UbPred детектували такі спільні сайти убіквітинування залишків лізину MGMT

у позиціях 31, 39, 49, 63, 67, 135, 156, 196 і 209 (табл. 3). Для передбачення сайтів убіквітинування програма BDM-PUB, як і PAAL, використовує Байєсівський імовірнісний підхід та збалансоване вилучення з порогом передбачення 0,3. Під час вибору високої специфічності сайтів ця програма детектувала сайт Lys67 (пори́г 2,7; дані не наведено). Програма UbPred використовує алгоритм випадкового лісу, що поєднує ідею пакування Бреймана та випадковий вибір характеристик [27]. Поро́г детекції потенційних сайтів цієї програми становив 0,3.

Як видно з даних, наведених у табл. 2 і 3, виявлені сайти убіквітинування MGMT людини є конкурентними для сумоїлування та ацетилювання. Залишок Lys63 передбачено обома використаними інструментами як потенційний для убіквітинування та сумоїлування (табл. 3, синій фон), тоді як залишки лізину в позиціях 31, 39, 49, 63, 67, 132, 135, 138, 156, 196 і 209 є конкурентними для всіх трьох посттрансляційних модифікацій (табл. 2, 3). Показано, що багато протеїнів часто сумоїлюються та убіквітуються за однаковими сайтами, однак дані модифікації не конкурують напряду за сайт, оскільки частина молекул із пулу цього протеїну в клітині може бути убіквітована, решта – сумоїлована [45]. Також показано, що убіквітинування та сумоїлування можуть стосуватися як різних сайтів, так одних і тих самих, виявляючи своєрідну взаємодію. Наприклад, сумоїлований протеїн може сприяти його фосфорилюванню, а останнє може бути необхідною умовою для убіквітинування тих самих сайтів [45]. Перемикання між різними модифікаціями однакових залишків лізину в протеїні-мішені показано й для сумоїлування і ацетилювання, що також може залежати від фосфорилювання [45].

Пошук *in silico* потенційних сайтів метилювання протеїну MGMT людини. Для біоінформативного аналізу потенційних сайтів метилювання залишків аргініну та лізину використали сервер MeMo, який базується на методі опорних векторів, що аналізує дані та розпізнає шаблони для класифікації та регресійного аналізу [28]. У дослідженому протеїні за допомогою MeMo детектували один потенційний сайт метилювання залишку аргініну в позиції 166. Експериментально показано лише метилювання залишку цистеїну активного центру (вплив на ензим MGMT описано вище). Найкраще вивчено метилювання гістонів. Показано, що воно призводить до змін хроматину, та, як наслідок, до

Таблиця 3. Передбачені сайти сумоїлування та убіквітинування в амінокислотній послідовності MGMT людини (NP_002403.2)

Сайт	База даних, програма			
	сумоїлування		убіквітинування	
	Phosida	SUMOsp	BDM-PUB	UbPred
Lys31	—	+	+	+
Lys34	—	+	—	—
Lys39	—	+	+	+
Lys49	—	+	+	+
Lys63	+	+	+	+
Lys67	—	+	+	+
Lys132	—	+	+	—
Lys135	—	+	+	+
Lys138	—	+	—	+
Lys156	—	+	+	+
Lys196	—	+	+	+
Lys209	—	+	+	+
Lys224	—	+	—	+

Наявність або відсутність потенційних сайтів сумоїлування та убіквітинування відповідно позначено «+» і «-»

епігенетичної репресії або активації генної експресії [49]. Чи метилується залишок Arg166 MGMT та, якщо так, як ця модифікація впливає на його активність? Ці питання потребують експериментального аналізу.

Передбачення сайтів фосфорилювання.

Однією з найпоширеніших посттрансляційних модифікацій протеїнів є фосфорилювання, тобто ензиматичне зворотне приєднання фосфатної групи до гідроксильної групи бічного ланцюга серину, тирозину та треоніну (так зване O-фосфорилювання) та рідше до гістидину та аргініну (N-фосфорилювання) чи аспрату і глутамату (A-фосфорилювання) [50, 51]. Приєднання негативно заряджених фосфатних груп призводить до конформаційних змін протеїну, які, в свою чергу, можуть змінювати його активність, функції.

Для передбачення сайтів фосфорилювання досліджуваного протеїну використали низку баз даних і програм. У табл. 4 наведено результати аналізу за допомогою семи різних пошукових інструментів, а саме: DISPHOS, GPS, KinasePhos, NetPhos, NetPhosK, Phosida і SCANSITE. Програма DISPHOS передбачає сайти фосфорилювання, використовуючи частоти специфічних до певної позиції амінокислот й інформацію про неструктуровані

ділянки [29]. GPS використовує алгоритм рейтингу мотивів, а також ієрархічно класифіковані протеїнові кінази для пошуку потенційних сайтів фосфорилювання та кіназ [30]. На відміну від них KinasePhos застосовує приховану модель Маркова для кожної групи послідовностей довкола амінокислотного залишку, базуючись на відомих сайтах фосфорилювання певними кіназами [31]. Передбачення сайтів цієї модифікації методом штучної нейронної мережі залежно від послідовності та структури протеїнів еукаріот проводиться серверами NetPhos і NetPhosK [32, 33]. На відміну від решти інструментів, використовуючи які можна передбачити потенційні до фосфорилювання сайти в амінокислотній послідовності, SCANSITE також передбачає мотиви зв'язування із такими доменами, як SH₂-, PDZ-, із доменами 14-3-3 та ін. [34]. Для пошуку мотивів програма використовує ентропійний підхід, що оцінює ймовірність відповідності сайту певному мотиву, використовуючи селективність значень кожної амінокислоти в досліджуваній послідовності.

Внаслідок проведеного аналізу амінокислотної послідовності MGMT людини виявили 23 потенційні сайти фосфорилювання, серед яких сім вже відомих. Зокрема,

Таблиця 4. Передбачені сайти фосфорилювання в амінокислотній послідовності MGMT людини (NP_002403.2)

Сайт	Програма						
	SCANSITE	GPS 2.1	Phosida	DISPHOS 1.3	KinasePhos	NetPhos2.0 (S, T, Y)	NetPhosK1.0
Ser13	+	+	+	+	+	+	–
Thr23	–	+	–	–	–	+	–
Thr41 [9]	–	+	+	–	–	+/- *	+
Thr42 [7,9]	+	+	+	–	+	+	+
Ser45 [7]	+	+	+	+/- *	+	+	+
Ser53 [7]	–	+	+	+/- *	–	+	+
Thr69	+	+	–	–	–	+/- *	–
Ser70 [7]	–	+	+	–	–	+	–
Thr94	–	+	–	–	–	? *	–
Tyr100	–	+	–	–	–	? *	–
Ser124	+	+	–	–	–	+/- *	–
Thr126	–	+	–	–	–	? *	–
Ser144	–	+	–	–	–	? *	–
Tyr145 [7]	–	+	+	–	–	+/- *	–
Ser182	–	–	+	–	–	–	+
Ser183	–	+	+	–	–	? *	–
Tyr189	–	+	–	–	–	? *	–
Ser190	–	+	–	–	–	? *	–
Ser215	–	+	+	+/- *	–	–	–
Ser216	–	+	+	+/- *	–	? *	+
Thr229	–	+	–	+/- *	–	+/- *	+
Ser230	+	+	+	+/- *	–	+	+
Ser232 [7,8]	–	+	+	+	+	+	+

* Низький рейтинговий показник. Наявність або відсутність потенційного сайту фосфорилювання відповідно позначено «+» і «–»; «+/-» і «?» – наявність потенційного сайту фосфорилювання малоймовірна внаслідок низького рейтингового показника

можливість фосфорилювання залишків серину в позиції 70 [7] і тирозину 145 [7] підтвердили із використанням трьох різних програм (GPS, NetPhos і Phosida), треоніну 41 [9] – чотирьох (GPS, NetPhos, NetPhosK і Phosida), серину 53 [7] – п'яти (DISPHOS, GPS, NetPhos, NetPhosK і Phosida), треоніну 42 [7, 9] і серину 232 [7, 8] – шести, а серину 45 [7] – усіх семи використаних програм (табл. 4). Фосфорилювання залишку треоніну в позиції 42 передбачили за допомогою інструментів GPS, KinasePhos, NetPhos, NetPhosK, Phosida і SCANSITE, для серину 232 позитивний результат надала програма DISPHOS, а SCANSITE не детектувала можливість цієї модифікації (табл. 4).

На рис. 2 наведено результати пошуку за допомогою програми SCANSITE потенційних сайтів фосфорилювання та зв'язування зі специфічними доменами. Серед них Pro78 і Pro111 є потенційними сайтами зв'язування з SH₃-доменом, а Val21 й Ile172 – із D-доменом партнерів, тоді як для сайту Pro113 зазначено лише можливість зв'язування із мітогенактивованою протеїн кіназою 3. Заслуговує на увагу той факт, що для сайту Ser13 ця програма передбачила зв'язування із протеїнами 14-3-3, фізичну взаємодію MGMT з яким показано в дослідженні [38].

Серед детектованих нами нових сайтів найпотенційнішими до такої модифікації є

ті, що локалізовані поза метилтрансферазним (37-121: Ser13 і Thr23) і ДНК-зв'язувальним (123-208: Ser215, Ser216, Thr229 і Ser230) доменами протеїну, визначеними програмою SMART (рис. 1, А, В), а також у ділянках із

внутрішньо нестабільною структурою, визначеними IUPred та DISPHOS (рис. 1, Б, табл. 4). Варто зазначити, що Val21 локалізований в мотиві зв'язування з D-доменом MAPK1, який імовірно перекривається з ділянкою

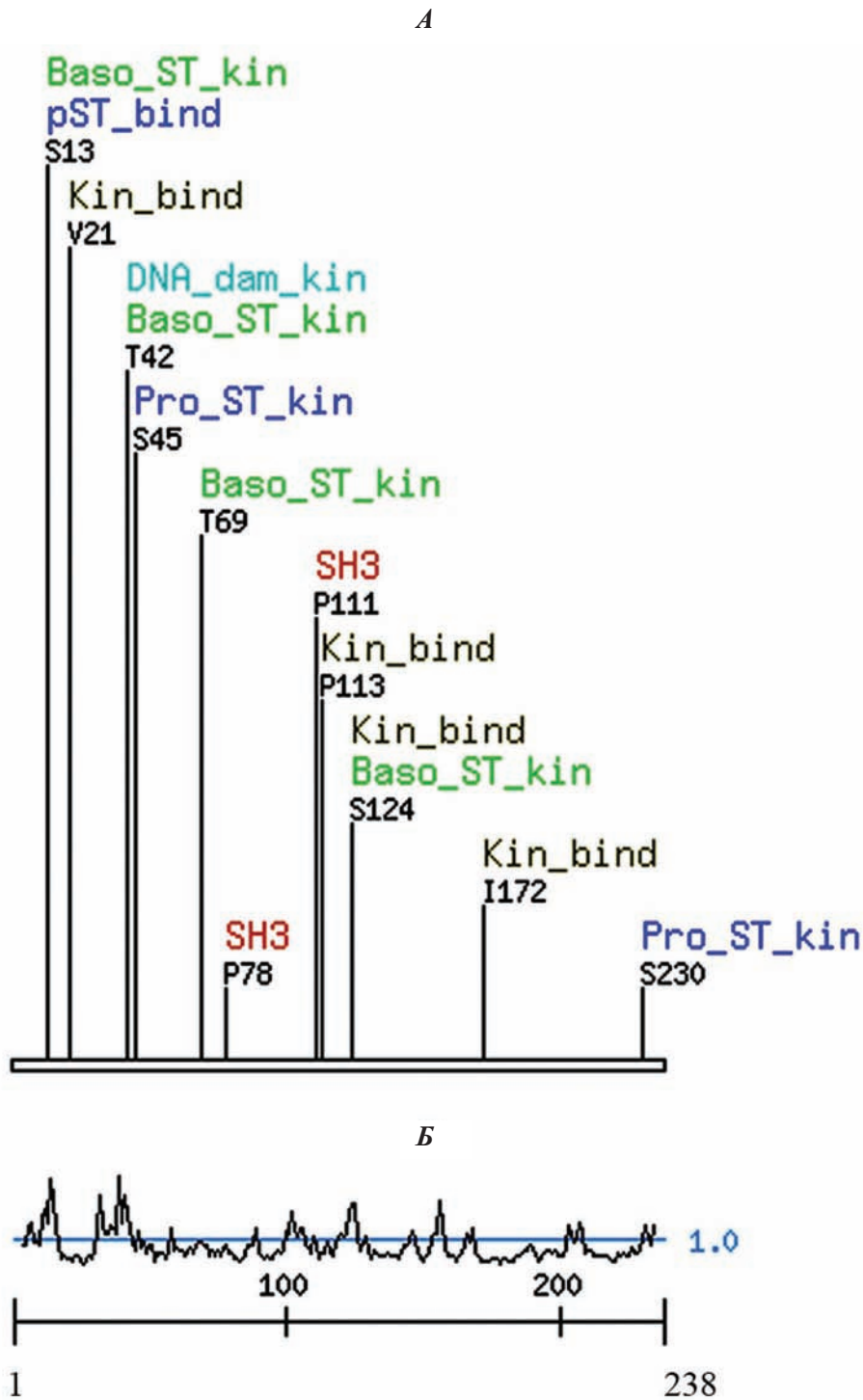


Рис. 2. Потенційні сайти фосфорилювання та зв'язування зі специфічними доменами в протеїні MGMT людини (NP_002403.2, 238 амінокислотних залишків), передбачені із використанням програми SCANSITE: А – передбачені сайти, Б – доступність амінокислотного залишку до поверхні протеїну

зв'язування із протеїнами (22–29), яку передбачено програмою Anchor (рис. 1, Б, В). Два нові потенційні сайти фосфорилювання (Ser13 і Ser230) визначені з використанням шести різних інструментів (DISPHOS, GPS, KinasePhos, NetPhos, Phosida і SCANSITE для Ser13; DISPHOS, GPS, NetPhos, NetPhosK, Phosida і SCANSITE для Ser230; табл. 4), тому саме вони є першими кандидатами для подальшого експериментального аналізу.

Не лише різні використані нами біоінформативні інструменти, а й також результати досліджень [7–9, 38] підтверджують імовірність модифікації передбачених сайтів. Наприклад, показано, що протеїн кіназа А, протеїн кіназа С і кazeїн кіназа ІІ фосфорилюють MGMT [7], а серед передбачених нами сайтів і кіназ, що їх модифікують, щонайменше два використані біоінформативні інструменти детектували залучення їх до модифікації сайтів Ser13 (протеїн кіназа С; GPS і Phosida), Thr42 (протеїн кіназа А; GPS, NetPhosK, Phosida і SCANSITE), Ser53 (kazeїн кіназа ІІ; NetPhosK і Phosida), Ser124 (протеїн кіназа С; GPS і SCANSITE). Фосфорилювання залишку Thr42 показано експериментально [7, 9]. За даними іншого дослідження MGMT людини зв'язується і з іншими кіназами, крім вже зазначеної протеїн кінази С [38]. Серед тих із них, які виявлено і нами біоінформативно, є ДНК залежна протеїн кіназа, циклінзалежна кіназа 2 (сайт Ser45; GPS і SCANSITE), мітогенактивована протеїн кіназа (сайт Ser45; GPS, NetPhosK і SCANSITE) та циклінзалежна кіназа 5 (сайт Ser45; GPS і SCANSITE).

Варто зазначити також, що виявлені серед партнерів MGMT тирозінова протеїн фосфатаза та інгібітор фосфатаз SET [38] підтверджують зворотний процес фосфорилювання цього протеїну в клітинах.

Відомо, що фосфорилювання протеїнів відіграє важливу роль у регуляції таких клітинних процесів, як метаболізм, сигналінг, диференціація, мембранний транспорт і т.д. [11]. Для протеїну MGMT людини показано, що його фосфорилювання зменшує ензиматичну активність, а дефосфорилювання, навпаки, збільшує, на основі чого припущено роль фосфорилювання в контролі активності репаративної функції ензиму в нормальних і ракових клітинах [7]. Показано, що фосфорилований MGMT щура має підвищену стійкість до протеолітичної деградації протеазами печінки порівняно з немодифікованим протеїном [13], тоді як фосфорилована метилтрансфераза людини втрачає активність, що

спекулятивно пов'язують з убіквітінуванням модифікованого протеїну та його протеолізом [7]. Це припущення необхідно підтвердити експериментально, однак воно цілком заслуговує на увагу, оскільки відомо, що від фосфорилювання протеїнів залежить їхнє убіквітінування та сумоїлування [45].

Отже, функціональну роль фосфорилювання MGMT на сьогодні вивчено недостатньо. В експериментах *in vitro* та в живих клітинах показано фосфорилювання MGMT людини та втрату ним репаративної функції, тоді як під час вивчення динаміки фосфорилювання протеома ядра у відповідь на пошкодження ДНК іонізуючим випромінюванням виявлено, що цей протеїн зазнає раннього фосфорилювання, ймовірно, кіназою АТМ [52]. На жаль, регуляторний вплив цієї модифікації MGMT невідомий. Нез'ясованим лишається також, чи впливає фосфорилювання цього протеїну на його подальші регуляторні посттрансляційні модифікації, наприклад, ацетилювання, убіквітінування, сумоїлування, а якщо так, то як саме.

Таким чином, за допомогою аналізу *in silico* в межах амінокислотної послідовності протеїну MGMT людини виявили низку потенційних сайтів посттрансляційних модифікацій, зокрема сайти ацетилювання, метилювання, сумоїлування, убіквітінування та фосфорилювання. Передбачені нами сайти потребують подальшого експериментального підтвердження модифікації та вивчення їхнього впливу на стабільність та репаративну активність MGMT.

БИОИНФОРМАТИВНЫЙ АНАЛИЗ ПОТЕНЦИАЛЬНЫХ САЙТОВ ПОСТТРАНСЛЯЦИОННЫХ МОДИФИКАЦИЙ O⁶-МЕТИЛГУАНИН-ДНК МЕТИЛТРАНСФЕРАЗЫ (MGMT) ЧЕЛОВЕКА

А. П. Яцышина¹, З. М. Нидоева²,
О. В. Підпала¹, Л. Л. Лукаш¹

¹Институт молекулярной биологии и генетики НАН Украины, Киев;
e-mail: a.p.iatsyshyna@imbg.org.ua;

²Национальный педагогический университет им. Н. П. Драгоманова, Киев, Украина

Используя анализ *in silico*, в последовательности протеина O⁶-метилгуанин-ДНК метилтрансферазы человека (MGMT) обнаружен ряд потенциальных сайтов посттрансляционных модификаций, в частности аце-

тилювання остатка Gly3 N-конца протеїна і внутрішніх остатків Lys132 і Lys135; метилювання Arg166; сумоїлювання Lys63 і убиквітиніювання Lys31, Lys39, Lys49, Lys63, Lys67, Lys135, Lys156, Lys196, Lys209. Також передбачили 16 нових потенціальних сайтів фосфорилювання остатків серина (позиції 13, 124, 144, 182, 183, 190, 215, 216 і 230), тирозина (позиції 100 і 189), треоніна (позиції 23, 69, 94, 126 і 229) і п'ять сайтів зв'язування з киназами і іншими протеїнами (Ser13 з 14-3-3, Val21 і Ile172 з D-доменом, Pro78 і Pro111 з SH₃-доменом, Pro111 з MAPK3). Деякі передбачені нами кинази є відомими партнерами MGMT людини, що підтверджує ймовірність модифікації даних сайтів. Потенціальні нові сайти потребують експериментального підтвердження модифікації і вивчення їх впливу на стабільність і репаративну активність даного ферменту.

Ключевые слова: O⁶-метилгуанін-ДНК метилтрансфераза (MGMT), посттрансляційна модифікація протеїна, ацетилювання протеїна, метилювання протеїна, сумоїлювання протеїна, убиквітиніювання протеїна, фосфорилювання протеїна.

BIOINFORMATIC ANALYSIS OF POTENTIAL POST-TRANSLATIONAL MODIFICATION SITES OF THE HUMAN O⁶-METHYLGUANINE-DNA METHYLTRANSFERASE (MGMT) PROTEIN

A. P. Iatsyshyna¹, Z. M. Nidoieva²,
O. V. Pidpala¹, L. L. Lukash¹

¹Institute of Molecular Biology and Genetics,
National Academy of Sciences of Ukraine, Kyiv;
e-mail: a.p.iatsyshyna@imbg.org.ua;

²M. P. Drahomanov National Pedagogical
University, Kyiv, Ukraine

S u m m a r y

Using *in silico* analysis a number of potential sites for post-translational modifications has been revealed within the human O⁶-methylguanine-DNA methyltransferase (MGMT) protein. In particular these were the acetylation of Gly3 residue in the N-terminus of protein and internal residues Lys132 and Lys135; Arg166 residue methylation; Lys63 SUMOylation and ubiquitination of Lys31, Lys39, Lys49, Lys63, Lys67, Lys135, Lys156, Lys196, Lys209. Also it has been predicted 16 novel potential phosphorylation sites of serine residues (positions 13, 124, 144, 182, 183, 190, 215, 216 and

230), tyrosine residues (positions 100 and 189) and threonine residues (positions 23, 69, 94, 126 and 229), as well as five binding sites for kinases and other proteins (Ser13 with 14-3-3, Val21 and Ile172 with D-domain, Pro78 and Pro111 with SH₃-domain, Pro111 with MAPK3). Some kinases predicted by the authors are known as partners of the MGMT protein, that confirms the probability of modification of the given sites. Potential sites require further experimental confirmation of modifications and investigation of their influence on stability and DNA-repair activity of this protein.

Key words: O⁶-methylguanine-DNA methyltransferase (MGMT), post-translational protein modification, protein acetylation, protein methylation, protein SUMOylation, protein ubiquitination, protein phosphorylation.

1. Kaina B., Christmann M., Naumann S., Roos W. P. // DNA Repair (Amst). – 2007. – 6, N 8. – P. 1079–1099.
2. Verbeek B., Southgate T. D., Gilham D. E., Margison G. P. // Br. Med. Bull. – 2008. – 85, N 1. – P. 17–33.
3. Pegg A. E. // Cancer Invest. – 1984. – 2, N 3. – P. 223–231.
4. Gerson S. L., Trey J. E., Miller K., Berger N. A. // Carcinogenesis. – 1986. – 7, N 5. – P. 745–749.
5. Margison G. P., Povey A. C., Kaina B., Santibáñez Koref M. F. // Ibid. – 2003. – 24, N 4. – P. 625–635.
6. Teo A. K., Oh H. K., Ali R. B., Li B. F. // Mol. Cell Biol. – 2001. – 21, N 20. – P. 7105–7114.
7. Srivenugopal K. S., Mullapudi S. R., Shou J. et al. // Cancer Res. – 2000. – 60, N 2. – P. 282–287.
8. Beausoleil S. A., Jedrychowski M., Schwartz D. et al. // Proc. Natl. Acad. Sci. USA. – 2004. – 101, N 33. – P. 12130–12135.
9. Imami K., Sugiyama N., Kyono Y. et al. // Anal. Sci. – 2008. – 24, N 1. – P. 161–166.
10. Srivenugopal K. S., Yuan X. H., Friedman H. S., Ali-Osman F. // Biochemistry. – 1996. – 35, N 4. – P. 1328–1334.
11. Ciesla J., Fraczyk T., Rode W. // Acta Biochim. Pol. – 2011. – 58, N 2. – P. 137–148.
12. Xu-Welliver M., Pegg A. E. // Carcinogenesis. – 2002. – 23, N 5. – P. 823–830.
13. Lim I. K., Park T. J., Paik W. K. // Biochem. J. – 2000. – 352, Pt 3. – P. 801–808.
14. Benson D. A., Karsch-Mizrachi I., Clark K. et al. // Nucleic Acids Res. – 2012. – 40, Database Issue. – D48–53.
15. Prieto C., De Las Rivas J. // Ibid. – 2006. – 34, Web Server Issue. – W298–302.

16. *Alfarano C., Andrade C. E., Anthony K. et al. // Ibid.* – 2005. – **33**, Database Issue. – D418–424.
17. *Stark C., Breitzkreutz B.J., Chatr-Aryamontri A. et al. // Ibid.* – 2011. – **39**, Database Issue. – D698–704.
18. *Brown K. R., Jurisica I. // Genome Biol.* – 2007. – **8**, N 5. – R95.
19. *Maglott D., Ostell J., Pruitt K. D. et al. // Nucleic Acids Res.* – 2011. – **39**, Database issue. – D52–57.
20. *Keshava Prasad T. S., Goel R., Kandasamy K. et al. // Ibid.* – 2009. – **37**, Database issue. – D767–772.
21. *Licata L., Briganti L., Peluso D. et al. // Ibid.* – 2012. – **40**, Database issue. – D857–861.
22. *Lee C., Atanelov L., Modrek B. et al. // Ibid.* – 2003. – **31**, N 1. – P. 101–105.
23. *Kiemer L., Bendtsen J. D., Blom N. // Bioinformatics.* – 2005. – **21**, N 7. – P. 1269–1270.
24. *Li A., Xue Y., Jin C. et al. // Biochem. Biophys. Res. Commun.* – 2006. – **350**, N 4. – P. 818–824.
25. *Gnad F., Gunawardena J., Mann M. // Nucleic Acids Res.* – 2011. – **39**, Database issue. – D253–260.
26. *Ren J., Gao X., Jin C. et al. // Proteomics.* – 2009. – **9**, N 12. – P. 3409–3412.
27. *Radivojac P., Vacic V., Haynes C. et al. // Proteins.* – 2010. – **78**, N 2. – P. 365–380.
28. *Chen H., Xue Y., Huang N. et al. // Nucleic Acids Res.* – 2006. – **34**, Web Server issue. – W249–253.
29. *Uversky V. N., Radivojac P., Iakoucheva L. M. et al. // Methods Mol. Biol.* – 2007. – **408**. – P. 69–92.
30. *Xue Y., Ren J., Gao X. et al. // Mol. Cell Proteomics.* – 2008. – **7**, N 9. – P. 1598–1608.
31. *Huang H. D., Lee T. Y., Tzeng S. W. et al. // Nucleic Acids Res.* – 2005. – **33**, Web Server issue. – W226–229.
32. *Blom N., Gammeltoft S., Brunak S. // J. Mol. Biol.* – 1999. – **294**, N 5. – P. 1351–1362.
33. *Blom N., Sicheritz-Pontén T., Gupta R. et al. // Proteomics.* – 2004. – **4**, N 6. – P. 1633–1649.
34. *Obenauer J. C., Cantley L. C., Yaffe M. B. // Nucleic Acids Res.* – 2003. – **31**, N 13. – P. 3635–3641.
35. *Ponting C. P., Schultz J., Milpetz F. et al. // Ibid.* – 1999. – **27**, N 1. – P. 229–232.
36. *Dosztányi Z., Csizmyk V., Tompa P. et al. // J. Mol. Biol.* – 2005. – **347**, N 4. – P. 827–839.
37. *Dosztányi Z., Mészáros B., Simon I. // Bioinformatics.* – 2009. – **25**, N 20. – P. 2745–2746.
38. *Niture S. K., Doneanu C. E., Velu C. S. et al. // Biochem. Biophys. Res. Commun.* – 2005. – **337**, N 4. – P. 1176–1184.
39. *Pegg A.E., Wiest L., Mummert C. et al. // Carcinogenesis.* – 1991. – **12**, N 9. – P. 1679–1683.
40. *Rasimas J. J., Dalessio P. A., Ropson I. J. et al. // Protein Sci.* – 2004. – **13**, N 1. – P. 301–305.
41. *Mogk A., Bukau B. // Science.* – 2010. – **327**, N 5968. – P. 966–967.
42. *Yang X. J., Seto E. // Mol. Cell.* – 2008. – **31**, N 4. – P. 449–461.
43. *Hwang C. S., Shemorry A., Varshavsky A. // Science.* – 2010. – **327**, N 5968. – P. 973–977.
44. *Glozak M. A., Sengupta N., Zhang X. et al. // Gene.* – 2005. – **363**. – P. 15–23.
45. *Wilkinson K. A., Henley J. M. // Biochem. J.* – 2010. – **428**, N 2. – P. 133–145.
46. *Geiss-Friedlander R., Melchior F. // Nat. Rev. Mol. Cell Biol.* – 2007. – **8**, N 12. – P. 947–956.
47. *Ciechanover A., Iwai K. // IUBMB Life.* – 2004. – **56**, N 4. – P. 193–201.
48. *Schaefer A., Nethe M., Hordijk P. L. // Biochem J.* – 2012. – **442**, N 1. – P. 13–25.
49. *Grewal S. I., Rice J. C. // Curr. Opin. Cell Biol.* – 2004. – **16**, N 3. – P. 230–238.
50. *Arena S., Benvenuti S., Bardelli A. // Cell Mol. Life Sci.* – 2005. – **62**, N 18. – P. 2092–2099.
51. *Hay R. T. // Mol. Cell.* – 2005. – **18**, N 1. – P. 1–12.
52. *Bennetzen M. V., Larsen D. H., Bunkenborg J. et al. // Mol. Cell Proteomics.* – 2010. – **9**, N 6. – P. 1314–1323.

Отримано 04.04.2012