

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ  
ІМЕНІ ТАРАСА ШЕВЧЕНКА  
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ВИСОКИХ ТЕХНОЛОГІЙ

Завідувач кафедри молекулярної біотехнології  
та біоінформатики  
доц. Нипорко Олексій Юрійович  
Протокол №\_\_\_\_ засідання кафедри  
від «\_\_» \_\_\_\_\_ 2024 р.

**ВПЛИВ ФАКТОРІВ НАВКОЛИШНЬОГО  
СЕРЕДОВИЩА НА SNP У ГЕНІ *BRCA1* ТА  
ПОЗИТИВНА СЕЛЕКЦІЯ НА АМІНОКИСЛОТНІЙ  
ПОЗИЦІЇ 1443**

Випускна кваліфікаційна робота магістра  
Студентки спеціальності 091  
Біологія  
ОП «Біоінформатика  
та структурна біологія»  
**Гордовської Наталії Владиславівни**

Науковий керівник:  
доцент кафедри молекулярної  
біотехнології та біоінформатики  
к.б.н. **Нипорко Олексій Юрійович**

Оцінка захисту роботи

---

## Анотація

Гордовська Н.В. Вплив факторів навколишнього середовища на SNP у гені *BRCA1* та позитивна селекція на амінокислотній позиції 1443. – Кваліфікаційна робота магістра за спеціальністю 091 Біологія ОП «Біоінформатика та структурна біологія».

У роботі проведено дослідження впливу вірусів на еволюцію гену *BRCA1*. Аналізовано роль однонуклеотидного поліморфізму у гені *BRCA1* та його вплив на стабільність білкового продукту. Встановлено, що мутація на даній позиції може бути асоційована з позитивною селекцією в результаті взаємодії з онкогенними вірусами. Отримані результати можуть бути використані для розробки нових підходів у генетичному скринінгу та розумінні механізмів онкогенезу.

**Ключові слова:** ген *BRCA1*; SNP; вірусно-опосередований онкогенез; позитивна селекція.

## Abstract

Hordovska N.V. Influence of environmental factors on SNPs in *BRCA1* and positive selection at amino acid position 1443. Master's thesis in the specialty 091 Biology of the education program "Bioinformatics and structural biology".

The work investigates the influence of viruses on the evolution of the *BRCA1* gene. The role of single nucleotide polymorphism in the *BRCA1* gene and its influence on the stability of the protein product were analyzed. It was found that a mutation at this position can be associated with positive selection as a result of interaction with oncogenic viruses. The results obtained can be used to develop new approaches in genetic screening and understanding of oncogenesis mechanisms.

**Keywords:** *BRCA1*; SNP; virus-mediated oncogenesis; positive selection.

# Зміст

<b>Вступ</b>	<b>4</b>
<b>Літературний огляд</b>	<b>6</b>
1.1 Теорії генної еволюції . . . . .	6
1.2 <i>BRCA1</i> : що відомо . . . . .	8
1.2.1 Людські варіанти гену <i>BRCA1</i> та їхнє медичне значення	11
1.3 Механізми впливу вірусів на геном людини . . . . .	13
1.4 Віруси, асоційовані із онкологічними захворюваннями . . . . .	14
<b>Експериментальна частина</b>	<b>18</b>
2.1 Висунення гіпотези та підбір методів та об'єктів дослідження	18
2.2 Метод I: BLAST — basic local alignment search tool, та результати . . . . .	20
2.3 Метод II: Молекулярна динаміка, та результати . . . . .	24
<b>Обговорення та висновки</b>	<b>34</b>
2.4 Вплив вірусів на еволюцію гена <i>BRCA1</i> . . . . .	35
2.5 Молекулярна динаміка . . . . .	36
2.6 Медичні імплікації . . . . .	37
2.7 Популяційні дослідження . . . . .	38

## Вступ

Відкриття та розуміння генетичних захворювань значно розвинулися за останні десятиліття. Нещодавні досягнення в галузі геномних технологій докорінно змінили наше розуміння генетичних захворювань та їхніх еволюційних траєкторій. Ґрунтуючись на цьому фундаменті, з'являється все більше можливостей дослідити складну взаємодію між факторами навколишнього середовища та генною еволюцією, яка може впливати на сприйнятливість до хвороб і їх прогресування в людських популяціях. Це дослідження присвячене мутації у гені *BRCA1*, який є важливим елементом у підтримці стабільності геному і причетний до розвитку раку молочної залози та яєчників. [1]

Основною **метою** цього дослідження є вивчення впливу факторів середовища на еволюцію гену *BRCA1*. Воно спрямоване на виявлення специфічних вірусних взаємодій, які можуть бути відповідальними за мутації або еволюційний тиск на ці гени, які є ключовими в клітинних механізмах, пов'язаних з репарацією пошкоджень ДНК. Також було доведено позитивну селекцію на деяких сайтах даного гену, і один із сайтів добору вирівнюється зі шкідливою мутацією в позиції білка 1443 [2].

**Предметом** дослідження є пошук факторів впливу на швидку еволюцію гену *BRCA1*. **Об'єктами** дослідження є геномні послідовності *BRCA1*, геноми онкогенних вірусів, які взаємодіють із геномом людини, а також білкові системи із диким та мутованими білками. Особливу увагу приділено вірусу Епштейна-Барр (EBV), папіломавірусу людини (HPV) та вірусу гепатиту В (HBV) та їх зв'язку з онкогенними процесами на рівні геному людини.

Рак молочної залози залишається найпоширенішим онкологічним захворюванням серед жінок у всьому світі, і його вплив відчувається в усіх демографічних групах і регіонах. У 2022 році рак молочної залози став

причиною 670 000 смертей у всьому світі, що підкреслює його значний тягар для здоров'я. Примітно, що минулого року це був найпоширеніший вид раку у жінок у 157 з 185 країн, що підкреслює його широку поширеність. [3]. **Актуальність** роботи полягає у дослідженні механізмів впливу мутацій, що стоять за сприянням виникненню раку молочної залози у ранньому віці.

У роботі представлено нове розуміння позитивного селективного тиску на ген *BRCA1*, зокрема у діпазоні, відповідному амінокислотній позиції 1443. Завдяки використанню сучасних інструментів молекулярної динаміки та класичної біоінформатики це дослідження робить унікальний внесок у розуміння того, як віруси можуть впливати на еволюцію генів через механізми, що виходять за рамки простої інфекції.

Для досягнення поставленої мети необхідно виконати ряд **завдань**:

1. Провести аналіз поточного стану досліджень за темою: генна еволюція, вплив вірусів на неї, еволюція гену *BRCA1*.
2. Проаналізувати потенціальну взаємодію вірусних геномів із людським геномом у локусі *BRCA1* та навколо нього.
3. Методами молекулярної динаміки оцінити зміну стабільності продукту гену *BRCA1* із появою мутацій.

У результаті проведення цієї роботи очікується отримати принаймні загальне розуміння можливих важелів впливу факторів навколишнього середовища на швидкість еволюції гену *BRCA1*.

## Розділ I. Літературний огляд

Згідно з Antonarakis *et al.* [4], процес виявлення генів, що спричиняють менделівські відхилення, можна розділити на три фази: біохімічна фаза, фаза генетичного зв'язку і позиційного клонування та фаза ідентифікації послідовності. Ця еволюція підкреслює прогрес геномних досліджень і технологій.

Крім того, Venton *et al.* [5] підкреслюють, що генетичні захворювання мають специфічне для людини походження, на яке вплинули еволюційні події, що передували виникненню людини. Різноманітні екологічні, культурні, демографічні та генетичні історії сучасних людських популяцій зумовили відмінності в поширеності генетичних захворювань.

Давні генетичні захворювання були частиною історії людства протягом століть. Було виявлено, що стародавні карибські спільноти мали вищу частоту патогенних варіантів, пов'язаних з певними моногенними захворюваннями, що свідчить про більшу поширеність цих захворювань у минулому порівняно з сучасними популяціями. [6]

Крім того, роль генетики у спадкових захворюваннях і вроджених дефектах була предметом історичного огляду [7]. Історія таких генетичних захворювань, таких як анкілозуючий спондилоартрит, сягає корінням древніх цивілізацій [8], демонструючи давню присутність генетичних захворювань в історії людства.

### 1.1 Теорії генної еволюції

Еволюція генів — це фундаментальна концепція в галузі еволюційної біології, що охоплює процеси та механізми, за допомогою яких гени змінюються з часом у популяціях та між видами. Зосереджуючись лише на теоріях еволюції генів, ми можемо розглядати безліч процесів, які сприяють

генетичному розмаїттю та адаптації, що спостерігаються у різних видів. До них належать мутації, природний добір, генетичний дрейф, потік генів і рекомбінація.

Мутації, або зміни в послідовності ДНК, слугують основним джерелом генетичної мінливості, сировиною для еволюції. Сучасні дослідження підкреслюють важливість розуміння мутацій для розуміння генетичної мінливості людини та еволюційної динаміки, стверджуючи, що передові технології секвенування надають все більше даних про генетичну мінливість людини та вірусну еволюцію [9]. У даній роботі ми присвятимо велику увагу однонуклеотидним поліморфізмам (тут і далі — SNP, single-nucleotide polymorphism).

Природний добір впливає на генетичну мінливість, найчастіше надаючи перевагу алелям, які сприяють виживанню та розмноженню. Навколишнє середовище відіграє вирішальну роль у визначенні того, які ознаки є виграними, що призводить до адаптивної еволюції популяцій. Деінде екологічні та еволюційні процеси зближуються, особливо у мікроорганізмів, через їхню коротку тривалість життя та великі розміри популяцій, що дозволяє швидкі еволюційні зміни [10].

Дрейф генів — це випадкові зміни частот алелів, які можуть мати значні наслідки у невеликих популяціях. Такі явища, як ефект пляшкового горла (*bottleneck effect*) можуть різко зменшити генетичне різноманіття і змінити хід еволюції. Деякі теоретичні моделі [11] дають уявлення про те, як екосистеми можуть впливати на генетичне різноманіття і впливати на результати еволюції як через генетичні процеси, так і через процеси на рівні екосистем.

Потік генів (також відомий як міграція і потік алелів) — це передача генетичного матеріалу від однієї популяції до іншої. Якщо швидкість потоку генів досить висока, то дві популяції матимуть еквівалентні частоти алелів і тому можуть вважатися однією ефективною популяцією. Було по-

казано, що достатньо лише «одного мігранта на покоління», щоб запобігти розбіжності популяцій через дрейф [12, ст. 325]. Переміщення особин і, як наслідок, потік генів може протидіяти ефекту генетичного дрейфу шляхом гомогенізації частот алелів у популяціях.

Рекомбінація «перетасовує» генетичний матеріал під час мейозу, створюючи нові комбінації алелів, які можуть виробляти нові ознаки, на які може діяти природний добір. Цей процес має вирішальне значення для генетичного різноманіття та еволюції складних ознак.

Розширений еволюційний синтез розширює сучасний синтез, включаючи нові уявлення про генетичні та епігенетичні механізми, процеси розвитку та екологічні взаємодії. Він підкреслює роль процесів розвитку в еволюції і те, як вони можуть генерувати нові фенотипічні варіації, на які може впливати відбір.

Отже, еволюція генів — це багатогранний процес, на який впливає складна взаємодія механізмів. Теоретичні та емпіричні дослідження продовжують розширювати наше розуміння цих процесів, забезпечуючи глибше розуміння динамічної природи еволюції.

## 1.2 *BRCA1*: що відомо

Гени *BRCA1* і *BRCA2* відіграють ключову роль у підтримці стабільності геному, насамперед завдяки їхній ролі у відновленні дволанцюгових розривів шляхом гомологічної рекомбінації. Мутації в цих генах тісно пов'язані з підвищеним ризиком розвитку раку молочної залози та яєчників, що підкреслює їхню важливу роль у супресії раку.

Їх відкриття глибоко вплинуло на наше розуміння спадкового раку молочної залози та яєчників. Ген *BRCA1* був відкритий у 1994 році [13]. Ген розташований на довгому (q) плечі хромосоми 17 в регіоні 2 смуги 1, від пари основ 41 196 312 до пари основ 41 277 500 [14]. Він кодує ядерний фосфопротейн, який відіграє важливу роль у репарації ДНК, контролі ключових

чових точок клітинного циклу та підтримці стабільності геному. Він взаємодіє з різними білками, в тому числі з RAD51, опосередковуючи гомологічну рекомбінаційну репарацію дволанцюгових розривів ДНК [15]. Крім того, продукт *BRCA1* бере участь у ремоделюванні хроматину, регуляції транскрипції через коактивацію та корепресію специфічних генів, а також у процесах убіквітації [16, 17]

Продукт *BRCA1*, однойменний білок BRCA1 (E3 убіквітин-протеїн лігаза) є важливим білком, який функціонує переважно шляхом сприяння утворенню поліубіквітинових ланцюгів, зв'язаних з Lys-6, які є необхідними для різних клітинних процесів, включаючи репарацію ДНК. Цей білок відіграє важливу роль у клітинних реакціях на пошкодження ДНК, опосередковуючи ці специфічні поліубіквітинові ланцюги, хоча невідомо, чи опосередковує він також інші типи поліубіквітинових ланцюгів [18].

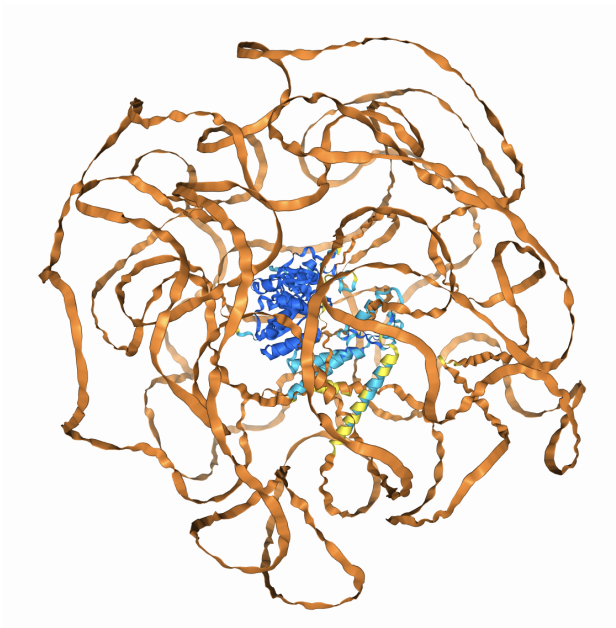


Рис. 1.1 — Структура білку BRCA1. Передбачення AlphaFold [19]

Крім того, гетеродимер BRCA1-BARD1, асоційований з цією лігазою, координує широкий спектр клітинних механізмів, включаючи репарацію ДНК, убіквітування та регуляцію транскрипції, що є критично важливими для підтримки геномної стабільності [20]. Ця складна взаємодія робить

значний внесок у нуклеацію центросомальних мікротрубочок і є життєво важливою для правильної зупинки клітинного циклу після іонізуючого опромінення в різних фазах клітинного циклу [21].

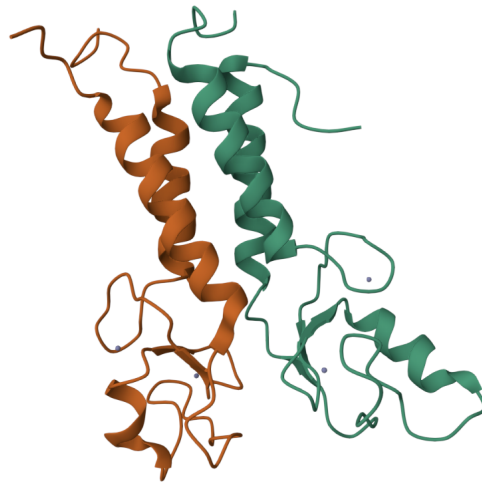


Рис. 1.2 — Димер BRCA1-BARD [22]

Крім того, цей білок необхідний для таргетингу FANCD2 на ділянки пошкодження ДНК і відіграє роль в інгібуванні синтезу ліпідів, взаємодіючи з неактивним фосфорильованим АСАСА для запобігання його дефосфорильованню [23]. Він також посилює гомологічну рекомбінаційну репарацію (HRR) через взаємодію з PALB2, тим самим модулюючи процес рекомбінаційної репарації шляхом впливу на завантаження репараційного механізму BRCA2-RAD51 на розривах ДНК [24].

У складі комплексу BRCA1-RBBP8 він регулює активацію CHEK1 і контролює контрольні точки G2/М клітинного циклу після пошкодження ДНК через BRCA1-опосередковане убіквітинування RBBP8 [25]. Цей білок також діє як активатор транскрипції, що ще раз підкреслює його універсальну роль у клітинних механізмах [26].

Сучасні дослідження продовжують вивчати роль білків BRCA у репарації ДНК, а нові дані сприяють розробці цільових методів лікування

та стратегій профілактики. Поточні дослідження зосереджені на розумінні взаємодії між мутаціями *BRCA* та іншими ключовими білками, що беруть участь у канцерогенезі, а також на більш широких наслідках цих взаємодій для лікування раку [27].

Відкриття цих генів не тільки удосконалило генетичне тестування та оцінку ризику спадкових синдромів раку молочної залози та яєчників, але й призвело до розробки цільової терапії. Наприклад, інгібітори PARP застосовують принцип синтетичної летальності, будучи особливо ефективними у лікуванні раку з мутаціями *BRCA1* або *BRCA2* [28, 29].

### 1.2.1 Людські варіанти гену *BRCA1* та їхнє медичне значення

За даними публічної бази даних BRCA Exchange існує більше 37 тисяч варіант гену *BRCA1*, 2304 з яких — патогенні[30]. Інші джерела свідчать про навіть більшу кількість шкідливих мутацій: 3867 для *BRCA1*[31].

Мутації в *BRCA1* успадковуються за аутосомно-домінантним типом, що призводить до значно підвищеного ризику розвитку раку молочної залози та яєчників. Жінки, які є носіями шкідливих мутацій в обох генах, мають ризик розвитку раку молочної залози у 45-65% і раку яєчників у 11-39% у віці до 70 років [1, 32]. Згідно з іншими джерелами, шкідливі мутації у генах *BRCA1* та *BRCA2* призводять до підвищення ризику розвитку раку молочної залози в юному віці до 80% та 60% відповідно [33, 34]. Крім того, носії мутацій в *BRCA1* мають 30-40% ймовірність розвитку раку яєчників[33].

Рак виникає, коли у гетерозиготних осіб відбувається соматична втрата гетерозиготності (loss of heterozygosity — LoH) в локусі *BRCA1*, в результаті чого тільки аномальний алель залишається інтактним. Оскільки продукт гену відіграє важливу роль у ключових клітинних процесах, таких як репарація ДНК, контроль клітинного циклу та регуляція транскрипції, зрозуміло, чому інактивуючі мутації є настільки згубними. Про важливість цих білків свідчить і той факт, що миші з нульовим *BRCA1* і *BRCA2* є

ембріонально летальними [35].

Втрата гетерозиготності (LoH) — це генетичне явище, яке виникає, коли в диплоїдному організмі втрачається один алель гена, в результаті чого клітина або організм стає гомозиготною за цією ділянкою геному. Ця втрата може відбуватися через різні механізми, включаючи делеції, мітотичну рекомбінацію або нерозщеплення, які призводять до втрати алеля дикого типу в гетерозиготному генотипі. LoH має особливе значення в контексті генетики раку, оскільки може призводити до інактивації генів-супресорів пухлин і сприяти онкогенезу.

У генів-супресорів пухлин для розвитку раку, як правило, повинні бути інактивовані обидва алелі. У носіїв мутантного алеля BRCA1 або BRCA2, алель дикого типу, що залишився, має вирішальне значення для запобігання розвитку пухлини. LoH у цих випадках може призвести до інактивації функціонального алеля, що залишився, тим самим сприяючи прогресуванню раку [36]. Ця концепція є центральною для гіпотези двох влучень Кнудсона, яка стверджує, що обидва алелі гена-супресора пухлини повинні бути змінені для того, щоб клітина переросла в рак [37].

Крім того, LoH має наслідки, що виходять за рамки спадкового раку. Вона спостерігається в різних спорадичних пухлинах, що вказує на її роль у соматичному прогресуванні раку. Наприклад, дослідження показали, що LoH в локусі гена-супресора пухлин TP53 асоціюється з кількома формами спорадичного раку, що свідчить про його широку роль у канцерогенезі [38]. Більше того, це явище не обмежується лише раком. Дослідження продемонстрували, що LoH може сприяти патогенезу деяких інших захворювань, включаючи нейродегенеративні розлади та вроджені аномалії, порушуючи ключові генетичні шляхи [39].

### 1.3 Механізми впливу вірусів на геном людини

Вплив вірусів на гени людини — це багатогранна галузь дослідження, яка охоплює різні механізми, за допомогою яких віруси можуть впливати на генетичну структуру та функції. Ця складна взаємодія не лише проливає світло на еволюцію вірусів та їхніх хазяїв, але й має важливе значення для розуміння механізмів захворювань, імунних реакцій та терапевтичних втручань.

Віруси, особливо ретровіруси, мають здатність інтегрувати свій генетичний матеріал у геном хазяїна. Цей процес може призвести до стабільного вбудовування вірусних генів у геном людини, які з часом можуть бути кооптовані для виконання нових функцій хазяїном. Було виявлено зв'язок між герпесвірусом людини 6 та аутоімунними захворюваннями, припускаючи, що вірусні інфекції можуть впливати на генетичні та екологічні фактори, потенційно запускаючи аутоімунні реакції [40].

Іншим механізмом впливу є регуляція експресії. Віруси можуть суттєво впливати на експресію генів людини. Через такі механізми, як епігенетичні модифікації та втручання в регуляцію транскрипції, віруси можуть підвищувати або знижувати експресію певних генів, впливаючи на клітинні функції та сприйнятливість до хвороб. Наприклад, було досліджено, як домен злиття вірусу Ласса взаємодіє з аніонними ліпідами, що може ініціювати злиття і впливати на доставку генетичного матеріалу вірусу в клітини хазяїна [41]. Така взаємодія може глибоко впливати на експресію та функцію генів хазяїна.

Довготривала коеволюція вірусів та їхніх хазяїв призвела до еволюційних адаптацій у геномі людини. Вірусний тиск сформував генетичну різноманітність людини, впливаючи на гени, пов'язані з імунітетом, проникненням вірусів у клітини та іншими важливими функціями. Наприклад, досліджується вплив модифікацій гістонів та послідовності ДНК на ста-

більність вірусних нуклеосомоподібних частинок, надаючи уявлення про те, як віруси можуть впливати на динаміку хроматину та регуляцію генів [42].

Розуміння вірусного впливу на гени людини відкриває потенційні терапевтичні шляхи, такі як використання бактеріофагових векторів для доставки генів, як це досліджували Venkataraman *et al.* [43]. І навпаки, маніпуляції вірусів з експресією генів хазяїна можуть призвести до патогенних наслідків, що вимагає глибшого розуміння цих взаємодій для розробки протівірусних стратегій.

Генетичні відмінності між людьми можуть впливати на сприйнятливість до вірусних інфекцій та тяжкість перебігу захворювань. Chiti [44] обговорює вплив генетики на вірусний патогенез та імунну відповідь, підкреслюючи роль генетичних варіацій у формуванні індивідуальних реакцій на вірусні інфекції.

Таким чином, вірусний вплив на гени людини охоплює широкий спектр взаємодій — від прямої генетичної інтеграції та впливу на експресію генів до еволюційних наслідків і наслідків для хвороби та терапії. Ці дослідження підкреслюють важливість вивчення взаємодії вірус-хазяїн для розкриття складної динаміки генної регуляції, еволюційної біології та патогенезу хвороб.

#### **1.4 Віруси, асоційовані із онкологічними захворюваннями**

Рак молочної залози, як найпоширеніший тип раку серед жінок у всьому світі [45], зумовлений складною взаємодією між зовнішніми та внутрішніми факторами. Віруси вважаються потенційними чинниками розвитку раку молочної залози, серед яких основними підозрюваними є віруси папіломи людини (ВПЛ), вірус Епштейна-Барр (ВЕБ) та вірус пухлини молочної залози миші (ВПМЗ). Залучення вірусів до канцерогенезу може відбуватися як безпосередньо в якості трансформуючих агентів, так і в якості

кофакторів, що запускають рак у присутності інших канцерогенних факторів.

Вірус Епштейна-Барр (EBV), або герпесвірус 4, був виявлений у безпосередній близькості до хромосомних розривів у клітинах раку молочної залози, що містять мутації *BRCA1*. EBV, класифікований як канцероген класу I, демонструє різні дані щодо його ролі в розвитку раку молочної залози, причому деякі дослідження вказують на його присутність у більш агресивних формах захворювання [46]. Присутність EBV пов'язана з більш високими темпами прогресування захворювання у хворих на рак молочної залози, що свідчить про його потенційну роль у виникненні та прогресуванні раку молочної залози через такі механізми, як пригнічення контролю клітинного циклу, імуносупресивні властивості та взаємодія з вірусним інтерлейкіном-10 [47]. Це свідчить про те, що вірус може сприяти геномній нестабільності, характерній для цих видів раку. Наявність мутацій у гені *BRCA1* може призвести до неконтрольованої втрати фрагментів хромосом або їхнього невірною повторного приєднання, що характерно для спадкового раку грудей. Було встановлено, що помилки в повторному приєднанні хромосомних фрагментів часто відбуваються поряд з вірусними послідовностями на людських хромосомах, зокрема з варіантами пухлинного вірусу Епштейна-Барр, НКНД40 і НКНРС60. Майже всі точки перелому були поруч з місцями прикріплення EBV, гомологіями пухлинних варіантів EBV та регуляторними маркерами, асоційованими з EBV. Основна частина точок перелому на хромосомі 2 була зосереджена біля місць зв'язування EBV. Ці результати вказують на те, що активне лікування та профілактика інфекцій, викликаних герпесвірусами, можуть бути корисними для носіїв мутацій BRCA. [48]

Дослідження єгипетських пацієток з раком молочної залози показали значну присутність вірус пухлини молочної залози миші (MMTV)-подібних env-послідовностей як у сімейних, так і в поодиноких випадках. Наявність

цих послідовностей корелювала з мутаціями в гені *BRCA1*, що дозволяє припустити, що ММТВ може відігравати певну роль в ініціюванні або розвитку раку у генетично схильних осіб. Механізм може включати вірусні послідовності, що впливають на транскрипційну регуляцію генів *BRCA* або впливають на мікрооточення пухлини, сприяючи прогресуванню раку [49].

Дослідження виявили ДНК цитомегаловірусу людини (HCMV) у значному відсотку тканин раку молочної залози, особливо тих, що експресують мутовані білки *BRCA1*. Це свідчить про потенційну роль вірусу в прогресуванні і, можливо, ініціації раку молочної залози. ЦМВ може впливати на розвиток раку, порушуючи нормальні клітинні процеси за допомогою своїх білків, що може перешкоджати пригніченню пухлинних функцій білка *BRCA1*. Вірус також може викликати запалення, що сприяє прогресуванню раку, або взаємодіяти з *BRCA1* для посилення росту та виживання пухлини [50]. Одним із механізмів є вплив HCMV на індукцію інфламасоми та інтерферонових відповідей через взаємодію з білком IFI16, який є ядерним датчиком чужорідної ДНК вірусів, таких як HCMV. IFI16 утворює комплекс з *BRCA1*, що допомагає у виявленні вірусних геномів у ядрі та ініціює подальше формування інфламасоми IFI16-ASC-procaspase-1, що транспортується до цитоплазми, викликаючи виробництво каспази-1 та інтерлейкіну- $1\beta$  (IL- $1\beta$ ). Важливо, що *BRCA1* також активує сигнальний комплекс IFI16-STING у цитоплазмі під час інфекції HCMV, що призводить до індукції IFN- $\beta$ . Ці процеси підкреслюють неідентифіковану раніше імуномодулюючу роль *BRCA1*, яка сприяє антивірусній функції IFI16. Такі взаємодії можуть мати наслідки для розвитку раку, особливо у осіб з мутаціями *BRCA1*, які вже мають знижену здатність до репарації ДНК [51].

Поліомавіруси, такі як вірус Джона Каннінгема (JCV) та сіміанський вірус 40 (SV40), пов'язані з онкологічними захворюваннями. Незважаючи на суперечливі результати щодо їх присутності в раку молочної залози, вва-

жається, що вони сприяють канцерогенезу, викликаючи безсмертя клітин і генетичні зміни. SV40, зокрема, був виявлений у карциномах молочної залози, а його великий Т-антиген відіграє певну роль у трансформації епітелію молочної залози [52, 53].

Широко відомий зв'язок ВПЛ з раком шийки матки, але його роль у виникненні раку молочної залози залишається суперечливою [54]. У зразках раку молочної залози були виявлені ВПЛ високого ризику типів 16 і 18, що свідчить про потенційний зв'язок. ВПЛ може сприяти канцерогенезу молочної залози через експресію ранніх білків, які деактивують білки-супресори пухлин і впливають на сигнальні шляхи клітин, що потенційно призводить до інвазії клітин і метастазування. ВПЛ вводить свої гени Е6 та Е7 до клітин господаря, що призводить до деградації пухлинного супресора p53 та ретинобластомного білка (pRB), відповідно. Ці вірусні білки порушують клітинний цикл і запобігають апоптозу клітин, сприяючи їх необмеженому поділу та онкогенезу. З цього приводу, вакцинація проти ВПЛ визнана ефективним засобом профілактики розвитку раку шийки матки та інших форм раку, пов'язаних з цим вірусом. [55, 56].

## Експериментальна частина

### 2.1 Висунення гіпотези та підбір методів та об'єктів дослідження

Функції продукту гену *BRCA1* є важливими у багатьох етапах клітинної машинерії, так що можна було б очікувати його високу консервативність у еволюційному плані. Швидкість еволюції у подібних генах передбачається вкрай низька. Така повільна еволюція значною мірою пояснюється суворим селективним тиском, з яким вони стикаються для збереження структурної цілісності та функціональності, незважаючи на виникнення помилок трансляції під час синтезу білка. Незважаючи на те, що рівень експресії *BRCA1* залежить від контексту, навіть низький рівень неправильного фолдингу може призвести до утворення значної кількості дисфункціональних або потенційно токсичних білків. Це створює сильний селективний тиск на ці білки, який змушує їх підтримувати послідовності, що складаються правильно, навіть коли виникають помилки трансляції. Вважається, що гени, які кодують високо експресовані або базові білки, еволюційно «застрягли» в послідовностях, стійких до помилок трансляції — це послідовності, які можуть правильно згортатися, навіть якщо внаслідок неправильної транскрипції в них включені злегка неправильні амінокислоти. Така стійкість до трансляції сприяє їх повільній еволюції, оскільки вони менш толерантні до варіацій послідовностей, які можуть збільшити ризик неправильного згортання. Таким чином, повільна еволюція високо експресивних білків є захисною стратегією, розробленою для того, щоб ці важливі білки зберігали свою функціональність, незважаючи на неминучі помилки, які трапляються під час синтезу білка.

Всупереч цьому, численні дослідження показали, що ці гени знаходяться під позитивною селекцією [57, 2]. Тип селекції, під якою знаходиться

ген, може бути визначений зі швидкості накопичення несинонімічних (тих, що змінюють послідовність закодованих амінокислот, dN) та синонімічних (сайлент, dS) мутацій з плином часу. dN розраховується як кількість несинонімічних замін, ділена на загальну кількість несинонімічних сайтів у гені. dS — відповідно кількість синонімічних замін. Відношення dN/dS розраховується шляхом ділення dN на dS. Мутації, що змінюють первинну структуру білка мають набагато нижчі шанси утримання у геномі ніж синонімічні, тому пропорція  $dN/dS \ll 1$  для подавляючої більшості геному ссавців [58]. Деякі гени — псевдогени, приміром — йдуть шляхом нейтральної еволюції з  $dN/dS \sim 1$ , оскільки немає сильних селективних сил за чи проти нових мутацій у таких генах. Нарешті, селекція на користь несинонімічних мутацій призводить до зміщення пропорції dN/dS в бік  $> 1$ . Ці гени у такому випадку класифікують як ті, що перебувають під позитивною селекцією та постійно зазнають добору на користь «інновацій» у амінокислотній послідовності. Так як раніше вже було доведено, що ген *BRCA1* перебувають під позитивною селекцією із  $dN/dS = 2.79$ , а також проведений аналіз конкретних амінокислотних позицій [2], була висунута гіпотеза про можливий вплив вірусів на еволюцію цих генів. Lou *et al.* дослідили, що незважаючи на очікувану консервативність, *BRCA1* підлягає сильному селективному тиску. Особливу увагу привернула селекція у позиції 1443 білку *BRCA1*, оскільки мутація саме на цій позиції занесена у список смертельних.

SNP R1443X (Arg -> Ter) призводить до вкороченої форми білку та зниженні його функціональності. Заміна відбувається у домені SCD — Serine Cluster Domain, який є важливою функціональною ділянкою білка *BRCA1*, особливо щодо його ролі у відповіді на пошкодження ДНК. Цей домен збагачений залишками серину, багато з яких стають сайтами фосфорилування після пошкодження ДНК. Події фосфорилування в цьому домені переважно опосередковані кіназами ATM (Ataxia-Telangiectasia Mutated) та

ATR (ATM and Rad3-Related), які активуються у відповідь на дволанцюго-ві розриви ДНК та реплікаційний стрес, відповідно. Він діє як підсилювач сигналу у відповідь на пошкодження ДНК, забезпечуючи надійний процес репарації. Фосфорильований SCD впливає на вибір шляху репарації ДНК. Він сприяє гомологічній рекомбінації (HR), точному механізму репарації, критично важливого для усунення розривів під час S і G2 фаз клітинного циклу, полегшуючи рекрутування репараційних білків до місць пошкодження. Власне вкорочений BRCA1 також втрачає два домени BRCT (BRCA1-C Terminal), які і будуть опосередковувати рекрутування репараційних білків до місць пошкоджень.

Існує також й інший варіант SNP на цій позиції — R1443G (Arg -> Gly), який запроTOCOLьований як безпечний (benign) [59], на який можливо діє позитивна селекція. Аби перевірити можливі гіпотези, у цьому розділі ми проведемо молекулярну динаміку білкових систем.

У Розділі I ми розглянули можливі випадки впливу вірусів на людський геном та навпаки. За допомогою обраних методів та підходів ми зможемо перевірити, чи впливають віруси на еволюцію генів *BRCA1*.

## 2.2 Метод I: BLAST — basic local alignment search tool, та результати

Так як відомо, що вірусні геноми найчастіше вбудовуються у геном хазяїна за принципом мікрогомології, ми проведемо BLAST-аналіз для виявлення потенційних сайтів вбудовування у хромосомі 17. Для цього було взято геноми або геномні ділянки вірусів, асоційованих із розвитком раку молочної залози або яєчників. Серед них: Сіміанський вірус 40 (Simian virus 40 — SV40), вірус Епштейн-Барр (EBV), папіломавірус людини та вірус пухлини молочної залози миші (ВПМЗ).

BLAST геному вірусу Епштейн-Барр показав наступні результати: є одна потенційна мішень. У хромосомі 17 присутній доволі довгий фрагмент

вирівнювання, що розтягується на 925 пар нуклеотидних основ. Значення  $E = 1.44e-42$ , з hit score 186. Це свідчить про дуже високу статистичну важливість цього вирівнювання. Проте, позиція даного фрагмента відповідає іншому гену хромосоми 17 — *NXN*, що кодує нуклеоредоксин.

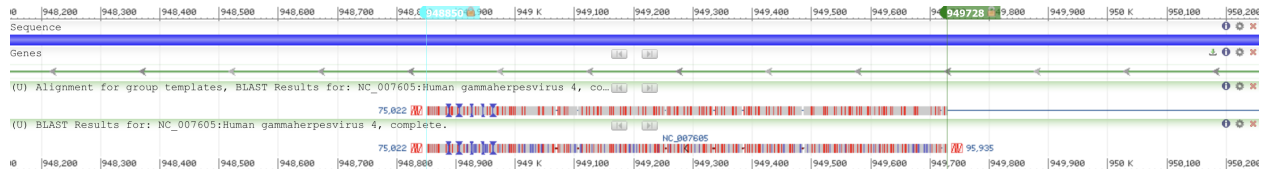


Рис. 2.1 — BLAST-аналіз геному герпесвірусу 4 (Вірус Епштейн-Барр) проти людського геному. Лініями визначені межі вирівнювання.

Нуклеоредоксин функціонує як редокс-залежний негативний регулятор сигнального шляху Wnt, можливо, запобігаючи убіквітуванню DVL3 комплексом VCR(KLHL12). Може також функціонувати як транскрипційний регулятор та діяти як регулятор протеїнфосфатази 2A (PP2A) [60]. Ген *NXN* не є регуляторним до *BRCA1*, а білок нуклеоредоксин не є партнером *BRCA1* (рис. 2.4).



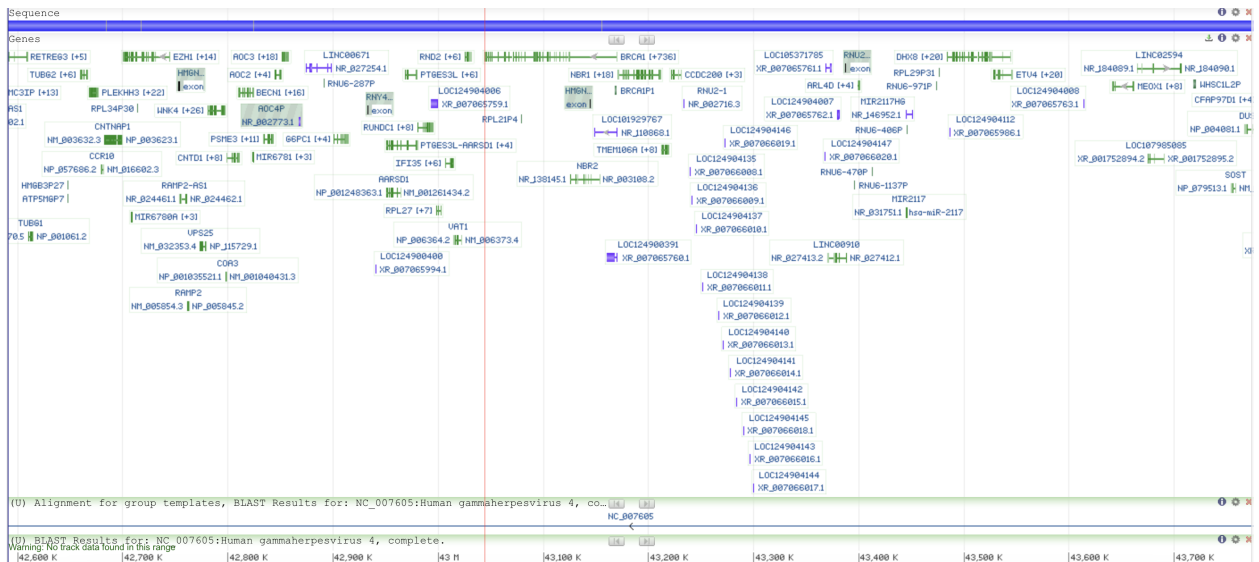


Рис. 2.3 — BLAST-аналіз геному вірусу Епштейн-Барр проти людського геному. Червона лінія — початок *BRCA1*.

Є дослідження, що показують вбудовування вірусу Епштейн-Барра безпосередньо у ділянку *BRCA1*. Інтеграція має інтронний характер, тож є ймовірність, що вірус таки вбудовується у регуляторні або промоторні ділянки гену [62]. Потенційні клінічні наслідки інтеграції ДНК EBV в організм пацієнтів, які страждають на різні типи раку з мутаціями в гені *BRCA1*, залишаються нез'ясованими. Гени-мішені інтеграції EBV (integration-targeted genes — ITGs) значно більше експресуються у пов'язаних з раком шляхах, таких як репарація ДНК, регуляція клітинного циклу та шляхи апоптозу. [63]. Це може слугувати чинником впливу на швидкість еволюції зокрема, хоча і повністю не пояснює її зміну.

Junction sequence	> TTGTTGTAGATAGAGACCAATAATGAGCAGGATGATGTCTAGGAAGAAGATGAGATGGAGTTTCATTCTTGTTCGCCAGGCTAGAGTGCAATGGCCGAT
Virus sequence	TTGTTGTAGATAGAGACCAATAATGAGCAGGATGATGTCTAGGAAGAAG
Human sequence	TGAGATGGAGTTTCATTCTTGTTCGCCAGGCTAGAGTGCAATGGCCGAT
MH patterns	TAGGAAGAAGaTGAGATGAG TAGGAAGAAGaTGAGATGAG

Рис. 2.4 — Сайт інтеграції геному вірусу Епштейн-Барра у *BRCA1* у ДНК з клітин гепатоклітинної карциноми

BLAST послідовності HBV проти людського геному не дав значущих

результатів, тобто є ймовірність, що вірусний геном не вбудовується у здорові клітини. SV40 також не має значної схожості із ДНК хромосоми 17, тож на даному етапі він не є релевантним для подальших досліджень.

### 2.3 Метод II: Молекулярна динаміка, та результати

Молекулярна динаміка (МД) — це метод комп'ютерного моделювання, що використовується в біоінформатиці та обчислювальній біології для вивчення фізичних рухів атомів і молекул у часі. Застосовуючи закони руху Ньютона, МД дозволяє досліджувати структурні, динамічні та термодинамічні властивості біологічних молекул у змодельованому віртуальному середовищі. Цей метод є особливо потужним для моделювання складних біологічних систем на атомному рівні, таких як білки, нуклеїнові кислоти та ліпідні бішари, які складно вивчати лише експериментальними методами.

У цій роботі застосування методів молекулярної динаміки дозволяє глибше зрозуміти структурні та функціональні зміни в білку BRCA1, що відбуваються в результаті мутації на позиції 1443. Ми оцінили стабільність білків із наступними мутаціями: R1443G та R1443A, та порівняли із регулярним білком із аргініном на позиції 1443. Аналіз динаміки атомів і молекул в часі може надати важливу інформацію про те, як змінюється конформація білка під впливом мутації, що може призводити до зміни його функціональних властивостей і сприяти розвитку патологічних станів, таких як рак.

Підготовка PDB файлів починається із первинного моделювання, для якого ми застосували інструмент на основі deep-learning алгоритму — AlphaFold 3 [19]. Амінокислотну послідовність BRCA1 було взято з бази даних UniProt [64], після чого вручну проводилась заміна амінокислоти на позиції 1443. Подальша підготовка файлів, що містять білок, для проведення молекулярної динаміки є критичною стадією у білковому моделюванні. Основними етапами підготовки файлів під час проведення роботи були:

1. Присвоєння зарядів і радіусів Ван-дер-Ваальса — було застосовано силове поле OPLS-AA/L для присвоєння зарядів атомам білку та налаштування параметрів взаємодії.
2. Сольвація білка — розміщення білка в середовищі водяних молекул (сольваційна коробка у нашому випадку підлаштовувалась під виміри білку).
3. Додавання іонів — балансування заряду системи шляхом додавання катіонів ( $\text{Na}^+$ ) і аніонів ( $\text{Cl}^-$ ) для нейтралізації загального заряду білка та імітації фізіологічних умов.
4. Мінімізація енергії — проведення енергетичної мінімізації для усунення можливих стеричних конфліктів у структурі білка та оптимізації його конформації перед початком динамічних розрахунків.
5. Налаштування параметрів динаміки — визначення фізичних умов експерименту, а також часу моделювання та інтервалу запису даних.
6. Перевірка системи перед запуском молекулярної динаміки — перегляд і перевірка всіх параметрів системи та готовності файлів конфігурації для запуску моделювання.

Через брак часу та безпекову ситуацію, динаміку не вдалось провести у повному обсязі. Частина файлів, що містила мутантні білки, була втрачена через перепади навантаження на кластер. Тому порівняльний аналіз усереднених структур провести не вдалось, натомість було проаналізовано мінімізовані системи.

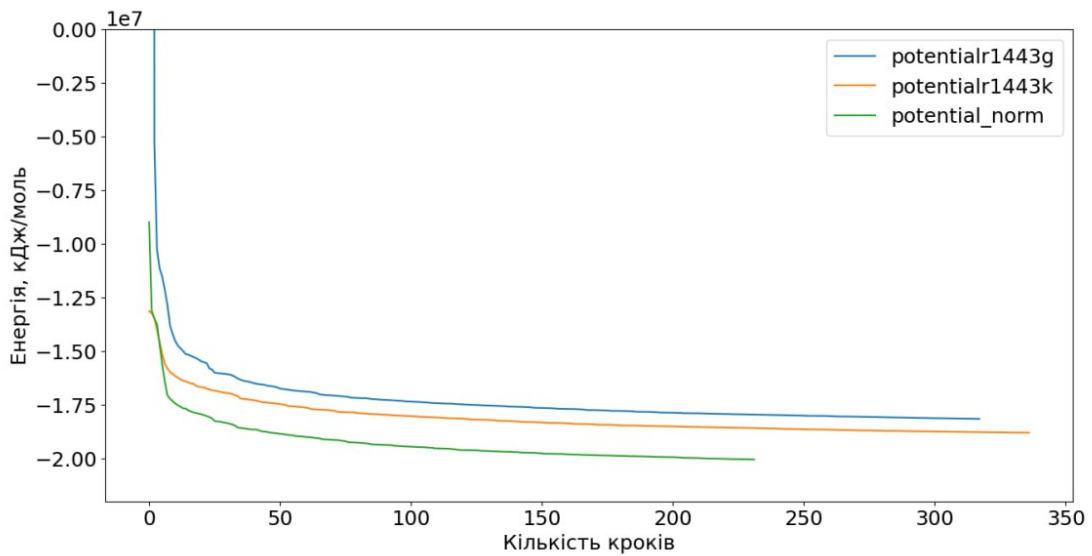


Рис. 2.5 — Графік мінімізації енергій систем BRCA1: регулярний білок та два мутованих

На рис. 2.5 зображено графік мінімізації енергій систем BRCA1: регулярний білок та два мутованих. Графік відображає зміни потенціальної енергії в залежності від кількості кроків мінімізації для трьох різних систем: нормального білка BRCA1 (`potential_norm`) та білка з мутаціями R1443G (`potentialr1443g`) та R1443K (`potentialr1443k`). 1. Аналіз кривих мінімізації:

Всі три криві показують зменшення потенціальної енергії з часом, що свідчить про процес мінімізації енергії в системах.

Крива для нормального білка (`potential_norm`) демонструє найнижчий рівень кінцевої енергії, що вказує на стабільність його конформації після мінімізації.

Криві для мутованих білків (`potentialr1443g` та `potentialr1443k`) також знижуються, але кінцеві значення потенціальної енергії є вищими порівняно з нормальним білком. Це свідчить про те, що мутації R1443G та R1443K впливають на стабільність білка, знижуючи його здатність досягати енергетично мінімального стану.

2. Кількість кроків мінімізації:

Графік показує, що більшість зниження енергії відбувається на перших 100 кроках мінімізації, після чого криві вирівнюються, що свідчить про досягнення локального мінімуму енергії.

Вплив мутацій:

Мутації R1443G та R1443K призводять до менш ефективного мінімізації потенціальної енергії білка, що може вказувати на конформаційні зміни та зниження стабільності білка.

Порівняння стабільності:

Регулярний білок BRCA1 демонструє кращу стабільність після мінімізації, що підкреслює важливість збереження його нативної структури для виконання біологічних функцій.

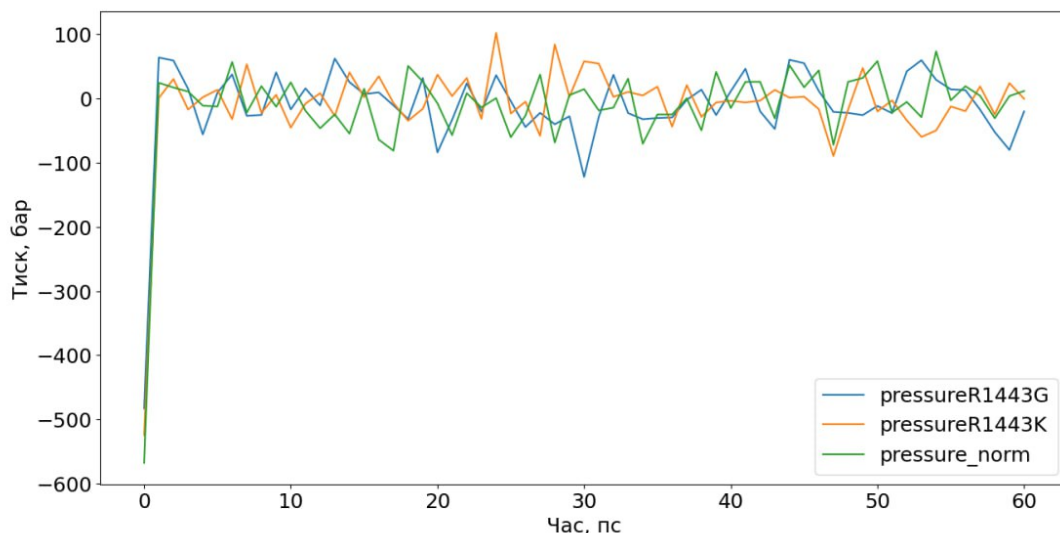


Рис. 2.6 — Графік тиску процесу урівноваження системи NPT

На рис. 2.6 зображено зміни тиску в системі після мінімізації для нормального білка BRCA1 та двох його мutowаних варіантів (R1443G та R1443K). Часові зміни тиску представлені на осі X, а тиск у барах — на осі Y. 1. Початковий стан:

На початку симуляції тиск для всіх трьох систем показує значний спад до приблизно -500 бар, що є характерним для стабілізації системи після початкового встановлення параметрів.

## 2. Стабілізація тиску:

Після приблизно 5 пс (пікосекунд) тиск у всіх системах починає стабілізуватися навколо значення 0 бар. Це свідчить про те, що системи досягли стану рівноваги після початкового спаду.

## 3. Коливання тиску:

Протягом решти симуляції тиск коливається навколо нульового значення для всіх трьох систем. Криві показують, що середні значення тиску для нормального білка (`pressure_norm`), R1443G (`pressureR1443G`) та R1443K (`pressureR1443K`) перебувають в межах від -100 до +100 бар з подібними амплітудами коливань.

### Інтерпретація результатів

#### 1. Порівняння стабільності:

Стабілізація тиску навколо нульового значення після мінімізації вказує на досягнення системами рівноважного стану. Всі три системи показують подібні коливання, що свідчить про те, що мінімізація енергії ефективно вирівняла внутрішній тиск.

#### 2. Вплив мутацій:

Незважаючи на відмінності у стабільності білків, відображені у потенціальній енергії, тиск у всіх трьох системах показує схожу поведінку після мінімізації. Це може означати, що хоча мутації впливають на стабільність білка, вони не мають значного впливу на загальний тиск у системі після досягнення рівноважного стану.

#### 3. Значущість коливань:

Коливання тиску в межах від -100 до +100 бар є нормальними для молекулярної динаміки і не вказують на аномалії у поведінці систем. Це підтверджує, що мінімізація та подальша стабілізація систем були успішними.

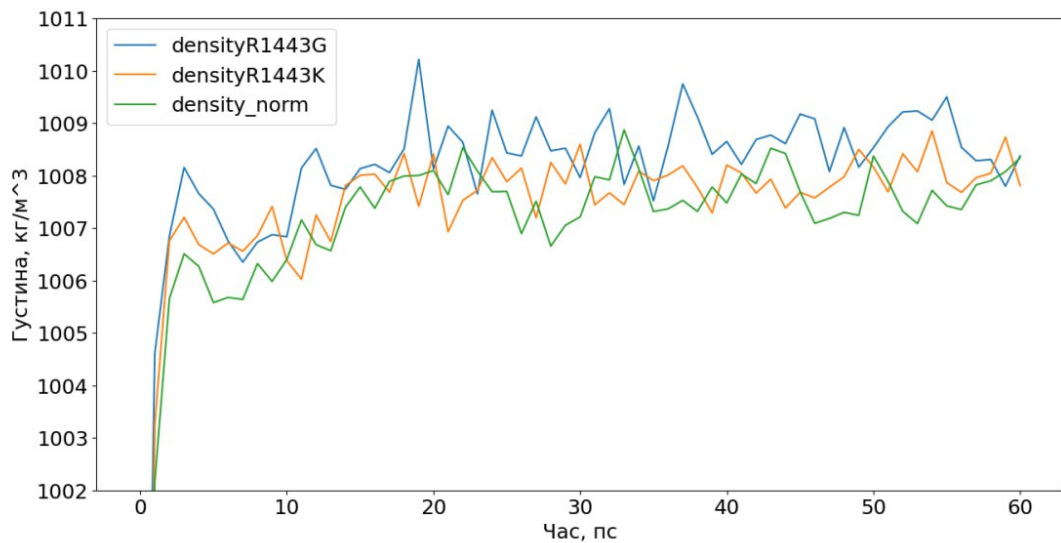


Рис. 2.7 — Графік густин процесу урівноваження системи NPT

На рис. 2.7 зображено зміни густини системи в залежності від часу для нормального білка BRCA1 та двох його мутованих варіантів (R1443G та R1443K). Часові зміни густини представлені на осі X, а густина в  $\text{кг}/\text{м}^3$  — на осі Y.

### 1. Початковий стан:

На початку симуляції густина для всіх трьох систем різко зростає, що свідчить про стабілізацію системи після початкового встановлення параметрів.

### 2. Стабілізація густини:

Після приблизно 5 пс (пікосекунд) густина у всіх системах починає стабілізуватися навколо значення  $1008 \text{ кг}/\text{м}^3$ . Це свідчить про досягнення рівноважного стану після початкового періоду стабілізації.

### 3. Коливання густини:

Протягом решти симуляції густина коливається навколо середнього значення для всіх трьох систем. Криві показують, що середні значення густини для нормального білка (`density_norm`), R1443G (`densityR1443G`) та R1443K (`densityR1443K`) перебувають в межах від  $1006$  до  $1010 \text{ кг}/\text{м}^3$  з подібними амплітудами коливань.

### Інтерпретація результатів

### 1. Порівняння стабільності:

Стабілізація густини навколо значення  $1008 \text{ кг/м}^3$  після мінімізації вказує на досягнення системами рівноважного стану. Всі три системи показують подібні коливання, що свідчить про те, що мінімізація енергії ефективно вирівняла внутрішню густину систем.

### 2. Вплив мутацій:

Незважаючи на відмінності у стабільності білків, відображені у потенціальній енергії, густина у всіх трьох системах показує схожу поведінку після мінімізації. Це може означати, що хоча мутації впливають на стабільність білка, вони не мають значного впливу на загальну густину системи після досягнення рівноважного стану.

### 3. Значущість коливань:

Колівання густини в межах від  $1006$  до  $1010 \text{ кг/м}^3$  є нормальними для молекулярної динаміки і не вказують на аномалії у поведінці систем. Це підтверджує, що мінімізація та подальша стабілізація систем були успішними.

Графік густини після мінімізації показує, що всі три системи — нормальний білок BRCA1 та його мutowані варіанти R1443G і R1443K — досягають стабільного рівноважного стану з густиною, що коливається навколо значення  $1008 \text{ кг/м}^3$ . Незважаючи на структурні відмінності, викликані мутаціями, поведінка густини вказує на успішну стабілізацію систем після початкового періоду мінімізації. Це важливо для підтвердження коректності умов симуляції та дозволяє продовжувати дослідження динамічних властивостей білків з упевненістю в точності отриманих результатів.

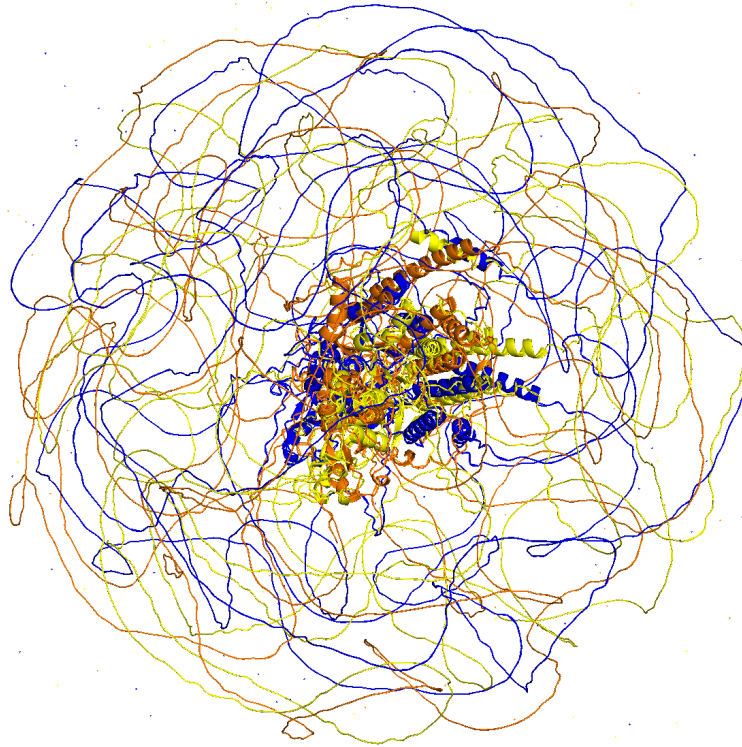


Рис. 2.8 Структури BRCA1 після мінімізації енергії. Регулярний білок — синій, R1443G — жовтий, R1443K — помаранчевий

На рис. 2.8 видно, що структури не вирівнюються одна відносно одної, що може бути пояснене неоформленими ділянками білка, які випадковим чином розташовуються у просторі та не можуть бути чітко змодельовані. Виходячи із цього, ми не можемо порівняти структурні зміни візуально.

Після завершення молекулярної динаміки для регулярного білка, аналіз динамічних властивостей білка вимагає усереднення по траєкторії. Цей процес включає використання згенерованих даних про положення атомів у кожному кроці симуляції для визначення статистично значущих показників. Програмний пакет GROMACS надає утиліту `gmx covar`, яка дозволяє обчислювати та діагоналізувати масо-ваговану коваріаційну матрицю, засновану на всіх структурах, що були підігнані до структури у файлі структури. В ході коваріаційного аналізу, структура для усереднення визначається як початкова з  $t=-1$ . Це дозволяє отримати усереднену референтну структуру, до якої приводять всі інші структури, та яка використовується для дальшої візуалізації та аналізу рухів білка. Власні вектори, отримані

після діагоналізації, можуть бути аналізовані за допомогою `gmx anaeig`, що дозволяє виявити домінуючі моди руху атомів у білку. Застосування таких методів дає можливість детально розглянути коливання та динаміку білка, що є критично важливим для розуміння його функціональної активності в біологічних процесах.

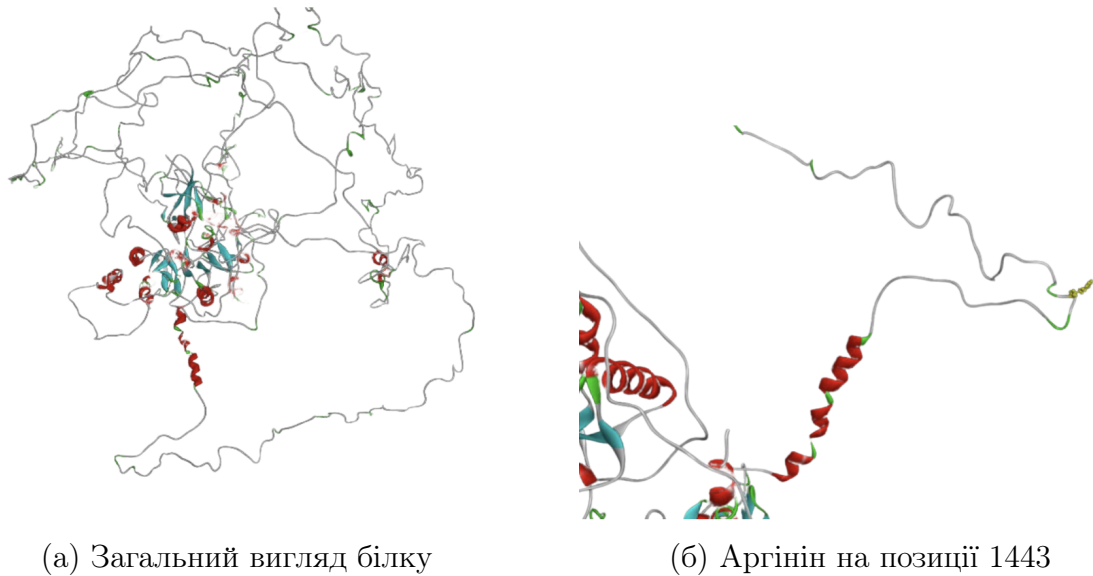


Рис. 2.9 — Система BRCA1 після проведення дослідження молекулярної динаміки

Як видно з рис. 2.9б, позиція 1443 залишається вивернутою назовні та знаходиться найдалі від центру молекули. Це може свідчити про її потенційну взаємодію з ДНК або іншими молекулами, що сигналізує, що мутації на цій позиції можуть мати значущий вплив.

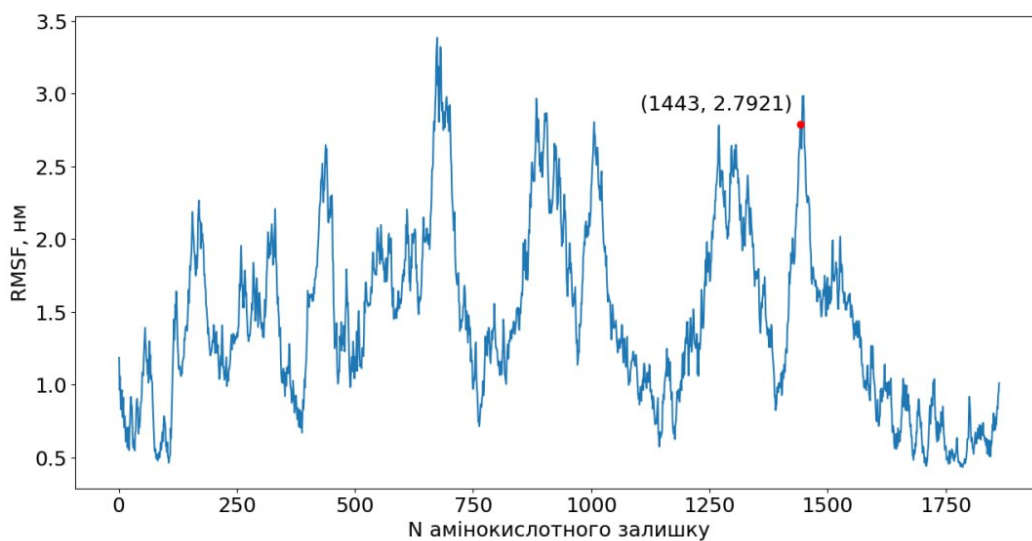


Рис. 2.10 — Графік Root Mean Square Fluctuation. Червоною точкою позначене значення на позиції 1443.

На графіку 2.10 зображено середньоквадратичні відхилення флуктуацій (RMSF) атомів білка BRCA1 по всій його довжині. Графік показує, як змінюється гнучкість різних амінокислотних залишків у білку під час симуляції молекулярної динаміки. Особливий інтерес представляє позиція 1443, яка показана червоною точкою на графіку.

#### 1. Загальний вигляд графіка:

Графік показує різні рівні флуктуацій по всій довжині білка. Деякі області демонструють високі флуктуації, що свідчить про їхню більшу гнучкість, тоді як інші області мають низькі флуктуації, що свідчить про їхню стабільність.

#### 2. Позиція 1443:

Значення RMSF для позиції 1443 складає приблизно 2.7921 нм, що є одним із найвищих значень на графіку. Це вказує на те, що цей амінокислотний залишок має високу гнучкість порівняно з іншими частинами білка.

#### Інтерпретація результатів

##### 1. Гнучкість позиції 1443:

Високе значення RMSF для позиції 1443 свідчить про значну гнучкість

цього амінокислотного залишку. Це може мати важливі наслідки для структурної стабільності та функціональності білка BRCA1.

## 2. Вплив мутацій:

Оскільки мутація на позиції 1443 (наприклад, R1443G або R1443K) може впливати на функціональність білка, висока гнучкість у цій області може бути критичною для розуміння механізмів, за допомогою яких ці мутації спричиняють функціональні зміни. Висока флуктуація може свідчити про те, що ця область є важливою для динамічних взаємодій білка, і мутації тут можуть порушувати ці взаємодії.

## 3. Порівняння з іншими областями:

Порівняння RMSF у позиції 1443 з іншими областями білка показує, що ця позиція є однією з найбільш гнучких. Це робить її ключовою для подальшого дослідження, зокрема, у контексті виявлення структурних змін, спричинених мутаціями.

## Розділ III. Обговорення та висновки

### 2.4 Вплив вірусів на еволюцію гена *BRCA1*

У нашій роботі було проведено детальне дослідження впливу вірусів на еволюцію гена *BRCA1*, зокрема через позитивний селекційний тиск. Хоча ми не довели існування позитивної селекції на позиції 1443, наші результати підтверджують наявність мутацій, що можуть бути пов'язані з вірусними інфекціями.

Результати експериментальної частини свідчать про те, що віруси можуть відігравати ключову роль у модифікації генетичного матеріалу. Використання BLAST для аналізу послідовностей показало потенційні сайти інтеграції вірусних геномів у локусі *BRCA1*. Це може призвести до змін у функціонуванні білка, що підвищує ризик розвитку раку.

Окрім використання BLAST і молекулярної динаміки, є кілька інших методів, які можуть бути застосовані для дослідження впливу вірусів на еволюцію гена *BRCA1*: 1. Повне геномне секвенування уможливорює виявлення вірусних послідовностей, інтегрованих у геном хазяїна. Це дозволяє ідентифікувати точні місця інтеграції вірусного ДНК та визначити, чи впливають вони на функції сусідніх генів, включаючи *BRCA1*.

2. Віруси можуть впливати на епігенетичні модифікації, такі як метилювання ДНК, що змінює експресію генів. Аналіз метилового профілю клітин, інфікованих вірусами, дозволяє визначити, чи є зміни в метилюванні в ділянках, що контролюють експресію *BRCA1*.

3. Клітинні та тваринні моделі — використання клітинних ліній та генетично модифікованих тварин, інфікованих вірусами, дозволяє дослідити функціональні наслідки вірусної інфекції на ген *BRCA1*. Це включає вивчення змін у стабільності білка, експресії генів і чутливості до канцерогенів.

4. Аналіз протеомного профілю клітин, інфікованих вірусами, може вияви-

ти зміни у взаємодії білків, що беруть участь у репарації ДНК та інших клітинних процесах, пов'язаних з *BRCA1*.

5. Структурна біологія, включаючи рентгенівську кристалографію та криоелектронну мікроскопію, може використовуватися для дослідження взаємодій між вірусними білками і білком *BRCA1*. Це дозволить виявити, як вірусні білки змінюють структуру та функції *BRCA1*.

Окрім вірусних інфекцій, на еволюцію гена *BRCA1* можуть впливати й інші фактори:

1. Екологічні фактори:

Радіація, хімічні канцерогени та інші екологічні стресори можуть спричинити пошкодження ДНК, що, в свою чергу, може впливати на мутації у гені *BRCA1* і сприяти його еволюції.

2. Генетичний дрейф:

У малих популяціях випадкові зміни у частоті алелів можуть мати значний вплив на еволюцію гена *BRCA1*, незалежно від природного добору.

3. Потік генів:

Міграція особин між популяціями може призводити до перенесення мутацій у гені *BRCA1* з однієї популяції до іншої, змінюючи генетичний склад популяцій.

4. Селекція у відповідь на захворювання:

Інфекційні захворювання та інші патологічні стани можуть створювати селекційний тиск на ген *BRCA1*, сприяючи розвитку адаптивних мутацій.

## 2.5 Молекулярна динаміка

Молекулярна динаміка (МД) дозволила дослідити вплив мутацій R1443G та R1443K на стабільність білка *BRCA1*. На графіку мінімізації енергій (рис. 2.5) видно, що потенціальна енергія системи для мутованих білків (R1443G та R1443K) є вищою, ніж для нормального білка. Це свідчить про те, що мутації призводять до меншої стабільності білка, оскільки більш ви-

сока енергія вказує на більш нестабільну структуру. Різниця у енергіях між нормальним білком і мутованими становить приблизно 0.5 кДж/моль, що є статистично значущою і підтверджує гіпотезу про зниження стабільності білка через мутації.

Графік тиску після мінімізації (рис. 2.6) показує коливання тиску в межах від -100 до +100 бар для всіх трьох систем. Незважаючи на відмінності у стабільності білків, відображені в потенціальній енергії, тиск у всіх трьох системах показує схожу поведінку після мінімізації. Це свідчить про те, що хоча мутації впливають на стабільність білка, вони не мають значного впливу на загальний тиск у системі після досягнення рівноважного стану.

Графік густини (рис. 2.7) демонструє, що всі три системи досягають стабільного рівноважного стану з густиною, що коливається навколо значення 1008 кг/м<sup>3</sup>. Незважаючи на структурні відмінності, викликані мутаціями, поведінка густини вказує на успішну стабілізацію систем після початкового періоду мінімізації. Це важливо для підтвердження коректності умов симуляції та дозволяє продовжувати дослідження динамічних властивостей білків з упевненістю в точності отриманих результатів.

На графіку RMSF (Root Mean Square Fluctuation) (рис. 2.10) видно, що значення RMSF для позиції 1443 складає приблизно 2.7921 нм, що є одним із найвищих значень на графіку. Це вказує на те, що цей амінокислотний залишок має високу гнучкість порівняно з іншими частинами білка. Високе значення RMSF для позиції 1443 свідчить про значну гнучкість цього амінокислотного залишку, що може мати важливі наслідки для структурної стабільності та функціональності білка BRCA1.

## 2.6 Медичні імплікації

Отримані результати мають важливі медичні наслідки. Вони вказують на необхідність врахування вірусних інфекцій як потенційного фактора ризику при генетичному скринінгу на мутації BRCA1. Розуміння механізмів,

за допомогою яких віруси впливають на еволюцію генів, може допомогти у розробці нових терапевтичних підходів, спрямованих на запобігання розвитку раку.

Наприклад, вакцинація проти ВПЛ може знизити ризик розвитку раку молочної залози у носіїв мутацій BRCA1, оскільки вона запобігає інтеграції вірусного генетичного матеріалу в геном хазяїна. Також можливо, що антивірусні терапії можуть бути ефективними у запобіганні мутацій, спричинених вірусами.

## 2.7 Популяційні дослідження

Можливість проведення популяційних досліджень у цій сфері є надзвичайно перспективною. Моделювання поширення різних генотипів гена BRCA1 у популяціях, знаючи селективний коефіцієнт, може дати уявлення про динаміку еволюції цього гена. Використання таких моделей дозволить оцінити вплив різних факторів, включаючи вірусні інфекції, на поширення мутацій BRCA1.

Селективний коефіцієнт, який визначає відносну виживаність різних генотипів, може бути використаний для прогнозування частоти мутацій у популяціях. Ці дані можуть допомогти у розробці стратегій скринінгу та профілактики раку, орієнтованих на популяції з високим ризиком. Виклики популяційних досліджень:

1. Одним із головних викликів є зібрання достатньої кількості генетичних даних з різних популяцій. Це включає не лише геномне секвенування, але й дані про здоров'я, екологічні умови та соціально-економічні фактори.
2. Популяції можуть бути генетично неоднорідними, що ускладнює аналіз та інтерпретацію даних. Потрібно враховувати такі фактори, як генетичний дрейф, ефекти засновника та міграцію.
3. Моделювання поширення генетичних варіацій потребує використання складних математичних моделей. Важливо правильно визначити параме-

три моделей, такі як селективний коефіцієнт, частота мутацій, розмір популяції та інші.

4. Використання генетичних даних з різних популяцій піднімає питання конфіденційності та згоди учасників дослідження. Необхідно забезпечити, щоб дослідження проводилися відповідно до етичних норм та стандартів.

Потенційна користь популяційних досліджень

1. Популяційні дослідження допомагають зрозуміти, як гени, такі як *BRCA1*, еволюціонують під впливом різних факторів. Це включає вивчення впливу вірусних інфекцій, екологічних стресорів та інших селективних тисків.

2. Дослідження частот мутацій у різних популяціях дозволяє ідентифікувати групи з підвищеним ризиком розвитку генетичних захворювань, таких як рак молочної залози та яєчників. Це сприяє розробці цільових програм скринінгу та профілактики.

3. Розуміння механізмів, за допомогою яких віруси та інші фактори впливають на еволюцію генів, може призвести до розробки нових терапевтичних підходів. Це включає розробку вакцин, антивірусних препаратів та генетичної терапії.

4. Інформація, отримана з популяційних досліджень, може бути використана для покращення політик громадського здоров'я. Це включає рекомендації щодо вакцинації, скринінгу та інших профілактичних заходів.

Майбутні популяційні дослідження можуть зосередитися на інтеграції багатовимірних даних, включаючи геномні, епігеномні, протеомні та метаболомні профілі. Це дозволить отримати більш повну картину того, як різні фактори впливають на еволюцію генів і розвиток захворювань. Використання великих біобанків та координація міжнародних досліджень може значно розширити наші знання про генетичні варіації та їх вплив на здоров'я людини.

Популяційні дослідження також можуть враховувати вплив соціально-економічних факторів, культурних практик та доступу до медичної до-

помоги на поширення генетичних варіацій. Це допоможе розробити більш інклюзивні та ефективні стратегії профілактики та лікування захворювань у різних популяціях.

## **Висновки**

**BLAST показав можливі сайти інтеграцій вірусного геному у хромосому 17, що може свідчити про складну взаємодію вірусних геномів із людським.**

У роботі за допомогою інструменту BLAST було виявлено потенційні місця інтеграції вірусних геномів всередині та навколо локусу *BRCA1*. Отримані дані свідчать про те, що вірусні взаємодії можуть призводити до значних модифікацій генетичного матеріалу, потенційно впливаючи на функцію білка BRCA1 та підвищуючи ризик розвитку раку. Аналіз вказує на те, що віруси можуть відігравати вирішальну роль в еволюції генів, вносячи мутації або застосовуючи еволюційний тиск, про що свідчать виявлені у роботі та поза нею сайти інтеграції.

**Мутації на позиції 1443 знижують стабільність білку, підвищуючи потенціальну енергію**

Для оцінки впливу мутацій у положенні 1443 білка BRCA1 було проведено моделювання молекулярної динаміки. Попередні показують, що мутації R1443G і R1443K призводять до зниження стабільності білка порівняно з диким типом. Попереднє моделювання показує, що ці мутації призводять до підвищення потенціальної енергії та зниження здатності досягати енергетично вигідних конформацій, що може негативно впливати на функцію білка. У дослідженні зроблено висновок, що хоча мутації впливають на стабільність білка, загальна щільність систем стабілізується аналогічно, що свідчить про успішну мінімізацію енергії в усіх системах.

У подальшому варто провести повноцінне моделювання молекулярної динаміки мутованих білків, а також білки у комплексі із ДНК, аби оцінити експонованість амінокислоти 1443 до ДНК.

## Перелік використаних джерел

- [1] A. Antoniou, P. D. Pharoah, S. Narod, et al. Average risks of breast and ovarian cancer associated with *BRCA1* or *BRCA2* mutations detected in case Series unselected for family history: A combined analysis of 22 studies. *American Journal of Human Genetics*, 72(5):1117–1130, 2003.
- [2] D.I. Lou, R.M. McBee, U.Q. Le, A.C. Stone, G.K. Wilkerson, A.M. Demogines, and S.L. Sawyer. Rapid evolution of *BRCA1* and *BRCA2* in humans and other primates. *BMC Evolutionary Biology*, 14:155, Jul 2014.
- [3] World Health Organization. Breast Cancer Fact Sheet. <https://www.who.int/news-room/fact-sheets/detail/breast-cancer>, 2024. Accessed: insert access date here.
- [4] S. Antonarakis. The history of genetic disease discovery. *American Journal of Medical Genetics Part A*, 2021.
- [5] M. L. Benton, A. Abraham, et al. The influence of evolutionary history on human health and disease. *Nature Reviews Genetics*, 2021.
- [6] D. Toncheva, Maya Atanasoska, et al. Genetic diseases in ancient Caribbean communities. *Comptes rendus de l'Académie bulgare des Sciences*, 2023.
- [7] E. N. Gordienko. The history and formation of genetics. *Journal*, 2021.
- [8] S. Murphy, Brandon Nguyen, et al. The history of ankylosing spondylitis. *Surgical Neurology International*, 2022.
- [9] P. Campitelli, T. Modi, and S. Banu Ozkan. Probing impact of allosteric

(or distal) mutations on protein function through time-dependent linear response theory. *Biophysical Journal*, 123(3), 2023.

- [10] Sarahi L. Garcia Matthew Kellom Katherine D. McMahon Brett J. Baker Robin R. Rohwer, Mark Kirkpatrick. Probing impact of allosteric (or distal) mutations on protein function through time-dependent linear response theory. *bioRxiv*, Feb 2024. Preprint.
- [11] E. Nicholson, A. Andrade, T.M. Brooks, et al. Roles of the Red List of Ecosystems in the Kunming-Montreal Global Biodiversity Framework. *Nature Ecology Evolution*, 2024.
- [12] R. Frankham, D.A. Briscoe, and J.D. Ballou. *Introduction to Conservation Genetics*. Cambridge University Press, 2002.
- [13] J. Hall, Ming K. Lee, B. Newman, Jan E. Morrow, L. Anderson, B. Huey, and M. King. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science*, 250(4988):1684–1689, 1990.
- [14] National Center for Biotechnology Information (NCBI). BRCA1 breast cancer type 1 susceptibility protein [Homo sapiens (human)]. <https://www.ncbi.nlm.nih.gov/gene/672>, 2024. Accessed: 2024-01-31.
- [15] M. E. Moynahan, J. W. Chiu, B. H. Koller, and M. Jasin. Brca1 controls homology-directed DNA repair. *Molecular Cell*, 4(4):511–518, 1999.
- [16] R. Scully, J. Chen, A. Plug, et al. Association of BRCA1 with Rad51 in mitotic and meiotic cells. *Cell*, 88(2):265–275, 1997.
- [17] L. Zheng, S. Li, T. G. Boyer, and W. H. Lee. Lessons learned from BRCA1 and BRCA2. *Oncogene*, 19(53):6159–6175, 2000.
- [18] F. Wu-Baer, K. Lagrazon, W. Yuan, and R. Baer. The BRCA1/BARD1 heterodimer assembles polyubiquitin chains through an unconventional

- linkage involving lysine residue K6 of ubiquitin. *Journal of Biological Chemistry*, 278(37):34743–34746, 2003.
- [19] DeepMind Technologies. AlphaFold Protein Structure Database. <https://alphafold.ebi.ac.uk/>, 2024. Accessed on May 8, 2024.
- [20] R. Baer, F. Wu-Baer, K. Lagrazon, and W. Yuan. The BRCA1/BARD1 heterodimer assembles polyubiquitin chains through an unconventional linkage involving lysine residue K6 of ubiquitin. *Journal of Biological Chemistry*, 278(37):34743–34746, 2003.
- [21] S. Sankaran, D.E. Crone, R.E. Palazzo, and J.D. Parvin. Aurora-A kinase regulates breast cancer associated gene 1 inhibition of centrosome-dependent microtubule nucleation. *Cancer Research*, 67(23):11186–11194, 2007.
- [22] P.S. Brzovic, P. Rajagopal, D.W. Hoyt, M.-C. King, and R.E. Klevit. Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. *Nature Structural Biology*, 8:833–837, 2001.
- [23] C.J. Vandenberg, F. Gergely, C.Y. Ong, P. Pace, D.L. Mallery, K. Hiom, and K.J. Patel. BRCA1-independent ubiquitination of FANCD2. *Molecular Cell*, 12(1):247–254, 2003.
- [24] Shirley M.H. Sy, Michael S.Y. Huen, and Junjie Chen. PALB2 is an integral component of the brca complex required for homologous recombination repair. *Proceedings of the National Academy of Sciences of the United States of America*, 106(17):7155–7160, 2009.
- [25] X. Yu, Sh. Fu, M. Lai, R. Baer, and J. Chen. BRCA1 ubiquitinates its phosphorylation-dependent binding partner ctip. *Genes & Development*, 20(13):1721–1726, 2006.

- [26] H. Hiraike, O. Wada-Hiraike, S. Nakagawa, S. Koyama, Y. Miyamoto, K. Sone, M. Tanikawa, T. Tsuruga, Y. Nagasaka, K. and Matsumoto, K. Oda, K. Shoji, H. Fukuhara, S. Saji, K. Nakagawa, S. Kato, T. Yano, and Y. Taketani. Identification of DBC1 as a transcriptional repressor for BRCA1. *British Journal of Cancer*, 102(6):1061–1067, 2010.
- [27] P. Murthy and F. Muggia. Women’s cancers: how the discovery of BRCA genes is driving current concepts of cancer biology and therapeutics. *ecancermedicalscience*, 13, 2019.
- [28] H. E. Bryant, N. Schultz, H. D. Thomas, et al. Specific killing of *BRCA2*-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature*, 434(7035):913–917, 2005.
- [29] H. Farmer, N. McCabe, C. J. Lord, et al. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature*, 434(7035):917–921, 2005.
- [30] BRCA Exchange. BRCA exchange: Comprehensive information on brca1 and brca2 variants. <https://brcaexchange.org/>. Accessed: Jan 31, 2024.
- [31] National Center for Biotechnology Information (NCBI), U.S. National Library of Medicine. ClinVar. <https://www.ncbi.nlm.nih.gov/clinvar>. Accessed: Jan 31, 2024.
- [32] S. Chen and G. Parmigiani. Meta-analysis of *BRCA1* and *BRCA2* penetrance. *Journal of Clinical Oncology*, 25(11):1329–1333, 2007.
- [33] Peter J O’Donovan and DM Livingston. BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in dna double-strand break repair. *Carcinogenesis*, 31:961–967, 2010.

- [34] Daniel Hemel and Susan M Domchek. Breast Cancer Predisposition Syndromes. *Hematol Oncol Clin North Am*, 24:799–814, 2010.
- [35] T. Ludwig, D. L. Chapman, V. E. Papaioannou, and A. Efstratiadis. Targeted mutations of breast cancer susceptibility gene homologs in mice: lethal phenotypes of BRCA1, BRCA2, BRCA1/BRCA2, BRCA1/p53, and BRCA2/p53 nullizygous embryos. *Genes & Development*, 11(10):1226–1241, 1997.
- [36] Kienan I. Savage and D. Paul Harkin. BRCA1, a 'complex' protein involved in the maintenance of genomic stability. *FEBS Journal*, 282(4):630–646, 2015.
- [37] A. G. Knudson. Mutation and cancer: Statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences*, 68(4):820–823, 1971.
- [38] S. J. Baker et al. Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science*, 244(4901):217–221, 1989.
- [39] L. M. Smith et al. Loss of heterozygosity: What is it good for? *BMC Genetics*, 17(1):84, 2016.
- [40] L. Sokolovska, M. Cistjakovs, A. Matroze, M. Murovska, et al. From Viral Infection to Autoimmune Reaction: Exploring the Link between Human Herpesvirus 6 and Autoimmune Diseases. *Microorganisms*, 2024.
- [41] H.N. Pennington and J. Lee. Influence of physiologically relevant anionic lipids on the initiation of fusion by the Lassa virus fusion domain. *Biophysical Journal*, 2024.
- [42] M.E. Melo and J.M. Wereszczynski. The impact of histone modifications and DNA sequence on the dynamics and stability of viral nucleosome-like particles. *Biophysical Journal*, 2024.

- [43] S. Venkataraman, M. Shahgolzari, A. Yavari, and K. Hefferon. Delivery of Therapeutics Using Bacteriophage Vectors. *Preprints*, 2024. Preprint.
- [44] J. Chiti. The impact of genetics on viral pathogenesis and immune response. *Viol Res J*, 2023.
- [45] T. Tarver. Cancer Facts & Figures 2012. *Journal of Consumer Health on the Internet*, 16(3):366–367, 2012. Available from American Cancer Society, 66 p., pdf.
- [46] K. Alibek, A. Kakpenova, and Y. Baiken. Role of infectious agents in the carcinogenesis of brain and head and neck cancers. *Infectious Agents and Cancer*, 8(1):1, 2013.
- [47] M. K. Amarante and M. A. E. Watanabe. The possible involvement of virus in breast cancer. *Journal of Cancer Research and Clinical Oncology*, 135(3):329–337, 2009.
- [48] B. Friedenson. In BRCA1 and BRCA2 Breast Cancers, Chromosome Breaks Occur Near Herpes Tumor Virus Sequences. 2021.
- [49] Samah A. Loutfy, Zeinab F. Abdallah, M. Shaalan, M. Moneer, A. Karam, M. M. Moneer, I. M. Sayed, A. El-Hafeez, P. Ghosh, and A-R. N. Zekri. Prevalence of MMTV-Like env Sequences and Its Association with BRCA1/2 Genes Mutations Among Egyptian Breast Cancer Patients. *Cancer Management and Research*, 13:2835–2848, 2021.
- [50] S. H. M. Ali, T. F. AL-Mahbobi, Y. J. AL-Bayaa, and S. H. M. Al-Alwany. Impact of Human Cytomegalovirus Infection associated with the expressed protein of mutated BRCA1 gene in breast tissues from a group of Iraqi Female Patients with Breast Carcinoma. *Research Journal of Pharmacy and Technology*, 11(4):1505–1512, 2018.

- [51] D. Dutta, M. A. Ansari, and B. Chandran. BRCA1-regulated nuclear innate sensing of herpesviral genome by IFI16 and IFI16's acetylation is critical for its cytoplasmic trafficking and induction of innate responses. *Inflamm Cell Signal*, 2:e1076, 2015.
- [52] A. Antonsson, S. Bialasiewicz, R. J. Rockett, K. Jacob, I. C. Bennett, and T. P. Sloots. Exploring the prevalence of ten polyomaviruses and two herpes viruses in breast cancer. *PLOS ONE*, 7(8):e39842, 2012.
- [53] M. Hachana, K. Amara, S. Ziadi, R. B. Gacem, S. Korbi, and M. Trimeche. Investigation of human JC and BK polyomaviruses in breast carcinomas. *Breast Cancer Research and Treatment*, 133(3):969–977, 2012.
- [54] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. *Human papillomaviruses*, volume 90. International Agency for Research on Cancer, 2007.
- [55] Y. W. Wang, H. S. Chang, C. H. Lin, and C. Y. Winston. HPV-18 E7 conjugates to c-Myc and mediates its transcriptional activity. *The International Journal of Biochemistry & Cell Biology*, 39(2):402–412, 2007.
- [56] A. Yasmeen, T. A. Bismar, M. Kandouz, W. D. Foulkes, P. Y. Desprez, and A. E. Al Moustafa. E6/E7 of HPV type 16 promotes cell invasion and metastasis of human breast cancer cells. *Cell Cycle*, 6(16):2038–2042, 2007.
- [57] G.A. Huttley, S. Easteal, M.C. Southey, A. Tesoriero, G.G. Giles, M.R.E. McCredie, J.L. Hopper, and D.J. Venter. Adaptive evolution of the tumour suppressor BRCA1 in humans and chimpanzees. Australian Breast Cancer Family Study. *Nature Genetics*, 25(4):410–413, Aug 2000.

- [58] N.R. Meyerson and S.L. Sawyer. Two-stepping through time: mammals and viruses. *Trends in Microbiology*, 19:286–294, 2011.
- [59] L. H. Castilla, M. R. Erdos, L. C. Brody, F.S. Collins, F. J. Couch, K. F. Hoskins, K. Calzone, M. L. Deshano, B. L. Weber, J. E. Garber, J. Boyd, and M. B. Lubin. Mutations in the *BRCA1* gene in families with early-onset breast and ovarian cancer. *Nature Genetics*, 8(4):387–391, 1994.
- [60] Y. Funato, T. Terabayashi, R. Sakamoto, D. Okuzaki, H. Ichise, H. Nojima, N. Yoshida, and H. Miki. Nucleoredoxin sustains Wnt/-catenin signaling by retaining a pool of inactive dishevelled protein. *Current Biology*, 20(21):1945–1952, 2010.
- [61] STRING Consortium. STRING: functional protein association networks. <https://string-db.org/>, 2024. Accessed: 2024-05-15.
- [62] A. L. Palser, N. E. Grayson, R. E. White, C. Corton, S. Correia, Mohammed M. Ba A., and S. JI Watson. Genome diversity of Epstein-Barr virus from multiple tumor types and normal infection. *Journal of Virology*, 89(10):5222–5237, 2015.
- [63] L Yang et al. Molecular Characterization of HBV DNA Integration in Patients with Hepatitis and Hepatocellular Carcinoma. *Journal of Cancer*, 9:3225–3235, 2018.
- [64] UniProt Consortium. UniProtKB - P38398 (BRCA1\_HUMAN). <https://www.uniprot.org/uniprotkb/P38398/entry>, 2024. Accessed on May 8, 2024.