

Київський національний університет імені Тараса Шевченка  
Міністерство освіти і науки України  
Київський національний університет імені Тараса Шевченка  
Міністерство освіти і науки України

Кваліфікаційна наукова  
праця на правах рукопису

**СТРАТІЙЧУК РОМАН ОЛЕКСАНДРОВИЧ**

УДК 577.352.4 + 577.322

**ДИСЕРТАЦІЯ**  
**ПРОГНОЗУВАННЯ МОЛЕКУЛЯРНИХ ВЗАЄМОДІЙ ХІМІЧНИХ**  
**СПОЛУК З РЕЦЕПТОР-КЕРОВАНИМ КАТІОННИМ КАНАЛОМ**  
**TRPC4 ТА ОЦІНКА ЇХ ФАРМАКОЛОГІЧНИХ ЕФЕКТІВ**

Спеціальність – 091 Біологія та біохімія

Галузь знань – 09 Біологія

Подається на здобуття наукового ступеня доктора філософії в галузі біології

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

\_\_\_\_\_ Р.О. Стратійчук

Науковий керівник:

Жолос Олександр Вікторович, доктор біологічних наук, професор

Київ – 2026

## АНОТАЦІЯ

**Стратійчук Р. О. «Прогнозування молекулярних взаємодій хімічних сполук з рецептор-керованим катіонним каналом TRPC4 та оцінка їх фармакологічних ефектів».** – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 091 «Біологія та біохімія» (09 – Біологія) – Київський національний університет імені Тараса Шевченка, Київ, 2026.

Канали транз'єнтного рецепторного потенціалу підродини канонічних (TRPC), насамперед TRPC4 і структурно найбільш близький до нього TRPC5, є рецептор-керованими катіонними каналами, що відіграють ключову роль у кальцієвій сигналізації та функціонуванні низки фізіологічних систем. У шлунково-кишковому тракті (ШКТ), TRPC4 опосередковує мускариновий катіонний струм ( $mI_{CAT}$ ) – ключовий механізм холінергічної регуляції скоротливої активності гладеньких м'язів, забезпечуючи близько 85% амплітуди цього струму в міоцитах клубової кишки миші. Поза ШКТ, TRPC4 залучений до сигнальних процесів у нейрональних, ендотеліальних та подоцитарних клітинах і розглядається як перспективна фармакологічна мішень при функціональних розладах моторики кишечника, фокально-сегментарному гломерулосклерозі, тривожно-депресивних та інших патологічних станах. Попри встановлену фармакологічну значущість TRPC4, наявність селективних лігандів цього каналу обмежена, а молекулярні механізми їхньої дії на мишачий TRPC4 вивчено недостатньо: для Pico145 (HC-608) дані кріо-електронної мікроскопії (кріо-ЕМ) доступні лише для людського TRPC5 (PDB ID: 6YSN) та TRPC1/TRPC4 (PDB ID: 8WPM), а для GFB-887, клінічно досліджуваного селективного інгібітора TRPC5, структурні дані у комплексі з представниками підгрупи TRPC1/4/5 відсутні взагалі. Тому встановлення особливостей молекулярних взаємодій Pico145 і GFB-887 з TRPC4 є актуальною задачею, розв'язання якої створює

структурно-функціональну основу для раціонального дизайну селективних модуляторів каналів цієї підродини.

Метою дисертаційного дослідження було встановлення особливостей молекулярних взаємодій TRPC4 з малими молекулами-інгібіторами Pico145 та GFB-887 шляхом поєднання методів молекулярного моделювання, машинного навчання, електрофізіології та тензометрії для функціональної оцінки цих взаємодій.

Для виконання роботи використовували наступні методи: ідентифікація сайтів зв'язування з застосуванням розробленого обчислювального гібридного підходу, що поєднує геометричне прогнозування порожнин програмою Focket з літературно-зумовленою пріоритизацією амінокислотних залишків з використанням великої мовної моделі GPT-4o-mini; молекулярний докінг (AutoDock 4); молекулярна динаміка повнорозмірних мембранних комплексів (GROMACS); прогнозування ADMET-параметрів власноруч розробленими моделями машинного навчання; реєстрація трансмембранних іонних струмів методом петч-клемп у конфігурації «ціла клітина» в ізольованих міоцитах клубової кишки миші; ізометрична тензометрія на сегментах тонкої кишки миші.

В межах дослідження розроблено гібридний метод ідентифікації та пріоритизації сайтів зв'язування, який поєднує геометричний аналіз порожнин білка з автоматизованим отриманням експериментально підтверджених амінокислотних залишків із наукової літератури за допомогою великих мовних моделей. Метод реалізовано як рецензований програмний інструмент з відкритим вихідним кодом, опубліковано у міжнародному фаховому виданні Bioinformatics (Oxford University Press) і розміщено у GitHub та Zenodo. Принциповою перевагою підходу є те, що він замінює тривалу і суб'єктивну процедуру ручного аналізу літератури автоматизованим відтворюваним протоколом і не залежить від наявності лігандів у досліджуваних PDB структурах. Застосування цього методу до кріо-ЕМ структури мишачого TRPC4 (PDB ID: 5Z96) дозволило виокремити чотири

біологічно релевантні сайти, кожен з яких представлений чотирма еквівалентними копіями: катіон-зв'язувальний сайт у ділянці S1–S4 VSL домену; сайт зв'язування фосфатидної кислоти на інтерфейсі порової спіралі однієї субодиниці та спіралі S6 іншої субодиниці;  $\text{Ca}^{2+}$ -зв'язувальний сайт у VSL-домени; та сайт зв'язування інгібіторів у VSL-домени. Важливо, що три з чотирьох ідентифікованих сайтів локалізовані в межах VSL-домену, що формує ключову структурно-регуляторну область каналу.

Уперше побудовано молекулярні моделі комплексів мишачого TRPC4 з інгібіторами Pico145 і GFB-887. Молекулярний докінг Pico145 було попередньо здійснено щодо референтної структури гетеротетрамеру TRPC1/TRPC4 людини (PDB ID: 8WPM) з метою відтворення експериментальної пози ліганду (енергія зв'язування  $-10,3$  ккал/моль). Шляхом подальшого докінгу Pico145 у апо-структуру мишачого TRPC4 (PDB ID: 5Z96) було визначено і охарактеризовано комплекс з енергією  $-9,9$  ккал/моль зі стабільним розміщенням ліганду у консервативній міжсубодиничній гідрофобній кишені трансмембранній ділянці білка: полярний фрагмент Pico145 формує водневий зв'язок з Trp573, ароматичний фрагмент фіксується через  $\pi$ - $\pi$  (T-shaped) взаємодію з Tyr523 та  $\pi$ -sulfur контакт з Cys524, гідрофобне оточення формується залишками Leu520, Leu527, Leu568, Phe565 і Phe572. Порівняння амінокислотного оточення між мишачим TRPC4 і людським TRPC1/TRPC4 виявило високий ступінь консервативності контактних залишків за наявності лише однієї напівконсервативної заміни Thr599 (миша) на Val624 (людина). Оцінка стабільності комплексу методом молекулярної динаміки тривалістю близько 215 нс підтвердила збереження ключових контактів ліганда з кишенею в усіх чотирьох симетричних копіях зі стаціонарним рівнем RMSD у діапазоні  $1,5$ – $2,5$  Å та утриманням  $1,5$ – $2,0$  водневих зв'язків на ліганд з оптимальною геометрією.

Потенційний сайт GFB-887 локалізовано у трансмембранному регіоні S1–S4 VSL-домену TRPC4. Отримані дані сумісні з кріо-ЕМ структурами



TRPC4 *Danio rerio* у комплексах з аналогами серії GFB – GFB-8438, GFB-8749 та GFB-9289 (PDB ID: 7B0S, 7B05, 7B16). В результаті молекулярного докінгу було отримано комплекс з енергетично сприятливою конформацією ліганду з енергією  $-8,11$  ккал/моль і мережею контактів з ключовими залишками Phe413, Tyr373, Tyr646, His369, Glu417, Asp438, Asn442, Arg491, Ser494, Met441, Leu492 та Leu495. Гетероциклічне ядро GFB-887 утворює  $\pi$ - $\pi$  stacking,  $\pi$ - $\pi$  T-shaped,  $\pi$ -alkyl та  $\pi$ - $\sigma$  взаємодії з гідрофобно-ароматичними залишками, тоді як полярні і електростатичні контакти забезпечують орієнтацію молекули у кишені. Молекулярна динаміка комплексу TRPC4:GFB-887 тривалістю близько 200 нс підтвердила стабільність полярних якорів зв'язування (His369, Asn442, Arg491, Ser494) і збереження гідрофобно-ароматичного оточення впродовж усієї траєкторії. Важливо підкреслити, що обидва інгібітори потенційно взаємодіють у трансмембранному регіоні каналу, але займають різні кишені – Pico145 у міжсубодиничній порожнині, GFB-887 у внутрішньосубодиничному сайті VSL-домену, відповідає двом різним молекулярним механізмам інгібування TRPC4.

Розроблено метод прогнозування параметрів абсорбції, розподілу, метаболізму, екскреції та потенційних токсичних ефектів (ADMET) на основі гібридних молекулярних представлень – конкатенації 512-вимірних ембедингів, отриманих шляхом навчання Mol2Vec моделі на даних з бази хімічних сполук ZINC20 (883 897 271 сполук) з класичними дескрипторами та фінгерпринтами. На стандартизованих бенчмарк-датасетах платформи Therapeutics Data Commons (TDC) архітектура моделі посідає перше місце у 10 з 16 бенчмарків. Результати оприлюднено на сервері препринтів bioRxiv. Із застосуванням цього методу вперше охарактеризовано прогнозований ADMET-профіль для GFB-887 та Pico145, для яких у відкритих джерелах неповні або відсутні відповідні експериментальні дані.

Попередній аналіз свідчить про те, що обидві сполуки демонструють ефективну кишкову абсорбцію та високе зв'язування з білками плазми, проте

розрізняються за метаболічним і токсикологічним профілями: GFB-887 не виявляє ознак інгібування основних ізоферментів CYP, тоді як Pico145 прогнозовано інгібує CYP3A4, CYP2C9, CYP2D6 та P-глікопротеїн і має високу прогнозовану ймовірність гепатотоксичності (DILI). Параметри DILI та блокування hERG для обох сполук визначено як пріоритетні для подальших експериментальних досліджень.

Експериментальну валідацію обчислювальних гіпотез виконано на двох рівнях організації: реєстрацією трансмембранних іонних струмів методом петч-клемп у конфігурації «ціла клітина» в ізольованих міоцитах клубової кишки миші (клітинний рівень) та ізометричною тензометрією на сегментах тонкої кишки миші (тканинний рівень). Методом петч-клемп у конфігурації «ціла клітина» в ізольованих міоцитах клубової кишки миші охарактеризовано кінетичні та потенціалзалежні особливості блокади Pico145 рецептор-керованого катіонного струму  $mI_{CAT}$ : інгібування розвивалось повільно з середньою постійною часу  $140 \pm 6$  с ( $n = 9$ ) з вираженою потенціалзалежністю дії цього ліганду – при  $-120$  мВ інгібування було менш ефективним порівняно з  $80$  мВ. Сукупність цих ознак, а саме висока афінність (діюча концентрація  $100$  пМ), повільна кінетика та чутливість до трансмембранного потенціалу є прямим функціональним свідченням того, що сайт зв'язування Pico145 локалізований у площині мембрани і потребує дифузії ліганду через ліпідне середовище, що повністю узгоджується з обчислювально передбаченим міжсубодиничним гідрофобним сайтом. Методом ізометричної тензометрії на ізольованих сегментах *ileum* миші вперше показано, що 20-хвилинна попередня інкубація з GFB-887 у концентрації  $10$  мкмоль/л пригнічує карбахол ( $50$  мкмоль/л)-індуковану скоротливу відповідь на  $\approx 40\%$  (до  $\approx 60\%$  контрольного рівня). Це є першим функціональним свідченням фармакологічної активності GFB-887 у TRPC4-залежних нативних тканинах, що свідчить на користь того, що TRPC4 є перспективною терапевтичною мішенню при порушеннях моторики ШКТ.

Практичне значення отриманих результатів полягає у тому, що побудовані молекулярні моделі комплексів TRPC4 з Pico145 і GFB-887 формують структурне підґрунтя для раціонального дизайну селективних модуляторів каналів підгрупи TRPC1/4/5. Розроблені обчислювальні інструменти, а саме гібридний метод пріоритизації сайтів зв'язування і ADMET-моделі, мають універсальний характер і не потребують перенавчання чи адаптації і можуть бути застосовані до інших мембранних та розчинних білкових мішеней та оцінки ADMET-властивостей досліджуваних сполук, що забезпечує їх самостійну наукову та практичну цінність та можливість інтеграції у дослідження інших наукових груп на ранніх етапах процесу розробки лікарських засобів.

Основні теоретичні положення, методичні підходи та практичні результати проведеного дисертаційного дослідження знайшли своє повне відображення у 6 наукових працях, з них: 1 стаття у міжнародному фаховому періодичному виданні, що індексується у наукометричних базах Scopus та Web of Science (із квартилем Q1), та 2 статті у фахових періодичних виданнях України. Апробацію результатів дисертаційної роботи було проведено в межах міжнародних науково-практичних конференцій з опублікуванням 3 тез доповідей.

**Ключові слова:** TRPC4; гладенькі м'язи; гладеньком'язові клітини; кальцієва сигналізація /  $\text{Ca}^{2+}$ -сигнал; молекулярний докінг; молекулярна динаміка; сайт зв'язування; комп'ютерне моделювання / комп'ютерна симуляція; біомеханічні параметри м'язового скорочення; водневий зв'язок; конформаційний аналіз

## ABSTRACT

**Stratiichuk R. O. «Prediction of molecular interactions of chemical compounds with the receptor-operated cation channel TRPC4 and evaluation of their pharmacological effects».** – Qualification scientific work on the rights of the manuscript.

Dissertation for the degree of Doctor of Philosophy in specialty 091 “Biology and Biochemistry” (09 – Biology) – Taras Shevchenko National University of Kyiv, Kyiv, 2026.

Transient receptor potential channels of the canonical subfamily (TRPC), primarily TRPC4 and its structurally closest relative TRPC5, are receptor-operated cation channels that play a key role in calcium signaling and in the functioning of a number of physiological systems. In the gastrointestinal (GI) tract, TRPC4 mediates the muscarinic cationic current ( $mI_{CAT}$ ) – a key mechanism of cholinergic regulation of smooth muscle contractile activity, accounting for about 85% of the amplitude of this current in mouse ileal myocytes. Beyond the GI tract, TRPC4 is involved in signaling processes in neuronal, endothelial, and podocyte cells and is considered a promising pharmacological target in functional intestinal motility disorders, focal segmental glomerulosclerosis, anxiety and depressive disorders, and other pathological conditions. Despite the established pharmacological significance of TRPC4, the availability of selective ligands of this channel is limited, and the molecular mechanisms of their action on mouse TRPC4 have been insufficiently studied: for Pico145 (HC-608), cryo-electron microscopy (cryo-EM) data are available only for human TRPC5 (PDB ID: 6YSN) and TRPC1/TRPC4 (PDB ID: 8WPM), while for GFB-887, a clinically investigated selective inhibitor of TRPC5, structural data in complex with members of the TRPC1/4/5 subgroup are entirely absent. Therefore, establishing the features of the molecular interactions of Pico145 and GFB-887 with TRPC4 is a relevant task, the solution of which creates a structural and functional basis for the rational design of selective modulators of channels of this subfamily.

The aim of the dissertation research was to establish the features of the molecular interactions of TRPC4 with the small-molecule inhibitors Pico145 and GFB-887 by combining methods of molecular modeling, machine learning, electrophysiology, and isometric tension recording for the functional evaluation of these interactions.

The following methods were used to carry out the work: identification of binding sites using a developed computational hybrid approach that combines geometric prediction of cavities with the Fpocket program and literature-driven prioritization of amino acid residues using the large language model GPT-4o-mini; molecular docking (AutoDock 4); molecular dynamics of full-length membrane complexes (GROMACS); prediction of ADMET parameters with in-house developed machine learning models; recording of transmembrane ionic currents by the patch-clamp method in the whole-cell configuration in isolated mouse ileal myocytes; isometric tension recording on segments of mouse small intestine.

Within the framework of the study, a hybrid method for the identification and prioritization of binding sites was developed, which combines geometric analysis of protein cavities with the automated retrieval of experimentally confirmed amino acid residues from the scientific literature using large language models. The method was implemented as a peer-reviewed open-source software tool, published in the international peer-reviewed journal *Bioinformatics* (Oxford University Press) and deposited in GitHub and Zenodo. A fundamental advantage of the approach is that it replaces the lengthy and subjective procedure of manual literature analysis with an automated, reproducible protocol and does not depend on the presence of ligands in the PDB structures under study. Application of this method to the cryo-EM structure of mouse TRPC4 (PDB ID: 5Z96) made it possible to distinguish four biologically relevant sites, each represented by four equivalent copies: a cation-binding site in the S1–S4 region of the VSL domain; a phosphatidic acid binding site at the interface of the pore helix of one subunit and the S6 helix of another subunit; a  $\text{Ca}^{2+}$ -binding site in the VSL domain; and an inhibitor binding site in the VSL domain. Importantly, three of the four identified

sites are localized within the VSL domain, which forms the key structural and regulatory region of the channel.

For the first time, molecular models of complexes of mouse TRPC4 with the inhibitors Pico145 and GFB-887 were constructed. Molecular docking of Pico145 was first carried out against the reference structure of the human TRPC1/TRPC4 heterotetramer (PDB ID: 8WPM) in order to reproduce the experimental pose of the ligand (binding energy  $-10.3$  kcal/mol). Through subsequent docking of Pico145 into the apo-structure of mouse TRPC4 (PDB ID: 5Z96), a complex with an energy of  $-9.9$  kcal/mol was determined and characterized, with a stable placement of the ligand in a conserved intersubunit hydrophobic pocket in the transmembrane region of the protein: the polar fragment of Pico145 forms a hydrogen bond with Trp573, the aromatic fragment is fixed through a  $\pi$ - $\pi$  (T-shaped) interaction with Tyr523 and a  $\pi$ -sulfur contact with Cys524, while the hydrophobic environment is formed by the residues Leu520, Leu527, Leu568, Phe565, and Phe572. Comparison of the amino acid environment between mouse TRPC4 and human TRPC1/TRPC4 revealed a high degree of conservation of the contact residues, with only a single semi-conserved substitution of Thr599 (mouse) for Val624 (human). Assessment of the stability of the complex by molecular dynamics of about 215 ns confirmed the retention of the key contacts of the ligand with the pocket in all four symmetric copies, with a stationary RMSD level in the range of 1.5–2.5 Å and the maintenance of 1.5–2.0 hydrogen bonds per ligand with optimal geometry.

The potential site of GFB-887 was localized in the transmembrane region S1–S4 of the VSL domain of TRPC4. The data obtained are consistent with the cryo-EM structures of *Danio rerio* TRPC4 in complexes with analogues of the GFB series – GFB-8438, GFB-8749, and GFB-9289 (PDB ID: 7B0S, 7B05, 7B16). As a result of molecular docking, a complex with an energetically favorable conformation of the ligand with an energy of  $-8.11$  kcal/mol and a network of contacts with the key residues Phe413, Tyr373, Tyr646, His369, Glu417, Asp438, Asn442, Arg491, Ser494, Met441, Leu492, and Leu495 was obtained. The

heterocyclic core of GFB-887 forms  $\pi$ - $\pi$  stacking,  $\pi$ - $\pi$  T-shaped,  $\pi$ -alkyl, and  $\pi$ - $\sigma$  interactions with hydrophobic-aromatic residues, while polar and electrostatic contacts ensure the orientation of the molecule in the pocket. Molecular dynamics of the TRPC4:GFB-887 complex of about 200 ns confirmed the stability of the polar binding anchors (His369, Asn442, Arg491, Ser494) and the preservation of the hydrophobic-aromatic environment throughout the entire trajectory. It is important to emphasize that both inhibitors potentially interact in the transmembrane region of the channel but occupy different pockets – Pico145 in the intersubunit cavity, GFB-887 in the intrasubunit site of the VSL domain – which corresponds to two different molecular mechanisms of TRPC4 inhibition.

A method was developed for predicting the parameters of absorption, distribution, metabolism, excretion, and potential toxic effects (ADMET) based on hybrid molecular representations – the concatenation of 512-dimensional embeddings obtained by training a Mol2Vec model on data from the ZINC20 chemical compound database (883,897,271 compounds) with classical descriptors and fingerprints. On the standardized benchmark datasets of the Therapeutics Data Commons (TDC) platform, the model architecture ranks first in 10 of 16 benchmarks. The results were made publicly available on the bioRxiv preprint server. Using this method, the predicted ADMET profile was characterized for the first time for GFB-887 and Pico145, for which the relevant experimental data in open sources are incomplete or absent.

Preliminary analysis indicates that both compounds demonstrate effective intestinal absorption and high plasma protein binding, but differ in their metabolic and toxicological profiles: GFB-887 shows no signs of inhibition of the main CYP isoenzymes, whereas Pico145 is predicted to inhibit CYP3A4, CYP2C9, CYP2D6, and P-glycoprotein and has a high predicted probability of hepatotoxicity (DILI). The DILI and hERG-blockade parameters for both compounds were identified as priorities for further experimental studies.

Experimental validation of the computational hypotheses was performed at two levels of organization: recording of transmembrane ionic currents by the

patch-clamp method in the whole-cell configuration in isolated mouse ileal myocytes (cellular level) and isometric tension recording on segments of mouse small intestine (tissue level). Using the patch-clamp method in the whole-cell configuration in isolated mouse ileal myocytes, the kinetic and voltage-dependent features of the Pico145 blockade of the receptor-operated cationic current  $mI_{CAT}$  were characterized: the inhibition developed slowly with a mean time constant of  $140 \pm 6$  s ( $n = 9$ ) and a pronounced voltage dependence of the action of this ligand at  $-120$  mV the inhibition was less effective compared with  $80$  mV. The combination of these features: high affinity (effective concentration  $100$  pM), slow kinetics, and sensitivity to the transmembrane potential is direct functional evidence that the Pico145 binding site is localized in the plane of the membrane and requires diffusion of the ligand through the lipid environment, which is fully consistent with the computationally predicted intersubunit hydrophobic site. Using the method of isometric tension recording on isolated segments of mouse *ileum*, it was shown for the first time that a 20-minute preincubation with GFB-887 at a concentration of  $10$   $\mu\text{mol/L}$  reduces the carbachol ( $50$   $\mu\text{mol/L}$ )-induced contractile response by  $\approx 40\%$  (to  $\approx 60\%$  of the control level). This is the first functional evidence of the pharmacological activity of GFB-887 in TRPC4-dependent native tissues, which supports the view that TRPC4 is a promising therapeutic target in GI tract motility disorders.

The practical significance of the results obtained lies in the fact that the constructed molecular models of the complexes of TRPC4 with Pico145 and GFB-887 form a structural basis for the rational design of selective modulators of channels of the TRPC1/4/5 subgroup. The developed computational tools, namely the hybrid method for the prioritization of binding sites and the ADMET models, are universal in nature and do not require retraining or adaptation and can be applied to other membrane and soluble protein targets and to the evaluation of the ADMET properties of compounds under study, which ensures their independent scientific and practical value and the possibility of integration into the research of other scientific groups at the early stages of the drug development process.



The main theoretical propositions, methodological approaches, and practical results of the conducted dissertation research were fully reflected in 6 scientific works, including: 1 article in an international peer-reviewed journal indexed in the Scopus and Web of Science scientometric databases (with quartile Q1), and 2 articles in peer-reviewed journals of Ukraine. The results of the dissertation work were presented at international scientific and practical conferences with the publication of 3 conference abstracts.

**Keywords:** TRPC4; smooth muscles; smooth muscle cells; calcium signaling /  $\text{Ca}^{2+}$ -signal; molecular docking; molecular dynamics; binding site; computer modeling / computer simulation; biomechanical parameters of muscle contraction; hydrogen bonding; conformational analysis

### Список публікацій здобувача

#### Праці, в яких опубліковані основні наукові результати дисертації:

1. **Stratiichuk R.**, Melnychenko M., Koleiev I., Voitsitskyi T., Husak V., Shevchuk N., Ostrovsky Z., Bdzhola V., Yesylevsky S., Starosyla S., Nafiiiev A. Leveraging large language models for literature-driven prioritization of protein binding pockets. *Bioinformatics*. 2025. V. 41 (8). P. btaf449. *(особистий внесок здобувача – обробка даних (формування набору білків і пов'язаних публікацій для бенчмарка), формальний аналіз результатів, валідація)*.
2. **Стратійчук Р. О.**, Дринь Д. О., Жолос О. В. Блокування TRPC4-каналів інгібітором Pico145: експериментальне та молекулярне дослідження. *Вісник Київського національного університету імені Тараса Шевченка. Біологія*. 2026. № 1 (104). С. 25–31. *(особистий внесок здобувача – концептуалізація дослідження, методологія експериментів із*

*молекулярного моделювання, проведення молекулярного моделювання, аналіз даних та написання чернетки рукопису).*

3. **Стратійчук Р. О.,** Дринь Д. О., Жолос О. В. Структурне моделювання взаємодії GFB-887 із TRPC4 та оцінка його впливу на карбахол-індуковану скоротливість тонкої кишки миші. Фармакологія та лікарська токсикологія. 2026. Т. 20, № 1. С. 45–58. *(особистий внесок здобувача – інформаційний пошук, концептуалізація дослідження, методологія експериментів із молекулярного моделювання, проведення молекулярного моделювання, аналіз даних, графічна підготовка результатів до опублікування та написання чернетки рукопису).*

### **Опубліковані праці апробаційного характеру**

1. **Stratiichuk R.,** Zholos A. ADMET prediction using molecular embeddings based on Mol2Vec: a comparative study within the TDC. XXII Міжнародна наукова конференція студентів та молодих вчених «Шевченківська весна: ДОСЯГНЕННЯ В НАУКАХ ПРО ЖИТТЯ». Київ, 23–25 квітня 2025.
2. **Stratiichuk R.,** Zholos A. Large language model-driven identification of biologically relevant protein binding pockets to accelerate drug discovery. IX Міжнародна наукова конференція «Психофізіологічні та вісцеральні функції в нормі і патології». Київ, 21–24 жовтня 2025.
3. **Stratiichuk R. O.,** Zholos O. V., Dryn D. O. Analysis of mouse TRPC4 inhibition by Pico145 and GFB-887 using molecular modeling, electrophysiology, and functional physiology methods. Український біофізичний тиждень 2026. Україна, 25–28 березня 2026.

## List of publication of the Ph.D candidate

### Publications containing the main scientific results of the dissertation:

1. **Stratiichuk R.**, Melnychenko M., Koleiev I., Voitsitskyi T., Husak V., Shevchuk N., Ostrovsky Z., Bdzhola V., Yesylevskyy S., Starosyla S., Nafiiev A. Leveraging large language models for literature-driven prioritization of protein binding pockets. *Bioinformatics*. 2025. V. 41 (8). P. btaf449. *(candidate's personal contribution – data processing (assembly of the set of proteins and associated publications for the benchmark), formal analysis of the results, validation)*.
2. **Stratiichuk R. O.**, Dryn D. O., Zholos O. V. Blockade of TRPC4 channels by the Pico145 inhibitor: an experimental and molecular study. *Bulletin of Taras Shevchenko National University of Kyiv. Biology*. 2026. No. 1 (104). P. 25–31. *(candidate's personal contribution – conceptualization of the study, methodology of the molecular modeling experiments, performing the molecular modeling, data analysis, and writing of the manuscript draft)*.
3. **Stratiichuk R. O.**, Dryn D. O., Zholos O. V. Structural modeling of the interaction between GFB-887 and TRPC4 and evaluation of its effect on carbachol-induced contractility of the mouse small intestine. *Pharmacology and Drug Toxicology*. 2026. Vol. 20, No. 1. P. 45–58. *(candidate's personal contribution – literature search, conceptualization of the study, methodology of the molecular modeling experiments, performing the molecular modeling, data analysis, graphical preparation of the results for publication, and writing of the manuscript draft)*.

**Published conference papers:**

1. **Stratiichuk R.**, Zholos A. ADMET prediction using molecular embeddings based on Mol2Vec: a comparative study within the TDC. *XXII International Scientific Conference of Students and Young Scientists «Shevchenko Spring: ADVANCEMENTS IN LIFE SCIENCES»*. Kyiv, April 23–25, 2025.
2. **Stratiichuk R.**, Zholos A. Large language model-driven identification of biologically relevant protein binding pockets to accelerate drug discovery. *IX International Scientific Conference “Psychophysiological and Visceral Functions in Norm and Pathology”*. Kyiv, October 21–24, 2025.
3. **Stratiichuk R. O.**, Zholos O. V., Dryn D. O. Analysis of mouse TRPC4 inhibition by Pico145 and GFB-887 using molecular modeling, electrophysiology, and functional physiology methods. *Biophysical Week of the Ukrainian Biophysical Society 2026*. Ukraine, March 25–28, 2026.

## ЗМІСТ

|                                                                          |           |
|--------------------------------------------------------------------------|-----------|
| <b>ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ.....</b>                                    | <b>20</b> |
| <b>ВСТУП.....</b>                                                        | <b>23</b> |
| <b>РОЗДІЛ 1. ОГЛЯД ЛІТЕРАТУРИ.....</b>                                   | <b>30</b> |
| 1.1. Підродина канонічних каналів транзйєтного рецепторного потенціалу.. | 31        |
| 1.1.1. Суперродина TRP: склад і спільні структурні риси.....             | 31        |
| 1.1.2. Підродина TRPC та її функціональні групи.....                     | 33        |
| 1.2. Структура TRPC4.....                                                | 34        |
| 1.2.1. Доменна організація мономеру TRPC4.....                           | 34        |
| 1.2.2. Тетрамерна архітектура та структурні дані.....                    | 36        |
| 1.2.3. Сайти зв'язування лігандів.....                                   | 41        |
| 1.3. Регуляція активності TRPC4.....                                     | 42        |
| 1.4. Фізіологічне значення TRPC4.....                                    | 44        |
| 1.4.1. Експресія та розподіл у тканинах.....                             | 44        |
| 1.4.2. TRPC4 у шлунково-кишковому тракті.....                            | 45        |
| 1.4.3. TRPC4 у центральній та периферичній нервовій системі.....         | 46        |
| 1.4.4. TRPC4 у серцево-судинній системі, нирках та інших системах.....   | 46        |
| 1.5. Патологічні стани, пов'язані з TRPC4.....                           | 47        |
| 1.5.1. Патологія шлунково-кишкового тракту.....                          | 47        |
| 1.5.2. Інші патологічні стани.....                                       | 48        |
| 1.6. Модулятори TRPC4.....                                               | 49        |
| 1.6.1. Загальна характеристика.....                                      | 49        |
| 1.6.2. Pico145 і GFB-887.....                                            | 51        |
| 1.7. In silico методи у розробці модулаторів TRPC4.....                  | 53        |
| <b>РОЗДІЛ 2. МАТЕРІАЛИ І МЕТОДИ.....</b>                                 | <b>58</b> |
| 2.1. Алгоритм ідентифікації та пріоритезації сайтів зв'язування.....     | 58        |
| 2.2. Молекулярний докінг.....                                            | 61        |
| 2.2.1. Підготовка молекул лігандів.....                                  | 62        |
| 2.2.2. Підготовка рецептора.....                                         | 62        |
| 2.2.3. Аналіз результатів молекулярного докінгу.....                     | 64        |
| 2.3. Моделювання молекулярної динаміки.....                              | 64        |

|                                                                                                                               |           |
|-------------------------------------------------------------------------------------------------------------------------------|-----------|
| 2.3.1. Веб-сервіс для генерації вхідних даних CHARMM-GUI.....                                                                 | 64        |
| 2.3.2. Створення мембрани за допомогою Membrane Builder.....                                                                  | 65        |
| 2.3.3. Створення системи для моделювання молекулярної динаміки.....                                                           | 66        |
| 2.3.4. Силове поле CHARMM36m.....                                                                                             | 68        |
| 2.3.5. Програма для моделювання молекулярної динаміки GROMACS.....                                                            | 68        |
| 2.3.6. Типи файлів у GROMACS.....                                                                                             | 69        |
| 2.3.7. Моделювання молекулярної динаміки у програмі GROMACS.....                                                              | 69        |
| 2.3.8. Аналіз та візуалізація результатів молекулярної динаміки.....                                                          | 71        |
| 2.4. Прогнозування ADMET-параметрів.....                                                                                      | 72        |
| 2.5. Експериментальне дослідження впливу Pico145 та GFB-887 на mICAT і скоротливу активність гладеньких м'язів кишечника..... | 72        |
| 2.5.1. Експериментальні тварини.....                                                                                          | 72        |
| 2.5.2. Матеріали і реактиви.....                                                                                              | 73        |
| 2.5.3. Виділення ізольованих клітин гладеньких м'язів тонкого кишечника...                                                    | 74        |
| 2.5.4. Реєстрація трансмембранних струмів.....                                                                                | 74        |
| 2.5.5. Реєстрація скоротливої активності гладеньких м'язів кишечника.....                                                     | 75        |
| 2.5.6. Обробка та аналіз даних.....                                                                                           | 76        |
| <b>РОЗДІЛ 3. ІДЕНТИФІКАЦІЯ ТА ПРІОРИТИЗАЦІЯ САЙТІВ ЗВ'ЯЗУВАННЯ TRPC4.....</b>                                                 | <b>78</b> |
| 3.1. Загальна схема гібридного алгоритму та формування бенчмарк-датасету...                                                   | 79        |
| 3.2. Геометричне виявлення кандидатних кишень зв'язування.....                                                                | 82        |
| 3.3. Отримання інформації про сайти зв'язування із застосуванням великих мовних моделей.....                                  | 83        |
| 3.3.1. Триетапний конвеєр: фільтрація, екстракція, уточнення.....                                                             | 83        |
| 3.3.2. Оптимізація системи запитів.....                                                                                       | 86        |
| 3.4. Алгоритмічне картування залишків та побудова об'ємного представлення                                                     | 87        |
| 3.4.1. Картування залишків на структури PDB та кластеризація.....                                                             | 87        |
| 3.4.2. Побудова сіткового представлення кишень.....                                                                           | 90        |
| 3.5. Оцінка ефективності розробленого методу.....                                                                             | 93        |
| 3.6. Застосування гібридного методу для аналізу сайтів зв'язування TRPC4.....                                                 |           |

100

|                                                                                                                          |            |
|--------------------------------------------------------------------------------------------------------------------------|------------|
| 3.6.1. Формування корпусу публікацій для BMM-опосередкованої екстракції сайтів зв'язування TRPC4.....                    | 101        |
| 3.6.2. Кишені-кандидати у структурі 5Z96 та їхня кластеризація.....                                                      | 103        |
| 3.6.3. Ідентифікація сайтів зв'язування у структурі TRPC4 миші та їх структурно-функціональна характеристика.....        | 105        |
| <b>РОЗДІЛ 4. МОЛЕКУЛЯРНЕ МОДЕЛЮВАННЯ ВЗАЄМОДІЇ PICO145 ТА GFB-887 З TRPC4.....</b>                                       | <b>120</b> |
| 4.1. Молекулярний докінг Pico145 до TRPC4.....                                                                           | 120        |
| 4.2. Молекулярна динаміка комплексу TRPC4:Pico145.....                                                                   | 125        |
| 4.3. Молекулярний докінг GFB-887 до TRPC4.....                                                                           | 133        |
| 4.4. Молекулярна динаміка комплексу TRPC4:GFB-887.....                                                                   | 138        |
| <b>РОЗДІЛ 5. ОЦІНКА ADMET-ПАРАМЕТРІВ.....</b>                                                                            | <b>144</b> |
| 5.1. Загальна схема методу.....                                                                                          | 145        |
| 5.2. Підготовка корпусу та генерація Mol2Vec-ембедингів.....                                                             | 147        |
| 5.3. Стратегії конструювання ознак та їх відбір.....                                                                     | 148        |
| 5.4. Навчання MLP-моделей та оптимізація гіперпараметрів.....                                                            | 151        |
| 5.5. Порівняння з існуючими підходами на бенчмарках TDC.....                                                             | 152        |
| 5.6. Оцінка ADMET-параметрів для Pico145 та GFB-887.....                                                                 | 157        |
| 5.6.1. ADMET-профіль Pico145.....                                                                                        | 157        |
| 5.6.2. ADMET-профіль GFB-887.....                                                                                        | 161        |
| <b>РОЗДІЛ 6. ЕКСПЕРИМЕНТАЛЬНА ВАЛІДАЦІЯ ВПЛИВУ МОДУЛЯТОРІВ НА СКОРОТЛИВІСТЬ ГЛАДЕНЬКИХ М'ЯЗІВ ТОНКОЇ КИШКИ МИШІ.....</b> | <b>166</b> |
| 6.1. Електрофізіологічне дослідження впливу Pico145 на mICAT.....                                                        | 166        |
| 6.2. Тензометричне дослідження впливу GFB-887 на скоротливість тонкої кишки.....                                         | 168        |
| 6.3. Співставлення результатів in silico та in vitro досліджень.....                                                     | 170        |
| <b>РОЗДІЛ 7. УЗАГАЛЬНЕННЯ РЕЗУЛЬТАТІВ РОБОТИ.....</b>                                                                    | <b>173</b> |
| <b>ВИСНОВКИ.....</b>                                                                                                     | <b>182</b> |
| <b>СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....</b>                                                                                   | <b>185</b> |

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

|         |   |                                                                                                                           |
|---------|---|---------------------------------------------------------------------------------------------------------------------------|
| BMM     | – | великі мовні моделі (large language models)                                                                               |
| Кріо-ЕМ | – | кріоелектронна мікроскопія                                                                                                |
| KX      | – | карбахол (carbachol)                                                                                                      |
| МД      | – | молекулярна динаміка                                                                                                      |
| ТМ      | – | трансмембранний (-i) (transmembrane)                                                                                      |
| ЦНС     | – | центральна нервова система                                                                                                |
| ШКТ     | – | шлунково-кишковий тракт                                                                                                   |
| ADMET   | – | absorption, distribution, metabolism, excretion, toxicity<br>(абсорбція, розподіл, метаболізм, екскреція і токсичність)   |
| ARD     | – | ankyrin repeat domain (анкіриновий повторний домен)                                                                       |
| АТФ     | – | adenosine triphosphate (аденозинтрифосфат)                                                                                |
| ВАРТА   | – | 1,2-bis(o-aminophenoxy)ethane-N,N,N',N'-tetraacetic acid<br>(1,2-біс(о-амінофенокси)етан-N,N,N',N'-тетраоцтова кислота)   |
| BBB     | – | blood–brain barrier (гематоенцефалічний бар'єр)                                                                           |
| CADD    | – | computer-aided drug design (обчислювальний дизайн лікарських засобів)                                                     |
| CaM     | – | calmodulin (кальмодулін)                                                                                                  |
| CH1     | – | connecting helix (сполучна спіраль, С-кінцевий домен)                                                                     |
| CH2     | – | coiled-coil domain (суперспіралізований С-кінцевий домен)                                                                 |
| CIRB    | – | calmodulin- and IP <sub>3</sub> -receptor binding region (ділянка зв'язування кальмодуліну та IP <sub>3</sub> -рецептора) |
| CoT     | – | chain-of-thought (ланцюжок міркувань)                                                                                     |
| CYP     | – | cytochrome P450 (цитохром P450)                                                                                           |
| DAG     | – | diacylglycerol (діацилгліцерол)                                                                                           |
| DILI    | – | drug-induced liver injury (лікарсько-індуковане ушкодження печінки)                                                       |
| DL      | – | deep learning (глибоке навчання)                                                                                          |



|                   |   |                                                                                                                              |
|-------------------|---|------------------------------------------------------------------------------------------------------------------------------|
| FSGS              | – | focal segmental glomerulosclerosis (фокально-сегментарний гломерулосклероз)                                                  |
| GABA              | – | gamma-aminobutyric acid (γ-аміномасляна кислота)                                                                             |
| GNN               | – | graph neural networks (графові нейронні мережі)                                                                              |
| GPCR              | – | G-protein-coupled receptors (рецептори, зв'язані з G-білками)                                                                |
| hERG              | – | human ether-à-go-go-related gene (калієвий канал hERG)                                                                       |
| HIA               | – | human intestinal absorption (людська кишкова абсорбція)                                                                      |
| HLH               | – | helix–loop–helix (спіраль-поворот-спіраль, мотив)                                                                            |
| IP <sub>3</sub>   | – | inositol 1,4,5-trisphosphate (інозитол-1,4,5-трисфосфат)                                                                     |
| MACCS             | – | molecular access system keys (структурні ключі MACCS)                                                                        |
| mI <sub>CAT</sub> | – | muscarinic cation current (мускариновий катіонний струм)                                                                     |
| ML                | – | machine learning (машинне навчання)                                                                                          |
| MLKL              | – | mixed lineage kinase domain-like protein (білок, подібний до домену кіназ змішаної лінії)                                    |
| MLP               | – | multilayer perceptron (багатошаровий перцептрон)                                                                             |
| NHERF             | – | Na <sup>+</sup> /H <sup>+</sup> exchanger regulatory factor (регуляторний фактор Na <sup>+</sup> /H <sup>+</sup> -обмінника) |
| PA                | – | phosphatidic acid (фосфатидна кислота)                                                                                       |
| PDB               | – | Protein Data Bank (банк даних просторових структур білків)                                                                   |
| PDZ               | – | PSD-95/DlgA/ZO-1 homology domain (білок-білковий домен PDZ)                                                                  |
| P-gp              | – | P-glycoprotein (P-глікопротеїн)                                                                                              |
| PIP <sub>2</sub>  | – | phosphatidylinositol 4,5-bisphosphate<br>(фосфатидилінозитол-4,5-бісфосфат)                                                  |
| PLC               | – | phospholipase C (фосфоліпаза C)                                                                                              |
| POPC              | – | 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine<br>(1-пальмітоїл-2-олеоїл-sn-гліцери-3-фосфохолін)                          |
| PPB               | – | plasma protein binding (зв'язування з білками плазми)                                                                        |
| PPI               | – | protein–protein interactions (білок-білкові взаємодії)                                                                       |

|                  |   |                                                                                                            |
|------------------|---|------------------------------------------------------------------------------------------------------------|
| RMSD             | – | root mean square deviation (кореневе середньоквадратичне відхилення)                                       |
| TDC              | – | Therapeutics Data Commons (база даних для розробки терапевтичних засобів)                                  |
| ToT              | – | tree-of-thought (дерево міркувань)                                                                         |
| TRP              | – | transient receptor potential (канали транз'єнтного рецепторного потенціалу)                                |
| TRPA             | – | transient receptor potential ankyrin (анкіринові TRP-канали)                                               |
| TRPC             | – | transient receptor potential canonical (канонічна підродина каналів транз'єнтного рецепторного потенціалу) |
| TRPM             | – | transient receptor potential melastatin (меластатинові TRP-канали)                                         |
| TRPML            | – | transient receptor potential mucolipin (муколіпінові TRP-канали)                                           |
| TRPP             | – | transient receptor potential polycystin (поліцистинові TRP-канали)                                         |
| TRPV             | – | transient receptor potential vanilloid (ванілоїдні TRP-канали)                                             |
| VD <sub>ss</sub> | – | volume of distribution at steady state (об'єм розподілу у стаціонарному стані)                             |
| VSL              | – | voltage-sensor-like domain (домен, подібний до сенсора напруги)                                            |

## ВСТУП

**Актуальність теми.** Канали транз'єнтного рецепторного потенціалу підродини канонічних (TRPC), насамперед TRPC4 і структурно найбільш подібний до нього TRPC5 є рецептор-керованими катіонними каналами, що відіграють ключову роль у кальцієвій сигналізації, регуляції судинного тону, функціонуванні серцево-судинної і нервової систем [1–4]. Порухення функції TRPC4 асоційовано з розладами моторики кишечника, зокрема з холінергічним збудженням і скороченням гладеньких м'язів шлунково-кишкового тракту (ШКТ), а також з порушенням регуляції ендотелій-залежного розслаблення гладеньких м'язів судин, гіпертрофією серця, розвитком онкологічних процесів та неврологічними розладами – посттравматичним стресовим розладом, тривожними станами, страхом і порушеннями соціальних взаємодій [3, 5–9].

Попри фармакологічну значущість TRPC4 як молекулярної мішені для лікування різних захворювань, наявність селективних лігандів цього каналу на сьогодні обмежена, а молекулярні механізми їхньої дії вивчено недостатньо. Pico145 – потужний інгібітор підгрупи TRPC1/4/5 з активністю у пікомолярному діапазоні; дані кріоелектронної мікроскопії (кріо-ЕМ) для комплексу TRPC5 людини з Pico145 підтвердили існування гідрофобного сайту зв'язування TRPC5, а подальші дослідження підтвердили, що сайт є консервативним для представників підродини TRPC1/4/5 [10, 11]. Оскільки структурні дані щодо зв'язування Pico145 з TRPC4 миші дотепер відсутні, оцінка таких взаємодій стала одним з завдань даного дослідження. GFB-887 – інгібітор TRPC5, що перебуває у II фазі клінічних випробувань для лікування хронічних захворювань нирок, зокрема фокально-сегментарного гломерулосклерозу (FSGS) [12]. Канали TRPC4 і TRPC5 є близькоспорідненими паралогами в межах підродини TRPC1/4/5 і поділяють приблизно 77% ідентичності амінокислотної послідовності [13]. Зважаючи на цю високу гомологію, а також на відсутність опублікованих структурних

даних щодо взаємодії GFB-887 з TRPC-каналами, дослідження потенційної активності цієї сполуки відносно TRPC4 є методологічно обґрунтованим. Таким чином, визначення особливостей молекулярних взаємодій хімічних сполук Pico145 і GFB-887 з TRPC4 є актуальною задачею, розв'язання якої створює структурно-функціональну основу для раціонального дизайну селективних модуляторів цього каналу та інших білків цієї родини.

**Мета і завдання.** Метою дисертаційного дослідження було встановлення особливостей молекулярних взаємодій TRPC4 з малими молекулами-інгібіторами Pico145 та GFB-887 шляхом поєднання методів молекулярного моделювання, машинного навчання, електрофізіології і тензометрії для функціональної оцінки цих взаємодій.

Відповідно до зазначеної мети було поставлено наступні завдання:

1. Проаналізувати сайти зв'язування TRPC4 з малими молекулами та ідентифікувати потенційні сайти зв'язування з інгібіторами Pico145 і GFB-887 з використанням комбінованого підходу, що поєднує геометричне прогнозування порожнин та літературні дані з метою їх пріоритизації.
2. Виконати молекулярний докінг хімічних сполук Pico145 та GFB-887 в структуру TRPC4 миші та оцінити стабільність отриманих комплексів методом молекулярної динаміки.
3. Оцінити ADMET-властивості Pico145 та GFB-887 за допомогою методів машинного навчання для оцінки його фармакологічного потенціалу.
4. Оцінити вплив Pico145 на індукований активований мускариновими рецепторами катіонний струм ( $mI_{CAT}$ ) у гладеньком'язових клітинах клубової кишки миші.
5. Охарактеризувати інгібуючий вплив GFB-887 на карбахол-індуковану скоротливість в тканинній моделі тонкої кишки миші *in vitro*.

**Об'єкт дослідження:** ліганди рецептор-керованого катіонного каналу TRPC4 та TRPC4-залежні іонні струми.

**Предмет дослідження:** молекулярні механізми зв'язування лігандів із TRPC4 та їх фармакологічні ефекти.

**Методи дослідження:** ідентифікація сайтів зв'язування з використанням обчислювального комбінованого підходу, молекулярний докінг, молекулярна динаміка, прогнозування ADMET-властивостей методами машинного навчання, петч-клемп, ізометрична тензометрія.

**Наукова новизна отриманих результатів.** У даній роботі вперше отримано наведені нижче результати:

1. Вперше побудовано молекулярні моделі комплексів мишачого TRPC4 з інгібіторами Pico145 і GFB-887; встановлено, що Pico145 зв'язується у консервативній міжсубодиничній гідрофобній кишені, що є характерною для підгрупи TRPC1/4/5 каналів родини TRPC, а GFB-887 – у внутрішній субодиничній кишені, яка розміщена в трансмембранній ділянці S1–S4 VSL-домену.
2. Охарактеризовано кінетичні та потенціалзалежні особливості блокади Pico145 рецептор-керованого катіонного струму  $mI_{CAT}$  в ізольованих міоцитах клубової кишки миші.
3. Експериментально показано, що інгібітор TRPC5 GFB-887 (10 мкмоль/л, 20-хвилинна інкубація) пригнічує карбахол-індуковану скоротливу відповідь сегментів тонкої кишки миші на  $\approx 40\%$  (до  $\approx 60\%$  контрольного рівня), що є функціональним свідченням фармакологічної активності GFB-887 у TRPC4-залежних нативних тканинах.
4. У зв'язку з обмеженням доступних підходів та методів, розроблено гібридний метод ідентифікації та пріоритизації сайтів зв'язування, який поєднує геометричний аналіз порожнин білка з автоматизованим отриманням експериментально підтверджених амінокислотних залишків із наукової літератури за допомогою великих мовних моделей (BMM).
5. Розроблено метод прогнозування параметрів абсорбції, розподілу, метаболізму, екскреції та токсичності (ADMET) для малих молекул, що посів перше місце у 10 із 16 бенчмарків платформи Therapeutics Data Commons. Із застосуванням розробленого методу вперше охарактеризовано прогнозний ADMET-профіль модуляторів TRPC4, у тому числі сполуки

GFB-887, для якої у відкритих джерелах відсутні відповідні експериментальні дані щодо ADMET, що створює підґрунтя для подальших цілеспрямованих експериментальних досліджень.

**Практичне значення отриманих результатів.** Результати дисертаційного дослідження мають комплексне прикладне значення для медичної хімії, фармакології мембранних білків та обчислювального етапу розробки лікарських засобів.

Побудовані молекулярні моделі комплексів TRPC4 з Pico145 і GFB-887 формують структурне підґрунтя для раціонального дизайну селективних модуляторів каналів підгрупи рецептор-керованих TRPC1/4/5 каналів. Локалізовані сайти зв'язування та ідентифіковані залишки, критичні для контакту з лігандами, можуть бути безпосередньо використані для побудови фармакофорних моделей, дизайну сполук другого покоління, а також для оптимізації фізико-хімічного та фармакокінетичного профілю кандидатних молекул. Структурно-функціональна характеристика TRPC4 у тонкій кишці миші формує доказову базу для подальшого розгляду каналу як перспективної терапевтичної мішені при порушеннях моторики шлунково-кишкового тракту, зокрема при синдромі подразненого кишечника та функціональній диспепсії.

Розроблений гібридний метод ідентифікації сайтів зв'язування реалізовано у формі рецензованого програмного інструменту з відкритим вихідним кодом, що супроводжується технічною документацією, прикладами застосування та куратованим бенчмарковим набором даних. Результати дослідження опубліковано у міжнародному фаховому виданні *Bioinformatics* (Oxford University Press) [14]. Вихідний код, набір даних та супровідні матеріали розміщено у публічному репозиторії GitHub та депоновано з присвоєнням DOI у відкритому архіві Zenodo [15, 16]. Метод не потребує перенавчання чи адаптації під конкретну мішень і застосовний до дослідження сайтів зв'язування будь-якого білка, для якого доступні просторова структура та достатня літературна база. Зазначені властивості

зумовлюють його придатність до безпосереднього використання академічними та індустріальними дослідницькими групами на ранніх етапах процесу відкриття лікарських засобів.

Робота щодо розробки ADMET моделей оприлюднена на сервері препринтів *bioRxiv* разом із описом архітектури, навчального протоколу та повного переліку молекулярних ознак, що забезпечує відтворюваність запропонованого підходу. Результати незалежної верифікації на бенчмарках платформи Therapeutics Data Commons демонструють конкурентоспроможність моделей та їх потенційну застосовність у прикладних задачах пріоритизації віртуальних бібліотек, фільтрації молекул за фармакокінетичним профілем та оцінки drug-likeness сполук-кандидатів [17]. Оскільки відповідна публікація наразі представлена у форматі препринту, отримані результати доцільно розглядати як основу для подальшої роботи у цьому напрямі.

Комплексний підхід даного дослідження, що поєднує ідентифікацію сайтів зв'язування, молекулярний докінг, молекулярну динаміку комплексів білок-ліганд, прогнозування ADMET-параметрів та функціональну валідацію на ізольованих нативних тканинах, апробований у цій роботі в контексті дослідження дії модуляторів активності TRPC4, формує методологічний шаблон структурно-орієнтованого пошуку модуляторів мембранних білків. Запропонований протокол може бути адаптований для інших представників родини TRP-каналів та інших родин іонних каналів. Обидва розроблені програмні інструменти мають універсальний характер і покривають два важливі етапи в пошуку лікарських засобів – ідентифікація сайтів зв'язування в білкових мішенях та оцінку ADMET-параметрів малих молекул. Це забезпечує їх самостійну наукову та практичну цінність, а також можливість інтеграції у дослідження інших наукових груп.

**Особистий внесок здобувача.** Дисертантом самостійно проведено інформаційний пошук та аналіз сучасного стану проблеми фармакологічної модуляції каналу TRPC4 і споріднених каналів групи TRPC1/4/5. Особисто

виконано обґрунтоване визначення потенційних сайтів зв'язування у структурі TRPC4 із застосуванням власне розробленого гібридного підходу, що комбінує геометричне прогнозування порожнин із літературно зумовленою пріоритезацією амінокислотних залишків на основі аналізу наукових публікацій за допомогою ВММ. Самостійно здійснено підготовку білкових структур, розробку та валідацію докінг-протоколу, та молекулярний докінг низькомолекулярних інгібіторів Pico145 і GFB-887 у кріо-ЕМ структуру мишачого TRPC4. Особисто проведено *in silico* прогнозування ADMET-властивостей досліджуваних сполук за допомогою моделей машинного навчання. Здобувачем самостійно виконано структурний аналіз отриманих комплексів TRPC4:ліганд, ідентифікацію ключових міжмолекулярних взаємодій та амінокислотних залишків, що визначають зв'язування інгібіторів, а також аналіз і порівняння обчислювальних результатів із наявними експериментальними даними кріо-ЕМ і функціональних досліджень.

Консультаційну допомогу та супровід у проведенні досліджень надавали: проф., д. б. н. О. В. Жолос (Київський національний університет імені Тараса Шевченка) – формування загальної концепції роботи, постановка наукових завдань, обговорення отриманих результатів та формулювання висновків; н. с., к. б. н. Д. О. Дринь (Інститут фізіології ім. О. О. Богомольця НАН України) – експериментальні електрофізіологічні дослідження методом петч-клемп та тензометрична реєстрація скоротливої активності тонкої кишки миші; к.б.н. С. А. Старосила (Receptor.AI Inc., Лондон, Сполучене Королівство) – консультативна та методична підтримка щодо застосування обчислювальних підходів у розробці лікарських засобів та інтерпретації результатів *in silico* досліджень. Низькомолекулярні інгібітори Pico145 та GFB-887 були люб'язно надані колегами з Університету Лідса (Сполучене Королівство).

**Апробація результатів дисертації.** Основні результати дисертаційного дослідження доповідалися на XXII Міжнародній науковій конференції



студентів та молодих вчених «Шевченківська весна: ДОСЯГНЕННЯ В НАУКАХ ПРО ЖИТТЯ» (23–25 квітня 2025, Київ); IX Міжнародній науковій конференції «Психофізіологічні та вісцеральні функції в нормі і патології» (21–24 жовтня 2025, Київ); Біофізичному тижні 2026, організованому українським біофізичним товариством (25–28 березня 2026, Україна).

**Зв'язок роботи з науковими програмами, планами, темами, грантами.** Дисертаційна робота виконана в рамках тематики наукових досліджень ННЦ «Інститут біології та медицини» Київського національного університету імені Тараса Шевченка, зокрема по кафедральній науково-дослідній роботі кафедри біофізики та нейробиології “Біофізичні ефекти і механізми дії електромагнітних полів, наноматеріалів та фармакологічних речовин на біологічні об’єкти” (січень 2023 р. - грудень 2025 р., науковий керівник О.В. Жолос).

Тема дисертаційної роботи затверджена на засіданні Вченої ради ННЦ «Інститут біології та медицини» Київського національного університету імені Тараса Шевченка, протокол №4 від 15 листопада 2022 року.

**Публікації.** За темою дисертаційного дослідження опубліковано 6 наукових праць, серед них 3 статті у наукових періодичних виданнях (з них 1 – у періодичному науковому виданні, проіндексованому в базах даних Web of Science Core Collection та Scopus (Q1), та 2 – у фахових виданнях, затверджених переліком МОН України), а також 3 тези доповідей у матеріалах наукових конференцій.

**Структура та обсяг дисертації.** Дисертаційна робота складається зі вступу, огляду літератури, матеріалів і методів досліджень, результатів дослідження, які викладено в 4 розділах, узагальнення, висновків, списку використаних літературних джерел (141 посилання). Дисертаційна робота викладена на 202 сторінках (з яких 162 сторінки основної частини) містить 31 рисунок та 11 таблиць.

## РОЗДІЛ 1

### ОГЛЯД ЛІТЕРАТУРИ

Канали транз'єнтного рецепторного потенціалу (TRP) є надродиною неселективних катіонних каналів, що функціонують як інтегратори сенсорних сигналів та забезпечують клітинну відповідь на широкий спектр хімічних, фізичних і біологічних стимулів – від змін температури, механічного навантаження та осмотичного тиску до дії ліпідних медіаторів, гормонів і ксенобіотиків [18]. Серед 6 підродин ссавців, TRPC канали посідають особливе місце як основна ланка рецептор-керованих катіонних струмів у збудливих і незбудливих клітинах. У межах підродини TRPC підгрупа TRPC1/4/5 характеризується високим ступенем структурної гомології та здатністю до гетеромеризації між собою. Канал TRPC4, що є домінуючим членом цієї підгрупи, опосередковує мускариновий катіонний струм ( $mI_{CAT}$ ), ключовий механізм холінергічної регуляції скоротливої активності гладеньких м'язів ШКТ [2, 3, 19, 20]. Поза ШКТ канал TRPC4 також залучений до низки сигнальних процесів у нейрональних, ендотеліальних та подоцитарних клітинах, що зумовлює інтерес до нього як до перспективної фармакологічної мішені [1, 7, 21–24].

Попри те, що особливості молекулярної будови TRPC4 з'ясовані за допомогою кріо-ЕМ, сайти зв'язування з фізіологічно релевантними модуляторами, такими як внутрішньоклітинний  $Ca^{2+}$ , фосфатидилінозитол-4,5-бісфосфат ( $PIP_2$ ), кальмодулін та гетеротримерні G-білки ( $G\alpha_{q/11}$  і  $G\alpha_{i/o}$ ), структурно охарактеризовані лише частково [25–27]. Значна частина відомих селективних модуляторів TRPC4/5, наприклад інгібітори Pico145 і HC-070, ML204 та GFB-887 виявлені у ході високопродуктивного скринінгу (HTS) [28–31]. У ході кріо-ЕМ досліджень було ідентифіковано дві ксенобіотик-зв'язувальні кишені у каналі, міжсубодиничну гідрофобну порожнину для ксантинів та кишеню у домені, що подібний до сенсора напруги (VSL) для піридазинових аналогів серії

GFB [11, 27, 32]. Однак раціональний дизайн нових модуляторів TRPC4 залишається обмеженим у зв'язку з відсутністю систематичного картування всіх функціонально релевантних зв'язувальних кишень каналу та браком експериментальних структурних даних щодо особливостей зв'язування хімічних сполук з ними.

Цю прогалину поглиблює відсутність автоматизованих інструментів пріоритезації сайтів зв'язування на основі інтеграції геометричних прогнозів з експериментальними літературними даними [14]. Методи обчислювального дизайну лікарських засобів (CADD) – алгоритмічна ідентифікація зв'язувальних кишень, молекулярний докінг, молекулярна динаміка (МД) та прогнозування ADMET-параметрів забезпечують інструментальну базу для цілеспрямованого пошуку TRPC4-селективних модуляторів [33, 34].

У цьому розділі систематизовано наявні у світовій літературі дані щодо структури TRPC4, механізмів його регуляції, фізіологічного значення та ролі у розвитку патологічних процесів, охарактеризовано сучасний фармакологічний репертуар модуляторів TRPC підродини і проаналізовано *in silico* методи, релевантні для задач, розв'язаних у цій дисертації.

## **1.1. Підродина канонічних каналів транз'єнтного рецепторного потенціалу**

### **1.1.1. Суперродина TRP: склад і спільні структурні риси**

Суперродина TRP у ссавців включає 28 представників, згрупованих на підставі амінокислотної подібності у 6 підродин: TRPC, ванілоїдні TRP-канали (TRPV), меластатинові TRP-канали (TRPM), анкіринові TRP-канали (TRPA), муколіпінові TRP-канали (TRPML), поліцистинові TRP-канали (TRPP) [1, 18]. Більшість каналів локалізована у плазматичній мембрані, проте представники підродини TRPML переважно експресуються

на мембранах ендосом і лізосом, де функціонують як внутрішньоорганельні  $\text{Ca}^{2+}$ -канали [18]. Окремі представники також локалізуються на мембранах ендо(сарко)плазматичного ретикулу, наприклад TRPM8 [18].

Усі TRP-канали є тетрамерними інтегральними мембранними білками. Кожна субодинаця містить шість трансмембранних (TM) сегментів (S1–S6), за топологією подібних до субодинаць потенціал-залежних калієвих каналів. Іон-провідна пора формується спіралями S5 і S6 чотирьох субодинаць з поровою петлею між ними, що утворює розташований на позаклітинному боці селективний фільтр. Сегменти S1–S4 формують VSL-домен, хоча у переважної більшості TRP-каналів він не має типової потенціал-чутливої функції. Внутрішньоклітинні N- та С-кінцеві домени різняться за довжиною і складом між підродинами: TRPC, TRPV і TRPA містять анкіринові повтори (ARD), тоді як TRPM – меластатинові гомологічні домени; TRPP і TRPML характеризуються подовженими позаклітинними петлями [18].

Спільною функціональною рисою більшості TRP-каналів є неселективна катіонна провідність, що забезпечує проходження іонів натрію ( $\text{Na}^+$ ), калію ( $\text{K}^+$ ) і кальцію ( $\text{Ca}^{2+}$ ). Механізми активації різноманітні: TRP-канали реагують на температуру, механічний тиск, осмолярність, зміни рН, наявність ліпідних вторинних посередників, зв'язування лігандів-малих молекул або пряму модуляцію G-білками [1, 18, 35]. Така функціональна пластичність зумовлює залученість TRP-каналів у процеси сенсорного сприйняття, запалення, больової чутливості, регуляції судинного тону, скоротливості гладеньких м'язів, розвитку і проліферації клітин, а також у патофізіологічні процеси – від нейродегенеративних захворювань до онкологічних трансформацій [1, 36].

### 1.1.2. Підродина TRPC та її функціональні групи

Підродина TRPC у ссавців представлена сімома генами (*TRPC1–TRPC7*). За структурною подібністю та функціональними характеристиками виділяють чотири підгрупи: TRPC1, TRPC2, TRPC4/5 та TRPC3/6/7 [1, 18]. TRPC2 у людей не експресується, оскільки відповідний ген є псевдогеном. Представники інших підгруп експресуються у широкому спектрі тканин: головному мозку, ендотелії, нирках, гладеньких м'язях, серці, наднирниках, плаценті та сітківці ока [18].

TRPC-канали активуються сигналами, що виникають через активацію фосфоліпази C (PLC) і ініціюються стимуляцією рецепторів, зв'язаних із G-білками (GPCR), або рецепторних тирозинкіназ. Класична парадигма активації передбачає залучення ізоформ PLC $\beta$  (через G<sub>q/11</sub>) або PLC $\gamma$  (через тирозинкінази), гідроліз PIP<sub>2</sub> з утворенням діацилгліцеролу (DAG) та інозитол-1,4,5-трисфосфату (IP<sub>3</sub>), підвищення концентрації цитозольного Ca<sup>2+</sup> через IP<sub>3</sub>-рецептори ендоплазматичного ретикулуму, а для деяких представників – пряму активацію DAG [1, 37, 38]. Крім того, для TRPC4/5 продемонстровано пряму модуляцію G $\alpha_{i/o}$ -субодиницями, що опосередковує інтеграцію сигналів від G<sub>q/11</sub>- і G<sub>i</sub>-шляхів у збудливих клітинах [39].

Канали підгрупи TRPC4/5 (TRPC4 і TRPC5) демонструють найвищий ступінь амінокислотної подібності у межах підродини (~77% ідентичності амінокислотної послідовності) та здатні формувати як гомотетрамери, так і гетеротетрамери між собою та з TRPC1 [13, 18]. TRPC1 поодиноці у клітинах ссавців не формує або слабо формує функціональні канали, виконуючи роль допоміжної субодиниці, що змінює біофізичні властивості TRPC4- та TRPC5-вмісних гетеромерів – зокрема, знижує Ca<sup>2+</sup>-провідність, модифікує форму вольт-амперної характеристики та змінює чутливість до фармакологічних модуляторів [1, 40]. TRPC4 і TRPC5 – неселективні катіонні канали з помірною Ca<sup>2+</sup>-провідністю і характерною активацією через Gq/11-PLC каскад із наступною чутливістю до іонів Ca<sup>2+</sup>; біофізичні

властивості, одноканальна провідність і воротний механізм цього мускаринового катіонного струму детально охарактеризовані у низці електрофізіологічних досліджень [1, 2, 41–44].

На відміну від TRPC4/5, канали підгрупи TRPC3/6/7 виявляють виражену DAG-чутливість і беруть участь передусім у регуляції судинного тону, серцевої функції та ниркової фізіології [1, 18]. TRPC1 часто опосередковує депо-кероване надходження  $\text{Ca}^{2+}$  у партнерстві зі STIM1 і Orai1, хоча ця його функція залишається дискусійною [30]. Таким чином, підгрупа TRPC4/5 виокремлюється з-поміж інших TRPC як спеціалізована ланка рецептор-керованих катіонних струмів у різних клітинах, включно з вісцеральними гладеньком'язовими клітинами ШКТ [2, 3, 19].

## **1.2. Структура TRPC4**

Хронологічно структурна характеристика TRPC4 відбулася у три послідовні етапи: отримання апо-структур (2018), структур у комплексах з низькомолекулярними модуляторами (2019–2020) та структур гетеромерних каналів і додаткових гомомерних конформацій (2025). Усі опубліковані на сьогодні кріо-ЕМ-моделі TRPC4 отримані з вирізаними цитозольно-дистальними ділянками С-кінцевого домену, які структурно не розв'язуються через властиву їм конформаційну гетерогенність, типову для внутрішньо неупорядкованих ділянок білків [10, 26, 27].

### **1.2.1. Доменна організація мономеру TRPC4**

Повнорозмірний мономер TRPC4 миші складається з 974 амінокислотних залишків і об'єднує два великі структурні модулі:

трансмембранний і цитозольний [25, 26]. Схематична доменна організація мономеру представлена на Рис. 1.1.

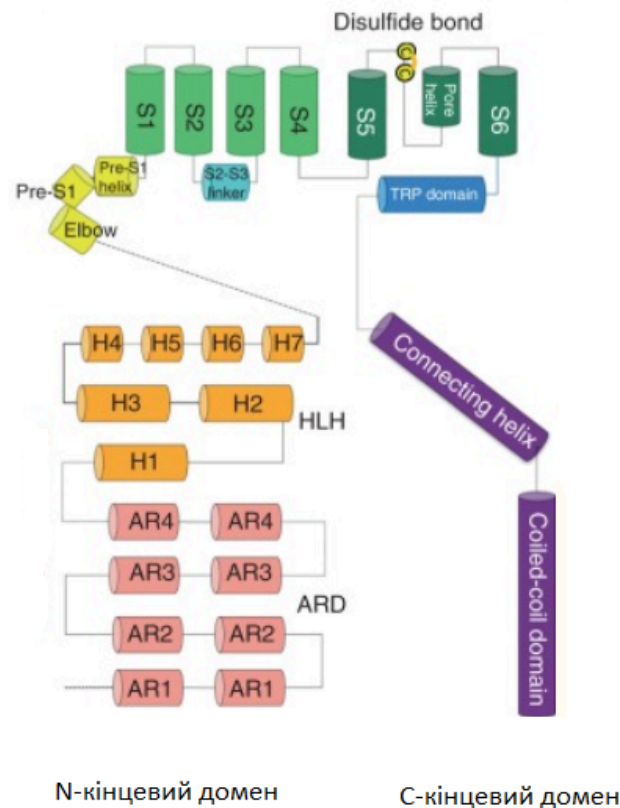


Рис. 1.1. Доменна організація мономеру TRPC4 миші [25].

ТМ-домен включає шість трансмембранних сегментів (S1–S6) і порову спіраль, яка разом з петлею між S5 і S6 формує селективний фільтр на позаклітинному боці мембрани [18]. Сегменти S1–S4 утворюють VSL-домен, організація якого гомологічна сенсору напруги потенціал-залежних катіонних каналів, хоча у TRPC4 він не виконує класичної функції гейтингу за рахунок потенціалу, проте натомість формує кишеню зв'язування для лігандів [26]. Сегменти S5 і S6 разом з поровою спіраллю формують центральний іон-провідний шлях. Між N-кінцем і S1 розташовано коротку ділянку «pre-S1 elbow» – два спіральні вигини, що сполучають цитозольний HLH-модуль (helix-loop-helix) із ТМ-доменом. Між S6 і C-кінцевим CH1 локалізовано TRP-домен – консервативний амфipатичний мотив, що пролягає паралельно до площини мембрани та бере участь як у стабілізації тетрамерної збірки, так

і у передаванні сигналу від периферичних регуляторних доменів до пори [1, 25].

Цитозольний N-кінцевий домен має форму перевернутого дзвону і складається з чотирьох ARD (AR1–AR4) та наступного за ним HLH-мотиву із семи спіралей (H1–H7) [25]. ARD і HLH забезпечують взаємодії з регуляторними білками –  $G\alpha_{i/o}$ ,  $IP_3$ -рецепторами та кальмодуліном (calmodulin, CaM) [45]. С-кінцевий домен складається зі сполучної спіралі CH1 і суперспіралізованого (coiled-coil) домену CH2. CH1 містить консервативний для всіх представників TRPC мотив CIRB (calmodulin- and  $IP_3R$ -binding site, сайт зв'язування кальмодуліну та  $IP_3$ -рецептора), локалізований одразу після TRP-домену і залучений у  $Ca^{2+}$ -залежну регуляцію активності каналу; спільні сайти зв'язування кальмодуліну та  $IP_3$ -рецептора на С-кінцях TRPC-каналів були ідентифіковані одними з перших [46]. CH2 містить три гепта-повтори з характерною періодичністю (a-b-c-d-e-f-g)<sub>n</sub>, у позиціях «a» та «d» яких розміщуються гідрофобні амінокислотні залишки (переважно Val, Ile, Leu, Phe), що забезпечують міжсубодиничні контакти у тетрамері [25].

У нативному повнорозмірному каналі С-кінцевий домен продовжується нерозв'язаною у кріо-ЕМ ділянкою (залишки приблизно 756–974), яка містить два додаткові  $G\alpha$ -зв'язувальні сайти, другий CaM-зв'язувальний домен, потенційний сайт зв'язування  $PIP_2$  і PDZ-зв'язувальний мотив Val-Thr-Thr-Arg-Leu у позиціях 969–974, що опосередковує взаємодію з  $Na^+/H^+$ -обмінним регуляторним фактором ( $Na^+/H^+$  exchanger regulatory factor, NHERF) і актиновим цитоскелетом [38, 47].

### 1.2.2. Тетрамерна архітектура та структурні дані

TRPC4 функціонує як тетрамер – гомотетрамер або гетеротетрамер з TRPC1 або TRPC5 [40]. Гомотетрамер TRPC4 має форму квадратної призми з



поперечними розмірами приблизно  $100 \times 100 \text{ \AA}$  і висотою близько  $120 \text{ \AA}$  уздовж осі, перпендикулярної до площини мембрани (Рис. 1.2.) [25].

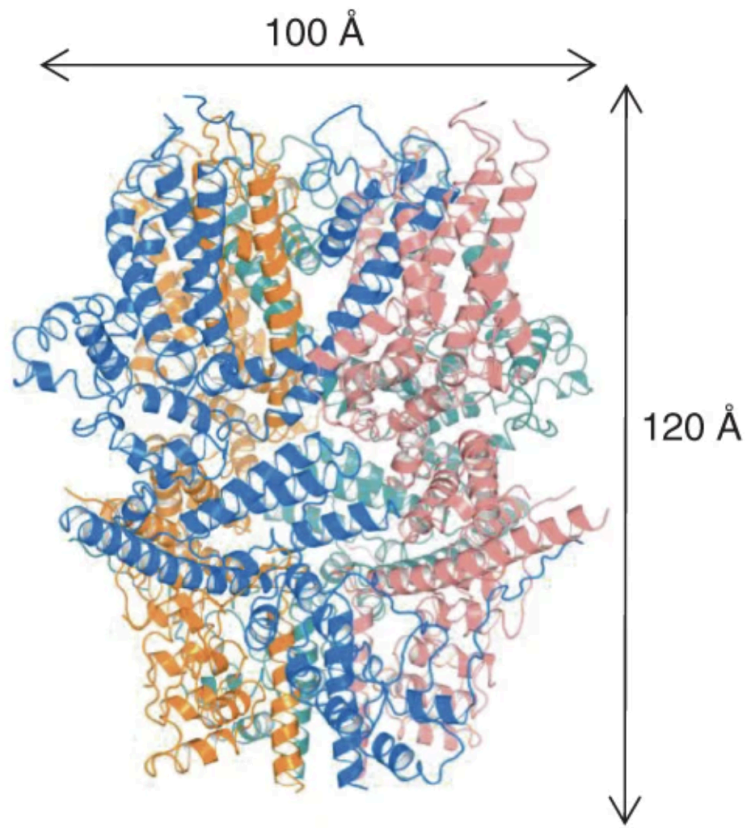


Рис. 1.2. Тетрамерна архітектура TRPC4 миші у мембрані (субодиниці показані різними кольорами) [25].

Чотири субодиниці організовані з симетрією  $C_4$ : ТМ-домени формують центральну мембранну частину з іон-провідним шляхом, а цитозольні домени виступають у цитозоль компактим «дзвоноподібним» блоком, що закриває нижні ворота. Сполучна спіраль СН1 кожного мономера входить у порожнину між двома сусідніми субодиницями і з'єднує їх, а чотири СН2 суперспіралізовані у вертикальну осьову структуру, що додатково ущільнює тетрамер на цитозольному боці [25]. Взаємодія N- і С-кінцевих доменів у межах мономера посилюється  $\pi$ - $\pi$  контактом Tyr271–Tyr706 і водневими зв'язками Glu264–Arg716 та Arg272–Arg702 [25].

Базовим для подальшого обчислювального моделювання у цій дисертації є запис в Protein Data Bank (PDB) з ідентифікатором 5Z96 –

кріо-ЕМ структура повнорозмірного TRPC4 миші в апо-стані з роздільністю 3,3 Å [25]. Структура депонована у Electron Microscopy Data Bank під номером EMD-6901. У скорочений конструктор увійшли амінокислотні залишки 15–755, що охоплюють усю ТМ-частину, N-кінцевий домен і проксимальну частину С-кінцевого домену (TRP-мотив, CIRB, CH1); дистальна частина С-кінця (756–974), яка містить другий CaM-сайт, частину Gα-інтерфейсів, потенційний PIP<sub>2</sub>-сайт і PDZ-мотив, у структурі не представлені. Структурно TRPC4 у 5Z96 представлений гомотетрамером у закритій конформації: селективний фільтр є відкритим, проте іон-провідний шлях повністю обмежений у цитоплазматичній ділянці залишками Ile617, Asn621 і Gln625 на С-кінці S6, що формують нижні ворота з найвужчою точкою на бічному ланцюгу Asn621 [25]. Консервативність Gln625 серед усіх представників підродино TRPC вказує на загальну роль цього залишку у формуванні іонної проникності. Селективний фільтр з найвужчою точкою 6,7 Å сформовано бічними ланцюгами Gly577; позаклітинна петля стабілізована дисульфідним зв'язком Cys549–Cys554 у вигині між S5 і поровою спіраллю [25]. Функціональне значення має залишок Gly503 у лінкері S4–S5: мутації у цій позиції переводять канал у конформацію постійно відкритого стану, що підкреслює роль лінкера S4–S5 у передаванні сигналу від VSL-домену до нижніх воріт [1, 25].

У 2018 році інша група вчених опублікували кріо-ЕМ структуру TRPC4 з *Danio rerio* в апо-стані з роздільністю 3,6 Å (PDB ID: 6G1K) [26]. Ця структура підтверджує базову архітектуру гомотетрамеру TRPC4 і демонструє високий ступінь консервативності між TRPC4 *Mus musculus* та *Danio rerio* на рівні доменної організації. У 2019 р. Duan *et al.* визначили структуру TRPC5 миші з роздільністю 2,8 Å (PDB ID: 6AEI), що встановила структурні засади функціональної подібності TRPC4 і TRPC5 у межах підгрупи TRPC4/5 [48].

Структурна характеристика комплексів TRPC-каналів з низькомолекулярними модуляторами розпочалася у 2020 р., Wright *et al.*

визначили кріо-ЕМ структуру людського TRPC5 у комплексі з інгібітором Pico145 (PDB ID: 6YSN) з роздільністю 3,0 Å, показавши, що ліганд зв'язується у міжсубодиничній ліпідній кишені, де витісняє ендogenous фосфоліпід [11]. Song *et al.* отримали структуру людського TRPC5 у комплексі із клемізолом (Clemizole) (PDB ID: 7D4P), продемонструвавши, що він зв'язується у VSL-подібному домені кожної субодиниці і стабілізує закриту конформацію каналу [32]. Паралельно Vinayagam *et al.* визначили низку структур TRPC4 *Danio rerio* у комплексах з піридазинон/піперазинон-похідними інгібіторами серії GFB – GFB-8438 (PDB ID: 7B0S), GFB-8749 (PDB ID: 7B05) і GFB-9289 (PDB ID: 7B16), а також у комплексі з кальмодуліном [27]. Усі три GFB-інгібітори займають ту саму міжсубодиничну кишеню у межах VSL-домену, а ключовими контактними залишками є Tyr373, Phe413, Met441, Arg491 (гідрофобні та електростатичні взаємодії), з варіабельною участю бічних ланцюгів Leu495 і His369 залежно від хемотипу ліганду [27]. CaM зв'язується з «rib»-спіраллю TRPC4 у ділянці CIRB і фіксує канал у закритій конформації через впорядкування раніше неупорядкованої ділянки [27].

У 2025 р. Won *et al.* опублікували першу групу структур людського каналу TRPC4 у гетеромерному і гомомерному контекстах: гетеротетрамер TRPC1/TRPC4 у стехіометрії 1:3 в апо-стані (PDB ID: 8WPL) і у комплексі з Pico145 (PDB ID: 8WPM), а також гомотетрамер TRPC4 людини в апо-стані (PDB ID: 8WPN) [10]. Присутність TRPC1 зумовлює асиметричну організацію іон-провідного шляху: селективний фільтр і нижні ворота набувають асиметрії, що корелює зі зниженою катіонною провідністю гетеромеру відносно гомомеру TRPC4. Pico145 у комплексі 8WPM локалізований у чотирьох еквівалентних міжсубодиничних кишенях, розташованих на інтерфейсах TRPC1–TRPC4 і TRPC4–TRPC4.

Перелік основних кріо-ЕМ структур каналів підгрупи TRPC1/4/5, доступних у PDB і використовуваних у подальших розділах дисертації, наведено у таблиці 1.1.

## Кріо-ЕМ структури каналів підгрупи TRPC1/4/5.

| PDB ID | Вид / канал                          | Стан        | Роздільна здатність, Å | Джерело                      | Рік  |
|--------|--------------------------------------|-------------|------------------------|------------------------------|------|
| 5Z96   | <i>Mus musculus</i> ,<br>TRPC4       | apo         | 3,3                    | Duan <i>et al.</i><br>[25]   | 2018 |
| 6G1K   | <i>Danio rerio</i> ,<br>TRPC4        | apo         | 3,6                    | Vinayagam <i>et al.</i> [26] | 2018 |
| 6AEI   | <i>Mus musculus</i> ,<br>TRPC5       | apo         | 2,8                    | Duan <i>et al.</i><br>[48]   | 2019 |
| 6YSN   | <i>Homo sapiens</i> ,<br>TRPC5       | + Pico145   | 3,0                    | Wright <i>et al.</i><br>[11] | 2020 |
| 7D4P   | <i>Homo sapiens</i> ,<br>TRPC5       | + clemizole | 2,7                    | Song <i>et al.</i><br>[32]   | 2021 |
| 7B0S   | <i>Danio rerio</i> ,<br>TRPC4        | + GFB-8438  | 3,6                    | Vinayagam <i>et al.</i> [27] | 2020 |
| 7B05   | <i>Danio rerio</i> ,<br>TRPC4        | + GFB-8749  | 3,8                    | Vinayagam <i>et al.</i> [27] | 2020 |
| 7B16   | <i>Danio rerio</i> ,<br>TRPC4        | + GFB-9289  | 3,15                   | Vinayagam <i>et al.</i> [27] | 2020 |
| 8WPL   | <i>Homo sapiens</i> ,<br>TRPC1/TRPC4 | apo         | 3,0                    | Won <i>et al.</i><br>[10]    | 2025 |
| 8WPM   | <i>Homo sapiens</i> ,<br>TRPC1/TRPC4 | + Pico145   | 3,43                   | Won <i>et al.</i><br>[10]    | 2025 |

|      |                                |     |      |                           |      |
|------|--------------------------------|-----|------|---------------------------|------|
| 8WPN | <i>Homo sapiens</i> ,<br>TRPC4 | apo | 2,82 | Won <i>et al.</i><br>[10] | 2025 |
|------|--------------------------------|-----|------|---------------------------|------|

### 1.2.3. Сайти зв'язування лігандів

Накопичені структурні дані дозволяють виділити кілька функціонально значущих сайтів зв'язування на поверхні каналу TRPC4 та його близького гомолога TRPC5, які мають відмінну топологію, біохімічну природу та набір характерних лігандів.

Найбільше охарактеризованим є міжсубодиничний сайт у межах VSL-домену, розташований на інтерфейсі між сегментами S1–S4 одного протомеру і S5–S6 сусіднього. У цій кишені зв'язуються піридазинон/піперазинон-похідні інгібітори серії GFB (GFB-8438, GFB-8749, GFB-9289 і споріднений клінічний кандидат GFB-887), а також клемізол [27, 32]. За структурами TRPC4 *Danio rerio* (PDB ID: 7B0S, 7B05, 7B16) ключовими контактними залишками є Tyr373, Phe413, Met441, Asn442, Arg491, Leu495 і Tyr646; додаткова стабілізація комплексу забезпечується бічними ланцюгами His369, Glu417, Asp438 і Ser494, участь яких варіює залежно від хемотипу ліганду [27]. Зв'язування інгібіторів цього хемотипу індукує переорієнтацію бічних ланцюгів кишені за механізмом «induced fit» і фіксує канал у закритій конформації через сигнал, що передається через TRP-домен до нижніх воріт [27].

Другим сайтом є консервативна ліпідзв'язувальна кишеня трансмембранного домену, прилегла до пори, у якій розміщуються ксантин-похідні інгібітори Pico145 і HC-070 [11, 32]. Wright *et al.* показали, що у комплексі людського TRPC5 з Pico145 (PDB ID: 6YSN) ліганд витісняє ендогенний фосфоліпід з цієї кишені, а ключовими контактними залишками є Leu521, Tyr524, Cys525, Leu572, Gln573, Trp577, Thr603, Gly606 і Val610 [11]. У гетеромері TRPC1/TRPC4 з Pico145 (PDB ID: 8WPM) та у гомомері

людського TRPC4 (PDB ID: 8WPN) аналогічна кишеня реалізується у чотирьох еквівалентних копіях на різних міжсубодиничних інтерфейсах зі збереженням складу ключових залишків [10].

Третій тип сайтів становлять катіон-координаційні ділянки всередині VSL-домену з цитоплазматичного боку. У структурі 5Z96 на цитозольному боці S1–S4 кишені ідентифіковано гідрофільну порожнину зі зв'язаним катіоном (ймовірно,  $\text{Na}^+$ ), координованим бічними ланцюгами Glu417, Gln420 (S2), Asn435 і Asp438 (S3) [25]. Розташування негативно заряджених Glu417 і Asp438 консервативне серед усіх представників підродини TRPC, що свідчить про загальну роль цього сайту. Додатково у 5Z96 виявлено до восьми ділянок електронної щільності, віднесених до фосфоліпідів і гемісукцинатів холестеролу, розташованих на поверхні взаємодії N-кінцевого домену і лінкера S4–S5 (чотири копії) та у проміжку між субодиницями біля порової спіралі (один фосфоліпід) [25]. Ці ліпідні контакти опосередковують стабілізацію тетрамерної збірки і є потенційними ділянками ліпід-залежної регуляції активності каналу.

### 1.3. Регуляція активності TRPC4

Регуляція активності TRPC4 є багаторівневою і залучає компоненти PLC-сигнального каскаду, пряму модуляцію G-білками, ліпідні другорядні посередники, іони  $\text{Ca}^{2+}$ , посттрансляційні модифікації та взаємодію з цитоскелетом [49, 50]. Ключова особливість TRPC4 – подвійна активація через Gq/11-PLC $\beta$  і Gi/o шляхи, яка забезпечує інтеграцію сигналів від різних GPCR-рецепторів у одну функціональну відповідь [51].

Активація через Gq/11-PLC каскад реалізується шляхом гідролізу  $\text{PIP}_2$  з утворенням DAG та  $\text{IP}_3$ . Показано, що саме активність PLC $\delta 1$  є вирішальною для активації повнорозмірного TRPC4 $\alpha$  у клітинних моделях, тоді як інші ізоформи PLC (зокрема PLC $\beta$ ) можуть лише підсилювати відповідь, але не

ініціювати її самостійно [52]. Інгібування PLC $\delta$ 1 малою ГТФазою RhoA пригнічує активацію TRPC4 і, ймовірно, пояснює відносно повільну кінетику активації mI<sub>CAT</sub> у нативних клітинах. PIP<sub>2</sub> сам по собі є інгібітором TRPC4 $\alpha$ ; делеція С-кінцевої ділянки 781–864 (ізоформа TRPC4 $\beta$ ) знімає інгібіторний ефект PIP<sub>2</sub>, що свідчить про пряму взаємодію фосфоліпиду з С-кінцевим доменом каналу [47]. Згодом було ідентифіковано додатковий тонічний тип PIP<sub>2</sub>-залежного інгібування, який характерний для обох ізоформ і не залежить від фрагмента з 84 амінокислот на С-кінці [52].

IP<sub>3</sub> регулює TRPC4 через прямий контакт IP<sub>3</sub>-рецептора з С-кінцевим мотивом CIRB каналу [1]. Паралельно до цього кальмодулін конкурує з IP<sub>3</sub>-рецептором за той самий сайт CIRB і у Ca<sup>2+</sup>-залежний спосіб інгібує канал у стані спокою; за зниження цитозольного Ca<sup>2+</sup> IP<sub>3</sub>-рецептор витісняє CaM, що призводить до реактивації каналу [53]. Ca<sup>2+</sup> сам по собі є біфазним регулятором TRPC4: помірне підвищення цитозольного Ca<sup>2+</sup> є тригером активації (через PLC та CaM/IP<sub>3</sub>R), тоді як високі концентрації Ca<sup>2+</sup> викликають десенситизацію каналу [1, 53].

Пряма модуляція TRPC4 G $\alpha_{i/o}$ -субодиницями є унікальною для підгрупи TRPC4/5 серед TRP-каналів [51]. G $\alpha_{i2}$ /G $\alpha_{i3}$  зв'язуються з N-кінцевим ARD у ділянці залишків Ile58-Tyr59-Tyr60 (за нумерацією TRPC5) і активують канал незалежно від PLC-шляху [1, 45, 54]. Таким чином, TRPC4/5 слугують «детекторами співпадіння» Gq- і Gi-сигналів, що інтегрують сигнали від різних GPCR у мускариновій нервовій передачі й деполяризації нативних клітин [45].

Чутливість TRPC4 до DAG динамічно регулюється Na<sup>+</sup>/H<sup>+</sup>-обмінним регуляторним фактором NHERF. Зв'язування NHERF з С-кінцем TRPC4 через PDZ-домени маскує DAG-чутливість каналу, а фосфорилювання треонінового залишку PDZ-зв'язувального мотиву протеїнкіназою С (PKC) призводить до дисоціації NHERF і відновлення DAG-чутливості [47, 55]. Додатковою ланкою регуляції є тирозинкінази: стимуляція епідермальним фактором росту

(EGF) спричиняє фосфорилування каналу, що у поєднанні з NHERF підсилює поверхневу експресію TRPC4 та чутливість до активаторів [1].

Вкорочена ізоформа TRPC4 $\beta$  містить внутрішню делецію 84 амінокислотних залишків у цитоплазматичному С-кінцевому домені (порівняно з повнорозмірною ізоформою TRPC4 $\alpha$ ), при цьому термінальний PDZ-зв'язувальний мотив зберігається. Ця делеція знімає PIP<sub>2</sub>-залежне інгібування каналу і супроводжується редукованою чутливістю до Gq/11- та Gi/o-опосередкованих сигнальних шляхів [47, 51]. Це підтверджує вирішальну роль С-кінцевого домену у рецептор-опосередкованій активації каналу і вказує на пластичність регуляторного профілю TRPC4 за рахунок альтернативного сплайсингу.

## **1.4. Фізіологічне значення TRPC4**

### **1.4.1. Експресія та розподіл у тканинах**

TRPC4 має широкий тканинний розподіл у ссавців. Найвищі рівні експресії виявлено у гладеньких м'язах кишечника, ендотелії судин, нирках (зокрема у подоцитах), кортиколімбічних структурах центральної нервової системи (ЦНС), наднирниках, сітківці ока та плаценті [1]. У ЦНС TRPC4 переважно експресований у корі, гіпокампі та базолатеральному ядрі амигдали, де разом з TRPC5 і TRPC1 опосередковує синаптичну пластичність і регулює збудливість нейронів [56]. У периферичній нервовій системі (ПНС) TRPC4 виявлено у дорсальних корінцевих гангліях, де він бере участь у формуванні ноцицептивної чутливості [1].



#### 1.4.2. TRPC4 у шлунково-кишковому тракті

Канал TRPC4 є ключовим молекулярним компонентом  $mI_{CAT}$  у вісцеральних гладеньких м'язах ШКТ. Експерименти з нокаутними TRPC4 мишами показали, що приблизно 85 % амплітуди  $mI_{CAT}$  у гладеньком'язових клітинах клубової кишки миші опосередковується саме TRPC4; залишкова частина припадає переважно на TRPC6, що діє у гетеромерних комплексах з TRPC4 [19, 20]. Відповідно, на подвійних TRPC4; TRPC6 мишах спостерігається практично повна відсутність  $mI_{CAT}$  і суттєве порушення моторики тонкої кишки *in vivo* [19].

Фізіологічний ланцюг, у якому TRPC4 виконує центральну роль у ШКТ, такий: ацетилхолін, що вивільняється з холінергічних нервових закінчень, активує  $M_2$ - та  $M_3$ -мускаринові рецептори на мембрані гладеньком'язових клітин [57]. Активація  $M_3$  через Gq/11-PLC $\beta$  гідролізує PIP $_2$  з утворенням DAG та IP $_3$ , тоді як  $M_2$  через Gi/o пригнічує аденілатциклазу і напряду модулює TRPC4 через G $\alpha_{i/o}$ -субодиниці [45, 53, 58–61]. Синергічна дія Gq- і Gi-шляхів активує TRPC4-опосередкований катіонний струм, що зумовлює деполяризацію плазматичної мембрани. Деполяризація активує потенціал-залежні Ca $^{2+}$ -канали L-типу, забезпечуючи вхід Ca $^{2+}$  і скорочення гладеньком'язових клітин [53].

Додатковим компонентом ШКТ-фізіології TRPC4 є його функціональна роль в інтерстиційних клітинах Кахаля, які задають пейсмейкерну активність кишечнику [62]. Клітини Кахаля генерують повільнохвильові потенціали, а холінергічне збудження гладеньких м'язів через TRPC4 модулює амплітуду та частоту цих хвиль, синхронізуючи нейроефекторну трансмісію з пейсмейкерною активністю [62]. Амплітуда і кінетика  $mI_{CAT}$  додатково модулюються внутрішньоклітинним вивільненням Ca $^{2+}$  через IP $_3$  рецептори, що є однією з ланок холінергічної відповіді гладеньких м'язів [63–65]. Цей локальний Ca $^{2+}$ -сигнал діє безпосередньо на мембранні іонні канали гладеньком'язової клітини [66]. Порушення цієї взаємодії спостерігається при

функціональних розладах ШКТ, включно із синдромом подразненого кишечника та післяопераційним парезом [66].

Селективна фармакологічна блокада TRPC4 відтворює фенотип генетичного нокауту: аплікація Pico145 у пікомолярних концентраціях пригнічує  $mI_{CAT}$  більш ніж на 80% і ослаблює карбахол-індуковані скорочення ізольованих сегментів клубової кишки миші [20]. Це надає експериментальне підтвердження концепції TRPC4 як фармакологічної мішені для корекції надмірної скоротливої активності ШКТ у клінічних станах.

#### **1.4.3. TRPC4 у центральній та периферичній нервовій системі**

У ЦНС TRPC4 бере участь у регуляції синаптичної пластичності, пресинаптичного гомеостазу  $Ca^{2+}$  та вивільнення нейротрансмітерів [56]. У гіпокампальних нейронах TRPC4 модулює короточасну пластичність швидких глутаматергічних синапсів [56], а у таламічних інтернейронах опосередковує серотонін-індуковане вивільнення  $\gamma$ -аміномасляної кислоти (GABA) через рецептор-опосередкований вхід  $Ca^{2+}$  [52]. У корі й амігдалі TRPC4/TRPC5 залучені у формування тривожних і депресивних станів, що обґрунтовує терапевтичний інтерес до селективних інгібіторів цієї пари каналів у лікуванні афективних розладів [67]. У ПНС підвищення експресії TRPC4 спостерігається під час регенерації аксонів дорсальних корінцевих гангліїв; інгібування каналу сповільнює ріст нейритів, що вказує на його роль у нейрональній пластичності та відновленні після травми [1].

#### **1.4.4. TRPC4 у серцево-судинній системі, нирках та інших системах**

У ендотелії судин TRPC4 опосередковує тромбін-індуковане підвищення проникності ендотелію, що є ключовим механізмом у патогенезі

травми судин і запалення тканин. У шлуночкових кардіоміоцитах цАМФ-залежний білок EPAC підсилює TRPC4-залежний вхід  $\text{Ca}^{2+}$ , що пов'язують з проаритмічним ефектом [1]. У нирках TRPC4 (частково у комплексі зі STIM1L) бере участь у регуляції цитоскелету подоцитів і формуванні депо-залежного входу  $\text{Ca}^{2+}$  у епітеліальних клітинах каналців [68]. У постнатальній диференціації міобластів TRPC4 разом з TRPC1 підтримує швидкі повторювані  $\text{Ca}^{2+}$ -сигнали, що перешкоджають втомі м'язів [1].

## **1.5. Патологічні стани, пов'язані з TRPC4**

### **1.5.1. Патологія шлунково-кишкового тракту**

З огляду на ключову роль TRPC4 у  $\text{mI}_{\text{CAT}}$  та холінергічному збудженні гладеньких м'язів ШКТ, порушення функціонування цього каналу потенційно пов'язане з функціональними розладами моторики [1, 5, 6]. Надмірна активація TRPC4-опосередкованих катіонних струмів може сприяти розвитку надмірної скоротливої активності та спастичних станів, тоді як недостатня активація – гіпомоторних і псевдообструктивних синдромів [1]. Крім фармакологічних агентів, на  $\text{mI}_{\text{CAT}}$  здатні впливати й інші ксенобіотики: показано, зокрема, що фулерени  $\text{C}_{60}$  відкривають канали TRPC4/TRPC6 через активацію мускаринових рецепторів і G-білків у гладеньких м'язах тонкого кишечника [69]. У моделях загального анестезу і післяопераційного парезу показано, що низка анестетиків і анксіолітиків пригнічує TRPC4-опосередкований  $\text{mI}_{\text{CAT}}$  і порушує моторику ШКТ, що може пояснювати клінічно значущий побічний ефект цих препаратів; зокрема, для кетаміну показано пригнічення  $\text{mI}_{\text{CAT}}$  з можливістю його відновлення агоністом TRPC4 (–)-енглерином А [1, 70–72].

Функціональне значення TRPC4 у формуванні вісцерального болю показано у моделях запалення та нейропатичного болю: TRPC4 миші демонструють знижену чутливість до вісцерального болю, а неселективні антагоністи TRPC4/5 (наприклад, ML204) пригнічують больову реакцію в цих моделях і зменшують механічну гіперчутливість при нейропатичному болю [28, 73]. Це формує експериментальне підґрунтя для розгляду TRPC4-селективних інгібіторів як потенційних засобів у лікуванні функціональних розладів ШКТ, асоційованих з вісцеральним болем (зокрема синдрому подразненого кишечника).

TRPC4 також обговорюється у контексті синдромів діареї, спричинених надмірним холінергічним збудженням, а також у патофізіології розладів моторики, асоційованих з кишковою нейропатією в діабеті та старінні [1]. Наявність селективних інструментальних блокаторів (Pico145) і клінічних кандидатів (HC-070, BI 1358894) створює перспективи для трансляційного вивчення TRPC4 як мішені для нових антиспастичних терапевтичних стратегій.

### 1.5.2. Інші патологічні стани

Найбільш розробленою клінічно мішенню у межах TRPC4/5-підгрупи є TRPC5 та його роль у патогенезі протеїнурії при FSGS та споріднених подоцитопатіях [31]. Активація TRPC5 у подоцитах через шлях Rac1 є первинним драйвером протеїнурії у низці моделей FSGS; селективний інгібітор GFB-887 захищає подоцити від протамін-індукованого пошкодження *in vitro* та ослаблює протеїнурію у гіпертензивних та негіпертензивних моделях FSGS у щурів [31]. На підставі цих даних розгорнуто клінічні випробування GFB-887 у пацієнтів з FSGS, діабетичною нефропатією, тощо (phase 2a TRACTION-2, NCT04387448) [12]. Паралельно у

клінічних дослідженнях перебуває BI 1358894 (Boehringer Ingelheim) – TRPC4/5-інгібітор для лікування афективних розладів [67].

У серцево-судинній системі TRPC4 розглядається як потенційна мішень для корекції аритмічних станів і ендотеліальної бар'єрної дисфункції (особливо при гострому ушкодженні легень) [1]. У нервовій системі TRPC4/1 беруть участь у епілептиформній активності у вигляді пачкових розрядів та викликаній епілепсією нейродегенерації: показано, що гетеромерні канали TRPC1/TRPC4 відіграють вирішальну роль у генерації епілептиформних розрядів, а їхній подвійний нокаут суттєво зменшує судомну активність і пов'язану із судомами загибель нейронів [40].

## 1.6. Модулятори TRPC4

### 1.6.1. Загальна характеристика

Ідентифікація та розробка модулаторів активності TRPC4 досягли значного розвитку за останнє десятиліття завдяки поєднанню високоефективного скринінгу, *in silico* підходів та валідації на моделях *in vitro* й *in vivo*. Відомі модулятори класифікуються за двома основними ознаками: напрям дії (активатор / інгібітор) і субтип-селективність (у межах підгрупи TRPC1/4/5 або між підгрупами TRPC4/5 та TRPC3/6/7).

Серед інгібіторів TRPC4 історично першим широко вживаним інструментом був ML204 – помірно потужний селективний антагоніст TRPC4/5 з IC<sub>50</sub> близько 1 мкмоль для TRPC4 і близько 65 % інгібуванням TRPC5 при 10 мкмоль/л [28]. Обмеженням ML204 є відносно низька потенція, слабший ефект на TRPC1-вмісні гетеромери та недостатня селективність до інших іонних каналів при високих концентраціях. Клемізол – це H1-гістаміновий антагоніст, який було перепрофільовано як інгібітор TRPC5; для нього визначено кріо-ЕМ структуру комплексу з TRPC5 людини

(PDB ID: 7D4P) [32]. Сполука ML204 була ідентифікована в межах ініціативи NIH Molecular Libraries Probe Production Centers Network шляхом високопродуктивного скринінгу (high-throughput screening, HTS) бібліотеки з понад 305 000 малих молекул із використанням кальцієвої флуоресцентної зчитувальної системи на клітинах HEK293 з гетерологічною експресією TRPC4, з подальшим вторинним функціональним підтвердженням методом петч-клемп [28]. Такий скринінг є дуже ресурсозатратним, натомість обчислювальні методи, реалізовані в цій роботі, дозволяють звузити простір пошуку до  $10^2$ – $10^3$  кандидатних сполук ще до експериментального тестування.

Наступним поколінням інгібіторів стали ксантин-похідні сполуки – Pico145 (також відомий як HC-608) і структурно споріднені HC-070 та AM237 [32, 74]. Усі три сполуки зв'язуються у консервативній ліпідзв'язувальній кишені трансмембранного домену і демонструють високу афінність щодо TRPC1/4/5-каналів при збереженні високої селективності відносно інших TRP- і потенціал-залежних каналів. Pico145 є «еталонним» інструментальним інгібітором для функціональних досліджень TRPC4/5, тоді як для HC-070 продемонстровано виражений анксиолітичний і антидепресантоподібний ефект у мишей, що обґрунтовує інтерес до TRPC4/5-інгібіторів як кандидатів для лікування тривожно-депресивних станів [32]. AM237, структурно близький до Pico145 та HC-070, виявляє наномолярну агоністичну активність на гомомерах TRPC5:C5 і водночас антагоністичну дію на гомомерах TRPC4:C4 і гетеромерах TRPC1:C5 та TRPC1:C4 – демонструє принципову можливість субтип-селективного перемикування на пряму дії у межах одного хемотипу [74]. Цей ксантин-похідний хемотип TRPC1/4/5-інгібіторів був отриманий унаслідок цілеспрямованої медико-хімічної оптимізації стартових сполук у лабораторії D. J. Beech (University of Leeds, Велика Британія) із залученням ітеративних циклів синтез–структура–активність та функціонального скринінгу на клітинах з гетерологічною експресією TRPC1/4/5 [74].

Окремий клас становлять піридазинон/піперазинон-похідні інгібітори серії GFB (Goldfinch Bio): GFB-8438, GFB-8749, GFB-9289 та клінічний кандидат GFB-887. Ці сполуки характеризуються наномольною потенцією щодо TRPC5, структурно визначеним сайтом зв'язування у VSL-доміні (PDB ID: 7B0S/7B05/7B16) і вираженим захисним ефектом на подоциту у моделях FSGS [27, 31]. BI 1358894 – TRPC4/5-інгібітор Boehringer Ingelheim, який у клінічних дослідженнях продемонстрував редукцію психологічних і фізіологічних відповідей на індуковані тривожно-панічні симптоми у здорових добровольців, а також знижену активацію ділянок мозку, залучених у емоційну обробку, за даними функціональної нейровізуалізації [67]. M084 – ще один селективний інгібітор TRPC4/5, для якого у мишей продемонстровано антидепресантоподібні та анксиолітикоподібні ефекти після гострого введення [56]. AC1903 – сполука, яка раніше розглядалась як селективний інгібітор TRPC5, однак нещодавні дослідження показали, що вона діє через додаткові мішені і не є істинно субтип-специфічною [1].

### 1.6.2. Pico145 і GFB-887

Pico145 (синоніми: HC-608; CAS 1628287-16-0) – ксантин-похідний інгібітор TRPC1/4/5-каналів, вперше охарактеризований Rubaiy *et al.* у 2017 році [29]. Згідно з первинним дослідженням, потенція Pico145 залежить від субтипу та механізму активації TRPC1/4/5 і лежить у діапазоні 9–1300 пмоль/л [29]. Зокрема, Pico145 блокує  $\text{Ca}^{2+}$ -вхід через TRPC5:C1 і TRPC4:C1 з  $\text{IC}_{50}$  0,199 і 0,03 нмоль/л [29]. Pico145 не впливає на TRPC3, TRPC6, TRPV1, TRPV4, TRPA1, TRPM2, TRPM8, а також на депо-залежне надходження  $\text{Ca}^{2+}$ , опосередковане Orai1, що робить його однією з найселективніших фармакологічних сполук у межах родини TRP. У гладеньком'язових клітинах клубової кишки миші середнє значення  $\text{IC}_{50}$  для інгібуючого ефекту Pico145 на  $\text{mI}_{\text{CAT}}$  становить близько 3,1 пмоль/л [20].

Структурні основи дії Pico145 з'ясовано Wright *et al.* у 2020 році на кріо-ЕМ структурі комплексу людського TRPC5 з Pico145 (PDB ID: 6YSN) з роздільністю 3,0 Å [11]. Pico145 зв'язується у консервативній ліпідній кишені трансмембранного домену, де витісняє ендogenous фосфоліпід. Стабілізація комплексу забезпечується поєднанням  $\pi$ - $\pi$  (T-shaped) взаємодії ксантинового ядра з Tyr524,  $\pi$ -sulfur контакту з Cys525, гідрофобно-ароматичних взаємодій з бічними ланцюгами Leu521, Leu572, Gln573, Thr603, Gly606 і Val610, а також водневого зв'язку з Trp577 [11]. Додаткове підтвердження консервативності сайту зв'язування у межах підгрупи TRPC1/4/5 отримано у 2025 році, коли Won *et al.* опублікували кріо-ЕМ структуру людського гетеротетрамеру TRPC1/TRPC4 у комплексі з Pico145 (PDB ID: 8WPM) у стехіометрії 1:3 [10]. У цій структурі Pico145 виявлено у чотирьох еквівалентних кишенях на інтерфейсах TRPC1–TRPC4 і TRPC4–TRPC4 з консервативним складом ключових контактних залишків. Гетеромер демонструє асиметричну архітектуру селективного фільтра і нижніх воріт, зумовлену присутністю TRPC1 [10].

GFB-887 – піперазинон/піридазинон-похідний інгібітор TRPC5, розроблений компанією Goldfinch Bio як подоцит-таргетований кандидат для лікування протеїнуричних захворювань нирок [31]. У первинних доклінічних дослідженнях показано, що GFB-887 захищає подоцити від протамін-індукованої втрати стресових волокон, відновлює їх після нокдауну синаптоподіну і ослаблює протеїнурію у моделях FSGS: деоксикортикостерон-ацетат/унілатерально нефректомованих щурів та пурамицин-аміноглікозидних (PAN) щурів без зміни артеріального тиску [31].

Клінічна трансляція GFB-887 представлена дослідженням TRACTION-2 (NCT04387448) – фаза 2а, плацебо-контрольоване, подвійно сліпе дослідження з триразовим дозовим нарощуванням у пацієнтів з FSGS, TR-MCD і діабетичною нефропатією [12]. У дослідженні «перший у людині» GFB-887 продемонстрував безпечний і добре переносимий профіль, дозопівставну фармакокінетику, сумісну з одноразовим добовим



дозуванням, і дозозалежне зниження сечового Ras1 – запропонованого біомаркера активності TRPC5-Ras1 шляху. Первинною кінцевою точкою TRACTION-2 є зміна протеїнурії після 12 тижнів лікування, з адаптивним проміжним аналізом для оптимізації подальшого набору пацієнтів [12].

### 1.7. *In silico* методи у розробці модуляторів TRPC4

Розробка лікарських засобів є складним, багатоетапним і дорогим процесом, у якому доклінічна стадія включає ідентифікацію та валідацію молекулярної мішені, ідентифікацію сполук-хітів, відбір та оптимізацію сполук-лідів, а також доклінічні дослідження *in vitro* та *in vivo* [75]. Методи CADD скорочують тривалість і вартість цього процесу за рахунок раціонального обмеження числа сполук, що передаються на експериментальне тестування [33, 34, 75, 76]. Сучасні CADD-підходи інтегрують структурну інформацію про білок-мішень, лігандні бібліотеки, методи машинного навчання (ML) і глибокого навчання (DL) та становлять основу для систематичного скорочення використання тваринних моделей на ранніх етапах доклінічної розробки [33, 34]. У межах цієї дисертації найвищу релевантність мають чотири класи *in silico* методів: ідентифікація сайтів зв'язування, молекулярний докінг, молекулярна динаміка та прогнозування ADMET.

Точна ідентифікація та пріоритезація сайтів зв'язування малих молекул на поверхні білка-мішені є фундаментальним елементом раціонального пошуку модуляторів, особливо для тетрамерних іонних каналів на кшталт TRPC4, у яких множинні міжсубодиничні кишені можуть приймати ліганди різних хемотипів [14]. Обчислювальні методи ідентифікації сайтів зв'язування поділяють на геометричні, енергетичні, еволюційно-консенсусні, ML/DL і гібридні. Геометричні методи виявляють порожнини (cavities) і щілини (clefts), здатні вмістити ліганд. Їх найпоширеніший представник

Focket використовує альфа-сфери на основі тесселяції Вороного, а до цього класу належать також CASTp 3.0 і CB-Dock2 [77–80]. Енергетичні методи (EASYMIFs, SITEHOUND) локалізують сприятливі для зв'язування ділянки за полями молекулярних взаємодій, але є обчислювально витратними. Еволюційно-консенсусні методи додають до геометрії аналіз збереженості послідовності, проте втрачають ефективність для білків з малим числом гомологів і для неконсервативних («cryptic») сайтів. Методи ML/DL (DeepSite, DeepPocket, PURESNet, PURESNetV2.0, FRASE-bot) інтегрують геометричні, енергетичні та еволюційні ознаки у нейромережевому предикторі, досягаючи вищої точності на складних поверхнях і прихованих кишнях ціною обчислювальних ресурсів і залежності від навчальних наборів [56, 81–86]. Гібридні методи (DELIA, PGocket, P2Rank) поєднують кілька підходів, що підвищує точність, але ускладнює налаштування [87–89]. Спільним недоліком більшості методів є надмірне передбачення кишень – більше, ніж виправдано експериментальними даними: з одного боку це корисно для виявлення прихованих і алостеричних сайтів, з іншого перешкоджає швидкій пріоритезації у комерційному пошуку ліків, а ручний відбір за публікаціями є трудомістким і схильним до помилок [14]. Нещодавно запропоновано гібридний підхід, який поєднує передбачення кишень Focket з літературно-зумовленою пріоритезацією за допомогою ВММ: модель витягує з наукових статей амінокислотні залишки експериментально підтверджених сайтів зв'язування, що слугують фільтром для передбачень Focket, після чого виконуються картування, злиття та уточнення об'ємного представлення кишень [14].

Молекулярна динаміка (MD) моделює часову еволюцію білкової системи через чисельне інтегрування рівнянь руху для всіх атомів у силовому полі [90]. Для мембранних білків, до яких належить TRPC4, стандартом є поєднання силового поля CHARMM36m (для білків та ліпідів) з модельним ліпідним бішаром на основі пальмітоїлолеоїлфосфатидилхоліну (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, POPC) або складнішими

сумішами, що відтворюють ліпідний склад нативних мембран [91]. Попередня підготовка системи (позиціонування білка у мембрані, сольватація, нейтралізація іонами) зазвичай виконується за допомогою CHARMM-GUI Membrane Builder [92]. MD-симуляції TRPC4 дозволяють охарактеризувати конформаційну пластичність білка, динаміку нижніх воріт і селективного фільтра, шлях проникнення іонів, ліпід-білкові контакти та конформаційну відповідь на зв'язування модуляторів. Додатково вільноенергетичні розрахунки (FEP; umbrella sampling; metadynamics) надають кількісні оцінки кінетичних і термодинамічних параметрів зв'язування лігандів, що є критичним для раціональної оптимізації сполук-лідів. Для обробки траєкторій використовують пакети GROMACS, NAMD і спеціалізовані бібліотеки на кшталт MolAR [93, 94].

Оцінка ADMET-властивостей є критичним етапом ранньої доклінічної розробки, що дозволяє зменшити відсоток провалів на останніх стадіях для програм по розробці лікарських засобів [34]. Експериментальні ADMET-дослідження надійні, але витратні й часозатратні, тому у високопродуктивному скринінгу їх доповнює *in silico* ADMET на основі моделей кількісного зв'язку структура-активність/властивість (QSAR/QSPR) [34]. Сучасні ADMET-моделі використовують широкий спектр молекулярних репрезентацій: мануально створені дескриптори (наприклад, Mordred-дескриптори), розширені циклічні відбитки (extended connectivity fingerprints, ECFP / Morgan fingerprints), графові нейронні мережі (GNN) та ембединги з мовних моделей, навчених на SMILES-репрезентаціях молекул (ChemGPT, ZairaChem) [95–97]. Високорозмірність вхідних ознак, типова для комбінованих дескрипторно-ембедингових представлень, вимагає автоматизованих процедур селекції ознак для контролю перенавчання, особливо у задачах з малими навчальними вибірками [98]. Ембединг Mol2Vec є некерованим навчанням на субструктурних векторних представлень за аналогією до Word2Vec, що є ефективним у комбінації з класичними дескрипторами [99]. Його розширена реалізація з доповненням

дескрипторами та автоматичною селекцією ознак посіла першість у 10 з 16 ADMET-завдань платформи Therapeutics Data Commons (TDC) [17]. TDC надає стандартизовані датасети, інструменти обробки та лідерборд для порівняння ADMET-моделей і охоплює 22 куровані набори з ключовими фармакокінетичними та токсикологічними параметрами: проникність через моношар клітин Caco-2, людську кишкову абсорбцію (HIA), інгібування Р-глікопротеїну (P-glycoprotein, P-gp), пероральну біодоступність, проникність через гематоенцефалічний бар'єр (BBB), зв'язування з білками плазми (PPB), об'єм розподілу у стаціонарному стані (VDss), інгібування та субстратність ізоформ CYP450 (CYP2C9, CYP2D6, CYP3A4), мутагенність за тестом Еймса (Ames mutagenicity), лікарсько-індуковане ушкодження печінки (DILI) та блокування hERG-каналу (human ether-à-go-go-related gene, hERG) – потенціал-залежного калієвого каналу серця, блокування якого асоційоване з розвитком аритмії типу torsades de pointes [100, 101]. Детальну архітектуру застосовуваних у цій дисертації ADMET-моделей та їх валідацію на TDC-бенчмарках наведено у Розділі 5.

Систематизація літературних даних щодо структури, регуляції, фізіологічного значення, патологічної ролі та фармакологічного профілю каналу TRPC4 дозволяє сформулювати такі узагальнюючі висновки. По-перше, TRPC4 є структурно і функціонально охарактеризованим тетрамерним неселективним катіонним каналом підгрупи TRPC1/4/5. Отримано кріо-ЕМ моделі (від 5Z96 у 2018 р. до 8WPL/8WPM/8WPN у 2025 р.), що охоплюють апо-стан, комплекси з низькомолекулярними модуляторами (Pico145, клемізол, GFB-8438/8749/9289) та гетеромерні канали TRPC1/TRPC4. По-друге, ключова роль TRPC4 у холінергічному збудженні гладеньких м'язів ШКТ (приблизно 85 % внесок у  $mI_{CAT}$ ) і у пацієнтів-релевантних патологіях (FSGS, тривожно-депресивні стани, епілепсія, онкологічні захворювання) підтверджує обґрунтованість розгляду TRPC4 як фармакологічної мішені. Наявність високоафінних інгібіторів пікомолярного (Pico145) і наномолярного (HC-070, GFB-887, BI 1358894)

діапазонів, частина з яких просунулася до клінічних досліджень (TRACTION-2 для GFB-887, Phase I для BI 1358894), створює передумови для подальшого структурно-керованого пошуку селективних TRPC4-модуляторів з оптимізованими ADMET-профілями. Попри значні досягнення у структурній характеристиці зв'язувальних сайтів TRPC4, залишається низка прогалин, що обмежують раціональний дизайн субтип-селективних модуляторів. По-перше, відсутні автоматизовані обчислювальні інструменти для пріоритезації сайтів зв'язування для каналів родини TRPC, а існуючі методи мають суттєві обмеження. По-друге, бракує структурно охарактеризованих сайтів, що є унікальними саме для TRPC4, тому в більшості випадків відомі модулятори є селективними саме для TRPC4/TRPC5. По-третє, прогнозування ADMET-параметрів досі рідко інтегрується з молекулярним докінгом та іншими *in silico* методами на ранніх етапах скринінгу, що підвищує ризик невдач на пізніх стадіях розробки лікарських засобів.

Систематизовані у цьому розділі структурно-функціональні та фармакологічні дані щодо TRPC4, разом із характеристикою доступних *in silico* інструментів, становлять базу, на якій будуються всі подальші методичні та експериментальні частини роботи.

## РОЗДІЛ 2

### МАТЕРІАЛИ І МЕТОДИ

У попередньому розділі систематизовано структурно-функціональні та фармакологічні характеристики TRPC4, а також окреслено *in silico* методи, релевантні для розв'язання дослідницьких завдань в межах цієї дисертації. У цьому розділі описано матеріали, обладнання та протоколи, використані в обчислювальній і експериментальній частинах роботи.

#### 2.1. Алгоритм ідентифікації та пріоритезації сайтів зв'язування

Для обґрунтованого визначення ділянок зв'язування малих молекул у структурі TRPC4 застосовували гібридний підхід пріоритезації кишень, що поєднує геометричний аналіз порожнин у просторовій структурі білка за допомогою програми Focket із літературно-зумовленою пріоритезацією на основі експериментально підтверджених амінокислотних залишків, витягнутих із наукових публікацій шляхом текстового аналізу з використанням BMM [14, 77]. Такий підхід забезпечує незалежність процедури ідентифікації сайту від хімічної структури ліганду, оскільки ґрунтується виключно на даних про білок-мішень.

**Дизайн експерименту.** Оскільки спеціалізованої процедури оцінювання для даної задачі в літературі не описано, використано дизайн самоабляційного експерименту (self-ablation) – порівнювалися три конфігурації конвеєра, що відрізняються складом активованих етапів:

1. Baseline – базова конфігурація з неоптимізованими запитами для етапів фільтрації та екстракції без етапу уточнення;
2. Optimized – конфігурація з оптимізованими запитами для фільтрації та екстракції без етапу уточнення;

3. Full pipeline – повна конфігурація, що доповнює Optimized етапом уточнення.

**Метрики на рівні кишень.** Для кількісної оцінки використано набір метрик на двох рівнях деталізації: на рівні кишень і на рівні амінокислотних залишків.

Точність за кількістю кишень (Pocket Number Accuracy) – частка публікацій, для яких кількість кишень, витягнутих ВММ, збігається з анотованою в еталоні. Для кожної публікації індикатор дорівнює 1 у разі збігу та 0 у протилежному випадку; підсумкова точність обчислюється як середнє арифметичне по датасету.

Повнота за кишенями (Pocket Recall) – частка анотованих кишень, для яких ВММ знайшла відповідну витягнуту кишеню:

$$\text{Pocket Recall} = n\_matched / n\_annotated,$$

де  $n\_matched$  – кількість анотованих кишень, для яких знайдено відповідну екстраговану кишеню;

$n\_annotated$  – загальна кількість анотованих кишень у еталоні.

Специфічність за кишенями (Pocket Specificity, TNR – true negative rate) відображає, як часто конвеєр коректно уникає передбачення кишені, коли жодної кишені не анотовано в еталонному наборі. Обчислюється як  $1 - \text{FPR}$  (False Positive Rate, частка хибнопозитивних результатів). FPR відображає частоту, з якою конвеєр генерує хибну кишеню, що не відповідає жодній анотованій кишені в еталоні. Визначається за формулою:

$$\text{Pocket Specificity} = 1 - n\_misses / n\_extracted,$$

де  $n\_misses$  – кількість витягнутих кишень, жодна з яких не відповідає жодній анотованій;

$n\_extracted$  – загальна кількість витягнутих кишень.

**На рівні залишків використано класичні метрики бінарної класифікації:** F1-оцінка, precision і recall, обчислені для списків залишків анотованих і витягнутих кишень. Для обрахування цих показників використовуються значення вірно класифікованих позитивних та негативних зразків (TP та TN), а також помилки першого та другого роду (FP та FN).

Recall – це відношення правильно передбачених позитивних спостережень до всіх позитивних спостережень. Recall обраховується за формулою:

$$\text{Recall} = TP / (TP + FN),$$

де TP – істинно позитивне значення;

FN – похибка другого роду.

Precision – це відношення правильно передбачених позитивних спостережень до загальної кількості прогнозованих позитивних спостережень. Precision обраховується за формулою:

$$\text{Precision} = TP / (TP + FP),$$

де TP – істинно позитивне значення;

FP – похибка першого роду.

F1-score – це середнє зважене значення recall та precision. Таким чином, ця оцінка враховує як помилкові позитивні, так і помилкові негативні значення. Інтуїтивно, ця метрика не така зрозуміла, як ассурасу, але F1-score зазвичай корисніший, якщо у тренувальній вибірці нерівномірний розподіл класів. F1-score обраховується за формулою:

$$\text{F1-score} = (2 \times \text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}),$$



де Precision – це відношення правильно передбачених позитивних спостережень до загальної кількості прогнозованих позитивних спостережень;

Recall – це відношення правильно передбачених позитивних спостережень до всіх позитивних спостережень.

**Цільові функції оптимізації.** Основними цільовими функціями під час оптимізації запитів були: максимізація F1-оцінки на рівні залишків, мінімізація FPR на рівні кишень, максимізація Pocket Number Accuracy та Pocket Recall. Інші метрики відстежувалися для повноти картини, але не були безпосередніми цілями оптимізації.

Детальний опис реалізації, параметрів та результатів роботи алгоритму на структурі TRPC4 миші (PDB ID: 5Z96) наведено у Розділі 3.

## 2.2. Молекулярний докінг

Молекулярний докінг виконували за допомогою програми AutoDock 4.2.6 [102], яка реалізує пошук оптимальної пози ліганду на основі сітки точок у межах визначеної області рецептора. У цьому методі молекула-мішень поміщається у сітку, після чого кожен атом ліганду послідовно розміщується у кожній точці сітки, та обчислюється енергія зв'язування між атомом ліганду і білком-мішенню; отримані значення зберігаються у файлі параметрів сітки. Основним методом конформаційного пошуку у AutoDock є генетичний алгоритм Ламарка (Lamarckian Genetic Algorithm, LGA) [103], під час роботи якого створюється популяція пробних конформацій, що з кожною генерацією зазнають мутацій та кросинговеру конформаційних параметрів; у підсумку з популяції обирається варіант з найменшою прогнозованою енергією зв'язування.

### 2.2.1. Підготовка молекул лігандів

Молекули лігандів Pico145 та GFB-887 готували до докінгу із застосуванням автоматизованого скрипту програмного пакета AutoDock Tools. Для кожної молекули ліганду розраховували часткові атомні заряди за методом Гастієра, після чого структуру зберігали у форматі PDBQT, який використовувався як вхідний файл для процедури докінгу.

### 2.2.2. Підготовка рецептора

Як рецепторну модель використовували кріо-електронно-мікроскопічну структуру мишачого каналу TRPC4 з ідентифікатором у банку білкових структур: 5Z96 [25]. Файл структури містить чотири ланцюги (A, B, C, D), кожен з яких відповідає окремій субодиниці тетрамерного каналу. Вибір саме цієї структури як рецепторної моделі обумовлений її високою роздільною здатністю, повним розв'язанням трансмембранного домену та відсутністю зв'язаного ліганду, що робить її оптимальною для дослідження механізму інгібування.

Для докінгу Pico145 використовували ланцюги A та B, що забезпечувало відтворення міжсубодиничної гідрофобної кишені трансмембранного домену, у якій за даними кріо-ЕМ зв'язується Pico145 у споріднених структурах TRPC1/TRPC4 та TRPC5 [10, 11]. Для докінгу GFB-887 використовували ланцюг A, оскільки сайт зв'язування лігандів серії GFB локалізується всередині VSL домену однієї субодиниці, що було показано у кріо-ЕМ структурах TRPC4 *Danio rerio* у комплексах з інгібіторами GFB-8438, GFB-8749 та GFB-9289 (PDB ID: 7B0S, 7B05, 7B16) [27].

Підготовку структури рецептора виконували у програмі BIOVIA Discovery Studio 4.5: структуру очищали від лігандів та молекул води. Далі у

середовищі AutoDock Tools до структури додавали полярні атоми гідрогену та розраховували часткові атомні заряди Коллмана. Підготовлену структуру зберігали у форматі PDBQT для подальшого використання.

Як референтні комплекси для оцінювання адекватності отриманих поз ліганду використовували кріо-ЕМ структуру гетеротетрамерного комплексу TRPC1/TRPC4 людини з Pico145 (PDB ID: 8WPM) для Pico145 та кріо-ЕМ структуру TRPC4 *Danio rerio* у комплексі з GFB-8438 (PDB ID: 7B0S) для GFB-887 [10, 27].

Для кожної процедури докінгу задавали кубічну область пошуку – грід-бокс, що визначав межі простору, у якому виконувався конформаційний пошук. Параметри грід-боксів обирали таким чином, щоб повністю охопити сайти зв'язування, попередньо визначені за допомогою описаного у підрозділі 2.1 методу ідентифікації та пріоритезації сайтів зв'язування. Роздільна здатність сітки у всіх випадках становила 0,375 Å; решта параметрів залишалися стандартними.

Для попередньої валідації докінг-протоколу молекулу Pico145 відтворювали у референтній структурі TRPC1/TRPC4:Pico145 (PDB ID: 8WPM) після видалення оригінального ліганду; кількість точок сітки вздовж осей x, y та z становила  $36 \times 40 \times 44$ , координати центру грід-боксу:  $x = 150,745$ ;  $y = 162,433$ ;  $z = 168,913$ . Для докінгу Pico145 у структуру 5Z96 параметри грід-боксу були такими:  $36 \times 38 \times 44$ , координати центру:  $x = 146,606$ ;  $y = 172,504$ ;  $z = 184,846$ . Для докінгу GFB-887 у структуру 5Z96 застосовували грід-бокс  $36 \times 34 \times 50$  із центром у точці:  $x = 189,849$ ;  $y = 151,628$ ;  $z = 172,226$ . Для розрахунку потенціалу використовували емпіричну скорингову функцію AutoDock 4. Отримані параметри сітки зберігали у файлі формату GPF.

### 2.2.3. Аналіз результатів молекулярного докінгу

За результатами кожної процедури докінгу отримували п'ять найбільш енергетично сприятливих конформацій ліганду. Фінальну позу обирали на основі комбінованої оцінки, що включала мінімальну прогнозовану енергію зв'язування та просторову відповідність пози ліганду референтній структурі (human TRPC1/TRPC4:Pico145 для Pico145 та zebrafish TRPC4:GFB-8438 для GFB-887). Візуалізацію отриманих комплексів та аналіз міжмолекулярних взаємодій (водневі зв'язки,  $\pi$ - $\pi$  stacking,  $\pi$ -alkyl,  $\pi$ - $\sigma$ ,  $\pi$ -катионні та  $\pi$ -аніонні взаємодії, ван-дер-ваальсові контакти) виконували у програмі BIOVIA Discovery Studio 4.5.

### 2.3. Моделювання молекулярної динаміки

Для незалежної перевірки стабільності комплексів, отриманих внаслідок молекулярного докінгу в умовах, наближених до фізіологічних, було проведено повноатомне моделювання молекулярної динаміки двох білок-лігандних комплексів – TRPC4:Pico145 та TRPC4:GFB-887 у мембранному середовищі. Обидві системи готували та симулювали за єдиним протоколом; відмінності стосувалися лише складу конкретної модельної системи і наведені окремо у підрозділі 2.3.3.

#### 2.3.1. Веб-сервіс для генерації вхідних даних CHARMM-GUI

Для уніфікованої підготовки вхідних файлів моделювання молекулярної динаміки використовували веб-платформу CHARMM-GUI [104]. Сервіс забезпечує графічний інтерфейс для генерації файлів координат, топологій та параметрів у форматах, сумісних із поширеними МД-пакетами, та

автоматизує типові підготовчі операції. Застосування CHARMM-GUI дозволяє мінімізувати технічні помилки на етапі побудови складних гетерогенних систем, зокрема мембранних білків у ліпідних бішарах, та забезпечує контроль якості розрахунків за загальноприйнятими стандартами.

### **2.3.2. Створення мембрани за допомогою Membrane Builder**

Мембранне оточення трансмембранного домену TRPC4 формували за допомогою модуля Membrane Builder платформи CHARMM-GUI [92], який автоматизує такі підготовчі операції, як визначення розміру системи, генерацію ліпідного бішару, водної фази та іонів, а також інтеграцію цих компонентів навколо білка. Для формування бішару застосовували метод заміщення, який передбачає попереднє розміщення псевдоатомів, що імітують ліпідні голівки, з подальшим почерговим заміщенням їх ліпідними молекулами з бібліотеки; цей метод забезпечує правильне упакування ліпідів навколо білка та включає автоматичну перевірку проникнення ліпідних хвостів у циклічні структури білка.

Ліпідне оточення відтворювали багатокомпонентним бішаром, склад якого моделює плазматичну мембрану ссавців. Основним матриксним фосфоліпідом обрано 1-пальмітоїл-2-олеоїл-sn-гліцеро-3-фосфохолін (POPC), який завдяки поєднанню одного насиченого та одного ненасиченого ацильного ланцюга відтворює характерні фізичні властивості плазматичної мембрани і є стандартним вибором для МД-моделювання мембранних білків [93]. Додатково до бішару введено 1-пальмітоїл-2-олеоїл-sn-гліцеро-3-фосфоетаноламін (POPE) та 1-пальмітоїл-2-олеоїл-sn-гліцеро-3-фосфо-L-серин (POPS) як фізіологічно значущі цвіттерійний і аніонний компоненти, а також холестерол, що моделює впорядкованість і механічні властивості бішару.

### 2.3.3. Створення системи для моделювання молекулярної динаміки

Перед збиранням повної МД-системи, виконували підготовку структури тетрамера TRPC4 миші. Кріо-ЕМ структура 5Z96 [25] не містить низки петельних ділянок, не розв'язаних експериментально через їхню локальну гнучкість. У межах цієї роботи наступні петлі кожної з чотирьох субодиниць тетрамера відновлювали шляхом кастомної пересадки відповідних фрагментів із моделі TRPC4, передбаченої за допомогою AlphaFold [105]: Met1–Tyr14, Lys119–Asp133, Arg175–Val187, Asp274–Asn284, Arg389–Leu392, Ala461–Leu462 та Pro661–Thr691. Пересадку виконували за допомогою структурного вирівнювання та ручного редагування структури у програмах BIOVIA Discovery Studio 4.5 та PyMOL [106]. Отриману повнорозмірну модель тетрамера TRPC4 використовували як вхідну структуру для всіх подальших етапів моделювання.

Для моделювання білок-лігандних комплексів вдоводана поза кожного ліганду переносилась із відповідного ланцюга, у якому виконувався докінг, у еквівалентну позицію кожної із чотирьох субодиниць повнорозмірного тетрамера шляхом структурного вирівнювання докованого ланцюга з відповідним ланцюгом тетрамерної моделі у програмах BIOVIA Discovery Studio 4.5 та PyMOL [106]. Таким чином у межах тетрамерної моделі TRPC4 відтворювалась структурна еквівалентність сайтів зв'язування у всіх чотирьох субодиницях.

Підготовлена повнорозмірна модель тетрамера TRPC4 у комплексі з лігандом завантажувалась у Membrane Builder CHARMM-GUI. Орієнтацію білка відносно мембрани визначали автоматичним розміщенням трансмембранного домену у площині бішару. Після цього систему доповнювали ліпідним бішаром, фазою води та іонами, що нейтралізують сумарний заряд системи та відтворюють фізіологічну іонну силу; іони розміщували методом Монте-Карло. Фосфоліпідні компоненти бішару розподілялися практично симетрично між верхнім та нижнім листками.

Окрім фосфоліпідів, що формують бішар, до складу обох систем було включено молекули холестерил-гемісукцинату, а до системи TRPC4:GFB-887 – також молекули фосфатидної кислоти (PA); ці ліпіди було розв’язано у вихідній кріо-ЕМ структурі 5Z96 (по одній молекулі кожного типу на субодиницю) і збережено при побудові моделі для відтворення нативного ліпідного оточення трансмембранного домену. Кількісний склад обох модельних систем наведено у табл. 2.1.

Таблиця 2.1.

Склад модельних систем для моделювання молекулярної динаміки комплексів TRPC4 з інгібіторами

| Компонент системи                  | TRPC4:Pico145       | TRPC4:GFB-887       |
|------------------------------------|---------------------|---------------------|
| Субодиниці тетрамера TRPC4         | 4                   | 4                   |
| Копії ліганду                      | 4                   | 4                   |
| POPC                               | 570                 | 560                 |
| POPE                               | 66                  | 140                 |
| POPS                               | 33                  | 56                  |
| Холестерол                         | 66                  | 84                  |
| Холестерил-гемісукцинат            | 4                   | 4                   |
| Фосфатидна кислота                 | –                   | 4                   |
| Молекули води                      | 127 560             | 140 556             |
| Іони Na <sup>+</sup>               | 374                 | 416                 |
| Іони Cl <sup>-</sup>               | 349                 | 364                 |
| Концентрація солі, M               | ≈0,15               | ≈0,15               |
| Розміри комірки, нм                | ≈16,4 × 16,4 × 19,0 | ≈17,2 × 17,2 × 19,0 |
| Тривалість продукційного етапу, нс | ≈215                | 200                 |

Розміри комірки наведено для стану системи на початок продукційного етапу.

В обох системах надлишок катіонів над аніонами компенсує сумарний від’ємний заряд білка та аніонних ліпідів POPS, забезпечуючи

електронейтральність комірки моделювання. Водні шари з обох боків бішару відокремлювали мембрану від її періодичних образів уздовж осі, нормальної до площини бішару.

#### **2.3.4. Силове поле CHARMM36m**

Для опису потенційної енергії білка, ліпідів, розчинника та іонів застосовували силове поле CHARMM36m [91]. CHARMM36m є оптимізованою версією CHARMM36, розробленою для коректного відтворення конформаційних властивостей не лише структурованих, але й внутрішньо неупорядкованих білків, що є важливим для адекватного моделювання гнучких цитозольних ділянок TRPC4. Поле забезпечує узгоджені параметри для взаємодій у системах білок–ліпід–вода, що є критичним для моделювання мембранних білків, та підтверджено довготривалими МД-симуляціями як здатне зберігати стабільність структурованих ділянок і коректно описувати поведінку неупорядкованих фрагментів.

#### **2.3.5. Програма для моделювання молекулярної динаміки GROMACS**

Розрахунки молекулярної динаміки виконували у програмному пакеті GROMACS версії 2024.3 [93] – одному з найпоширеніших інструментів для МД-симуляцій біомолекулярних систем. GROMACS забезпечує високоефективне інтегрування рівнянь руху для систем з великою кількістю атомів та підтримує паралельні обчислення на багатоядерних процесорах і графічних прискорювачах. Основним обчислювальним модулем є `gmx mdrun`, який виконує мінімізацію енергії, екваїлібрування та власне МД-симуляцію;



вхідний бінарний файл запуску формується препроцесором `gmx grompp` на основі файлів топології, координат, параметрів інтегратора та індексів.

### 2.3.6. Типи файлів у GROMACS

У процесі підготовки та виконання МД-симуляцій використовували стандартний набір файлових форматів GROMACS. Файл топології (`.top`) містить опис типів молекул, їхньої кількості, ковалентних і фіксованих взаємодій, атомних типів, зв'язків, кутів та торсійних кутів. Файл координат стартової конфігурації (`.gro`) містить позиції, а за необхідності й швидкості атомів. Файл параметрів інтегратора (`.mdp`) задає налаштування МД-симуляції (температура, тиск, кількість кроків, тип термостата та баростата, алгоритми обчислення електростатичних і ван-дер-ваальсових взаємодій). Файл індексів (`.ndx`) задає групування атомів для застосування обмежень або окремих термостатів. Єдиний бінарний вхідний файл запуску (`.tpr`) формується модулем `gmx grompp` комбінуванням файлів `.top`, `.gro`, `.mdp` та `.ndx`. Результуючі файли симуляції: файл журналу з подіями симуляції (`.log`), файли траєкторій – повноточний (`.trr`) та стиснутий (`.xtc`), файл енергій системи (`.edr`) та файл контрольної точки для відновлення симуляції (`.cpt`).

### 2.3.7. Моделювання молекулярної динаміки у програмі GROMACS

МД-моделювання виконували з використанням вхідних файлів, згенерованих CHARMM-GUI, за послідовністю команд, рекомендованою цим сервісом. Загальна схема включала три функціональні етапи.

На першому етапі проводили мінімізацію енергії стартової конфігурації системи методом найшвидшого спуску (*steepest descent*) з метою усунення стеричних конфліктів та локальних надлишків енергії.

На другому етапі виконували багатоступеневе еквілібрування системи, що складалося з шести послідовних стадій із поступовим послабленням позиційних обмежень, накладених на основний ланцюг білка, бічні ланцюги та ліпіди, а також дихедральних обмежень. Часовий крок інтегрування на перших трьох стадіях становив 1 фс, на трьох наступних – 2 фс. Така схема дозволяла поступово узгодити кінетичну та потенційну енергії та вивести систему у стабільний стан, придатний для продуктивної симуляції.

На третьому етапі проводили продуктивну МД-фазу в ізобарно-ізотермічному ансамблі (NPT, з фіксованими числом частинок, тиском і температурою), який обраний як такий, що найкраще відповідає фізіологічним умовам у клітині. Часовий крок інтегрування продукційного етапу становив 2 фс.

Температуру системи (303,15 K) підтримували термостатом стохастичного перемасштабування швидкостей (velocity-rescale) зі сталою зв'язку 1 пс, застосованим окремо до груп розчиненої речовини, мембрани та розчинника. Тиск (1 бар) підтримували баростатом стохастичного перемасштабування комірки (C-rescale) у напівізотропному режимі зі сталою зв'язку 5 пс, що є стандартним вибором для мембранних систем. Ковалентні зв'язки за участю атомів водню фіксували алгоритмом LINCS. Дальні електростатичні взаємодії обчислювали методом частинково-сіткового Евальда (PME); радіус обрізання електростатичних та ван-дер-ваальсових взаємодій у прямому просторі становив 1,2 нм, причому ван-дер-ваальсові взаємодії згладжували функцією зсуву сили (force-switch), починаючи з 1,0 нм.

Тривалість продукційного етапу, відібраного для аналізу, становила приблизно 215 нс для комплексу TRPC4:Pico145 та 200 нс для комплексу TRPC4:GFB-887.

### 2.3.8. Аналіз та візуалізація результатів молекулярної динаміки

Аналіз отриманих траєкторій виконували за допомогою стандартних аналітичних утиліт GROMACS: кореневе середньоквадратичне відхилення (RMSD) положень атомів білка та ліганду від стартової конформації обчислювали за допомогою `gmx rms` для оцінки загальної стабільності системи; параметри стану системи – потенційну, кінетичну та повну енергію, температуру і тиск, за допомогою `gmx energy` для контролю врівноваження та стабільності NPT-ансамблю протягом траєкторії; кількість водневих зв'язків між лігандом і білком у сайті зв'язування – за допомогою `gmx hbond`. Тим самим інструментом отримували також часові профілі числа донорних і акцепторних атомів та розподіли відстаней донор–акцептор і кутів H–D–A. Водневий зв'язок реєстрували за геометричним критерієм: відстань донор–акцептор  $\leq 0,35$  нм і кут H–D–A  $\leq 30^\circ$  (тобто відхилення D–H...A від лінійності  $\leq 30^\circ$ ).

Оскільки модельна система містила чотири еквівалентні копії ліганду, відповідні параметри (RMSD, кількість водневих зв'язків, число донорів/акцепторів, розподіли відстаней і кутів) розраховували для кожної копії окремо, а на графіках наводили усереднене по чотирьох копіях значення; затінена ділянка навколо усередненої кривої відображає стандартне відхилення між копіями. Оскільки чотири копії ліганду перебувають у межах однієї білкової структури і не є статистично незалежними, цей розкид слід розглядати як описову характеристику варіабельності між копіями, а не як стандартну похибку незалежних повторів. Кадри для аналізу зберігали кожні 100 пс; для згладжування високочастотних флуктуацій часових профілів застосовували ковзне середнє з вікном 50 кадрів (5 нс).

Побудову графіків здійснювали засобами бібліотеки `matplotlib` (Python), візуалізацію та рендеринг молекулярних структур – у програмі VMD з використанням рендерера Tachyon.

## 2.4. Прогнозування ADMET-параметрів

Прогнозування ADMET-параметрів для досліджуваних сполук (Pico145 та GFB-887) виконували із застосуванням власної моделі машинного навчання, яку детально описано у нашому препринті та розділі 5 цієї дисертації [17]. Модель базується на поєднанні ембедингів Mol2Vec [99] із традиційними фізико-хімічними молекулярними дескрипторами та натренована на експериментальних ADMET-наборах даних платформи Therapeutics Data Commons [100]. У межах цього розділу прогнози для Pico145 та GFB-887 отримували із застосуванням стандартних налаштувань моделі, без додаткового перенавчання. Детальний опис архітектури моделі, набору дескрипторів, процедури навчання, та повний перелік прогнозованих ADMET-параметрів наведено у Розділі 5.

Для оцінювання якості моделей використовували метрики, передбачені відповідними бенчмарками TDC. Для регресійних задач застосовували середню абсолютну похибку (MAE) та коефіцієнт рангової кореляції Спірмена, для задач бінарної класифікації – площу під ROC-кривою (ROC-AUC) та площу під кривою «precision–recall» (PR-AUC); вибір конкретної метрики для кожного параметру відповідає стандарту лідерборда TDC. Детальні числові значення метрик для всіх 16 задач наведено в оригінальній публікації [17].

## 2.5. Експериментальне дослідження впливу Pico145 та GFB-887 на $mI_{CAT}$ і скоротливу активність гладеньких м'язів кишечника

### 2.5.1. Експериментальні тварини

Експериментальні дослідження виконували на мишах-самцях лінії Albino B6, яких утримували у стандартних умовах віварію Інституту

фізіології ім. О. О. Богомольця НАН України. Для експериментів із реєстрації трансмембранних струмів методом петч-клемп (досліди з Pico145) використовували тварин віком 3 місяці з масою тіла 25–30 г. Для експериментів з ізометричної тензометрії (досліди з GFB-887) використовували мишей віком 3 місяці з масою тіла 30–35 г.

Усі процедури здійснювали з дотриманням положень Європейської конвенції щодо захисту хребетних тварин, що використовуються для дослідних та інших наукових цілей (Страсбург, 1986) [107], а також Закону України «Про захист тварин від жорстокого поводження» від 21.02.2006 № 3447-IV [108].

### **2.5.2. Матеріали і реактиви**

Неорганічні солі, буферні сполуки та ферменти (колагеназа типу IA, інгібітор трипсину, бичачий сироватковий альбумін), а також карбахол та інші стандартні реагенти виробництва Sigma-Aldrich (США). Сполуки GFB-887 та Pico145 люб'язно надали колеги з Університету Лідса (Сполучене Королівство).

У експериментах застосовували такі розчини.

Розчин Кребса для ферментативного виділення клітин (ммоль/л): NaCl – 120; глюкоза – 12; HEPES – 10; KCl – 6; pH 7,4 (корекція NaOH), з додаванням колагенази типу IA (1 мг/мл), інгібітора трипсину (1 мг/мл) та BSA (1 мг/мл).

Модифікований розчин Кребса для первинного вилучення тонкої кишки під час тензометричного дослідження (ммоль/л): NaCl – 120; глюкоза – 12; HEPES – 10; KCl – 6; CaCl<sub>2</sub> – 2,5; MgCl<sub>2</sub> – 1,2; pH 7,4 (корекція NaOH).

Розчин Кребса–Рінгера для перфузії гладеньком'язових препаратів у тензометричних експериментах (ммоль/л): NaCl – 133; NaHCO<sub>3</sub> – 16,3;

$\text{NaH}_2\text{PO}_4$  – 1,38;  $\text{KCl}$  – 4,7;  $\text{MgCl}_2$  – 1,05; глюкоза – 11,5;  $\text{CaCl}_2$  – 2,73; HEPES – 10; pH 7,4 (корекція  $\text{NaOH}$ ).

Зовнішній цезієвмісний розчин для реєстрації трансмембранних струмів методом петч-клемп (ммоль/л):  $\text{CsCl}$  – 120; глюкоза – 12; HEPES – 10; pH 7,4 (корекція  $\text{CsOH}$ ).

Внутрішньопіпеточний розчин (ммоль/л):  $\text{CsCl}$  – 80;  $\text{MgATP}$  – 1; креатин – 5; глюкоза – 5; BAPTA (1,2-bis(o-aminophenoxy)ethane-N,N,N',N'-tetraacetic acid) – 10; HEPES – 10;  $\text{CaCl}_2$  – 4,6; pH 7,4 (корекція  $\text{CsOH}$ ). Концентрація вільного  $[\text{Ca}^{2+}]$  у піпеточному розчині стабілізована на рівні приблизно 100 нмоль/л буферною системою BAPTA/ $\text{CaCl}_2$ .

### **2.5.3. Виділення ізольованих клітин гладеньких м'язів тонкого кишечника**

Виділення ізольованих міоцитів тонкого кишечника виконували під бінокулярним мікроскопом. Тканину тонкої кишки розсікали на фрагменти шириною приблизно 1 мм, які піддавали ферментативній обробці у розчині Кребса з додаванням колагенази типу IA (1 мг/мл), інгібітора трипсину (1 мг/мл) та BSA (1 мг/мл). Інкубацію здійснювали протягом 17 хв за температури 36 °C. По завершенні інкубації тканину промивали, міоцити ізолювали шляхом механічної тритурації, після чого аліквоту (0,5 мл) суспензії клітин наносили на покривні скельця у розчині Кребса.

### **2.5.4. Реєстрація трансмембранних струмів**

Реєстрацію іонних струмів через плазматичну мембрану ізольованих міоцитів виконували методом петч-клемп у конфігурації whole-cell з

використанням підсилювача AxoPatch 200B та програмного комплексу pClamp 8 (Molecular Devices, США).

Для селективної реєстрації мускарин-індукованого катіонного струму застосовували симетричні цезієві розчини (склад наведено у підрозділі 2.5.2), а внутрішньоклітинну концентрацію вільних іонів кальцію стабілізували на рівні приблизно 100 нмоль/л за допомогою буферної системи ВАРТА/ $\text{CaCl}_2$  у внутрішньопіпеточному розчині [2]. Усі реєстрації виконували за кімнатної температури (20–22 °C). Активацію  $mI_{\text{CAT}}$  індукували аплікацією карбахолу у концентрації 50 мкмоль/л, що забезпечувало одночасну стимуляцію мускаринових рецепторів підтипів  $M_2$  та  $M_3$ .

Мембранний потенціал утримували на рівні –40 мВ. Для побудови стаціонарної вольт-амперної характеристики застосовували ramp-протокол з безперервною зміною потенціалу у діапазоні від 80 до –120 мВ тривалістю 6 с, який повторювали з інтервалом 30 с. Записані струми фільтрували при частоті зрізу 2 кГц та оцифровували з частотою дискретизації 10 кГц.

Досліджувані концентрації Pico145 (1 та 100 пмоль/л) додавали до зовнішнього розчину на фоні активованого карбахолом  $mI_{\text{CAT}}$ ; ефект блокатора оцінювали за змінами амплітуди струму при потенціалах –40 мВ, –120 мВ та 80 мВ.

#### **2.5.5. Реєстрація скоротливої активності гладеньких м'язів кишечника**

Скоротливі реакції гладеньких м'язів кишечника оцінювали методом тензометричної реєстрації в ізометричному режимі із застосуванням ємнісних датчиків напруження та програмного забезпечення LabScribe3 (World Precision Instruments Inc., США).

Попередньо у миші виконували лапаротомію, після чого ділянку тонкої кишки швидко вилучали та негайно переносили у модифікований розчин

Кребса. Вилучений кишечник розрізали на сегменти довжиною 1 см, які переносили у камеру, заповнену розчином Кребса–Рінгера. Препарати фіксували між двома сталевими гачками: один гачок закріплювали стаціонарно, інший з'єднували з тензодатчиком. Для забезпечення оптимальної сили скорочення сегменти піддавали попередньому пасивному розтягненню силою приблизно 0,5 г, після чого тензодатчик встановлювали у нульове положення. Перфузію камери розчином Кребса–Рінгера, підігрітим до 37 °C, здійснювали зі сталою швидкістю 1,5 мл/хв за допомогою чотириканального перистальтичного насоса IPS ISM 930 (Ismatec, Німеччина). Перед початком реєстрації препарати адаптували протягом 60 хв у перфузійному розчині.

Збереження функціональної активності препарату перед застосуванням агоніста мускаринових рецепторів карбахолу підтверджували тестом з гіперкалієвим розчином ( $K^+$  60 ммоль/л), який викликав виражене скорочення та свідчив про фізіологічну інтактність тканини. Після відмивання гіперкалієвого розчину аплікація карбахолу (50 мкмоль/л) спричиняла пікову фазу скоротливу відповідь із подальшим тонічним скороченням (плато). Після контрольного тесту з карбахолом препарат відмивали у розчині Кребса–Рінгера протягом 20 хв, після чого до перфузійного середовища додавали GFB-887 у концентрації 10 мкмоль/л. Після 20-хвилинної інкубації у розчині з GFB-887 повторно аплікували карбахол у тій самій концентрації (50 мкмоль/л), що й у контрольному тесті. Скоротливу відповідь у присутності блокатора порівнювали з контрольною у тому самому препараті.

#### **2.5.6. Обробка та аналіз даних**

Аналіз експериментальних результатів, статистичну обробку та графічну візуалізацію даних електрофізіологічних експериментів виконували із застосуванням програмних пакетів pClamp версії 9.2 (Molecular Devices,



США) та Origin версії 8.5 (OriginLab Corporation, США). Обробку записів тензометричних експериментів та побудову відповідних графіків здійснювали у ПЗ LabScribe3 (World Precision Instruments Inc., США) та Origin 8.5 (OriginLab Corporation, США).

Для характеристики кінетики блокади  $mI_{\text{CAT}}$  постійну часу розвитку інгібування ( $\tau$ ) визначали шляхом апроксимації часового ходу спаду амплітуди струму моноекспоненційною функцією виду

$$I(t) = I_0 \cdot e^{(-t/\tau)},$$

де  $I(t)$  – амплітуда струму в момент часу  $t$ ;

$I_0$  – початкова амплітуда;

$\tau$  – постійна часу.

Оскільки в окремих клітинах постійна часу характеризувалася значною варіабельністю, апроксимацію виконували на усереднених нормалізованих значеннях струму; отримане значення  $\tau = 140 \pm 6$  с ( $n = 9$ ) відповідає результату підгонки усередненої кривої, де  $\pm 6$  с – стандартна похибка апроксимації.

Для тензометричних експериментів амплітуду карбахол-індукованого скорочення оцінювали як силу скорочення (у грамах) у контрольних умовах та на фоні дії GFB-887. Ступінь пригнічення скоротливої відповіді розраховували як відсоток від контрольного рівня, який приймали за 100%. Оскільки контрольну та експериментальну відповіді реєстрували на тому самому препараті, для оцінки статистичної значущості відмінностей застосовували парний  $t$ -критерій Стьюдента. Відмінності вважали статистично значущими за рівня значущості  $p < 0,05$ . Кількісні результати наведено як середнє значення  $\pm$  стандартна похибка середнього ( $\text{mean} \pm \text{SEM}$ );  $n$  позначає кількість незалежних експериментів. Статистичну обробку тензометричних даних виконували у ПЗ Origin 8.5 (OriginLab Corporation, США).

### РОЗДІЛ 3

## ІДЕНТИФІКАЦІЯ ТА ПРІОРИТИЗАЦІЯ САЙТІВ ЗВ'ЯЗУВАННЯ

### TRPC4

Достовірна ідентифікація та пріоритизація сайтів зв'язування на поверхні білкових мішеней є важливим кроком у раціональному дизайні лікарських засобів. Попри широкий арсенал обчислювальних методів виявлення кишень, усі вони демонструють спільну тенденцію – передбачають істотно більше потенційних кишень, ніж це може бути підтверджено експериментальними даними. Внаслідок цього дослідник, що ставить завдання докінгу у відомий сайт зв'язування певної мішені, стикається з надлишковим набором частково перекритих передбачень, серед яких необхідно обрати один чи кілька біологічно релевантних прогнозів і відкинути або об'єднати решту. Це супроводжується паралельним аналізом літературних джерел щодо об'єкта дослідження, що є ручним та часозатратним процесом. Окрім того, варто підкреслити, що геометричні межі сайтів зв'язування рідко явно анотовано як у публічних базах структурних даних, так і в наукових публікаціях, тоді як правильне визначення таких меж є необхідною передумовою коректного налаштування процедури докінгу.

На практиці єдиним способом обрати серед передбачуваних кишень ту, що відповідає експериментально підтвердженому сайту зв'язування, залишається ретельний аналіз наукової літератури з подальшим зіставленням описаних амінокислотних залишків із конкретною структурою PDB. Такий ручний етап має три принципові обмеження. По-перше, він є трудомістким і часто є вузьким місцем в процесі пошуку ліків. По-друге, термінологія та спосіб подання даних про кишені суттєво відрізняються між публікаціями, що збільшує ризик помилок при зведенні інформації з кількох джерел. По-третє, опис сайту у статті майже ніколи не містить готового об'ємного представлення – для докінгу його необхідно реконструювати окремо на

основі списку залишків і геометрії білка. Отже, методологічною прогалиною, яка безпосередньо мотивувала розробку описаного нижче методу, є відсутність автоматизованого конвеєра (pipeline), який поєднував би геометричне виявлення кишень з літературно-керованим відбором біологічно релевантних серед них.

### 3.1. Загальна схема гібридного алгоритму та формування бенчмарк-датасету

Запропонований у цій роботі метод є гібридним і об'єднує два незалежні джерела інформації про сайти зв'язування: геометричне виявлення кандидатних кишень на основі структури білка з PDB та отримання інформації про експериментально підтверджені сайти з наукових публікацій за допомогою BMM. Загальну архітектуру розробленого конвеєра подано на рис. 3.1.

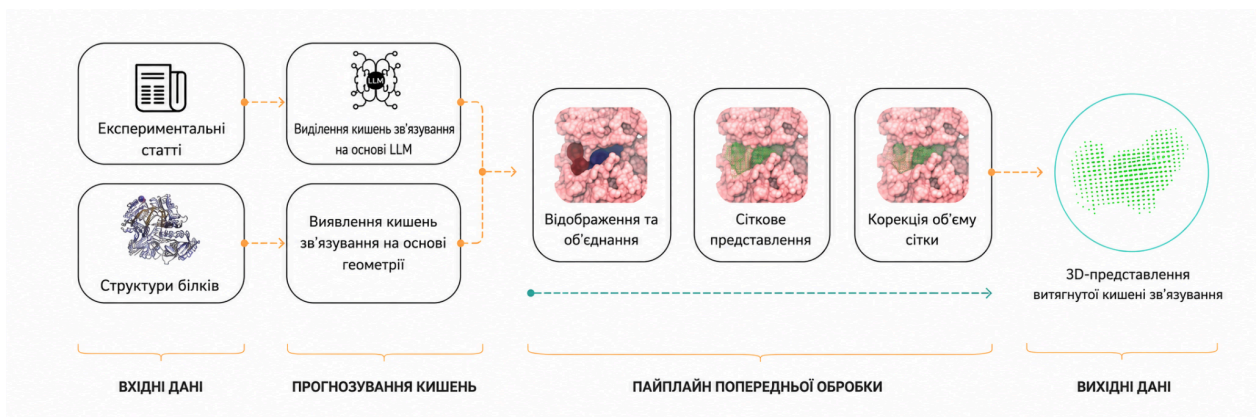


Рис. 3.1. Загальна схема запропонованого гібридного конвеєра виявлення і пріоритизації сайтів зв'язування [14].

На вході конвеєр приймає тривимірну структуру білка та набір наукових статей. Геометричний модуль генерує множину потенційних кишень на основі аналізу поверхні структури, водночас текстовий модуль на

основі ВММ ідентифікує у літературі описи малих молекул, що зв'язуються з мішенню, і витягує списки амінокислотних залишків, які формують відповідні сайти. Результати двох модулів об'єднуються на подальших етапах: отримані списки залишків використовуються, як фільтр для відбору геометрично передбачених кишень, після чого обрані кишені переводяться у об'ємне сіткове представлення та проходять низку кроків постобробки.

Концептуально даний підхід співвідноситься з наявними базами біологічно релевантних кишень, зокрема BioLip2 та Q-BioLip [109, 110]. Ці бази надають зручний інтерфейс для огляду сайтів зв'язування у структурах PDB, проте визначають склад кишені за відстанню між амінокислотними залишками білка та атомами ко-кристалізованого ліганда. Такий критерій має два суттєві обмеження. По-перше, він непридатний до структур без ко-кристалізованого ліганда, частка яких у практичних задачах пошуку ліків залишається значною. По-друге, критерій відстані до ліганда не враховує перехідних динамічних контактів, а також непрямих свідчень про склад кишені, що надходять із сайт-спрямованого мутагенезу чи флуоресцентної спектроскопії. Розроблений нами метод позбавлений обох обмежень: він збирає інформацію про залишки кишені безпосередньо з літератури незалежно від наявності ко-кристалізованого ліганда і здатний інтегрувати дані, отримані різнорідними експериментальними техніками, а також агрегувати описи одного і того ж сайту з кількох незалежних публікацій для різних за розміром і хімічним класом лігандів [14].

Для систематичної оцінки якості роботи запропонованого методу було сформовано публічно доступний бенчмарк-датасет, адаптований саме до задачі отримання інформації про сайти зв'язування з літератури. До датасету увійшло 10 різнотипних білкових мішеней: каталітична субодиниця ДНК-полімерази  $\alpha$ , тирозинкіназа ABL1, серотоніновий рецептор 5-HT<sub>2A</sub>, мускариновий ацетилхоліновий рецептор M<sub>2</sub>, людський потенціалзалежний натрієвий канал Nav1.7, натрієвий канал американського таргана, рецептор  $\gamma$ -аміномасляної кислоти типу GABA<sub>A</sub>, ГТФаза KRas,

дигідрооротатдегідрогеназа та mixed lineage kinase domain-like protein (MLKL) [14]. Для кожної мішені додатково було підібрано від двох до п'яти рецензованих наукових публікацій і відповідні тривимірні структури з PDB. Сумарно датасет охоплює 31 статтю.

Принциповим методологічним рішенням при формуванні датасету була імітація реального робочого сценарію користувача: публікації відбирали безпосередньо з топу результатів веб-пошуку за назвою мішені без накладання додаткових фільтрів за якістю, рівнем деталізації чи темою [14]. Це дозволяє оцінювати конвеєр у близьких до практичних умов, коли дослідник насправді не має змоги наперед відсіяти всі нерелевантні джерела. Відібрані публікації розподілено на три групи, що відображають типові комбінації інформації про сайти зв'язування, з якими стикається кінцевий користувач: перша група складається з 11 статей, що не містять опису сайтів зв'язування взагалі й виступають негативним контролем для виявлення хибнопозитивних передбачень; друга група це 17 статей описують один сайт зв'язування для своєї мішені – це найпростіший рівень, що демонструє базову здатність BMM розуміти структурно-біологічні концепції; в третю групу увійшли 3 статті, які описують кілька сайтів у межах однієї мішені й дозволяють оцінити здатність моделі точно розподіляти конкретні залишки між окремими кишнями [14].

Фінальний бенчмарк-датасет було збережено у вигляді набору файлів JavaScript Object Notation (JSON), по одному на кожну публікацію. Кожен файл містить анотацію сайтів зв'язування малих молекул, прямо описаних у тексті статті, у формі структурованих записів, що включають назву сайту, його текстовий опис і список амінокислотних залишків, які його формують. Для попередньої конвертації PDF-файлів статей у формат, який буде сприйматись BMM, було використано інструмент GROBID, що забезпечує високу якість виділення структури документа, а саме заголовків, розділів, підписів до рисунків і таблиць зі збереженням порядку тексту [111]. Незначне ручне редагування застосовувалося лише для уніфікації форматування

заголовків і видалення нерелевантних для задачі секцій, таких як подяки чи авторські внески. Розроблений бенчмарк-датасет та код оприлюднено у відкритому репозиторії GitHub та в архіві Zenodo [15, 16].

### 3.2. Геометричне виявлення кандидатних кишень зв'язування

Геометричний модуль конвеєра відповідає за генерацію первинного набору кишень-кандидатів без використання жодної інформації з літератури. Задачу сформульовано як локалізацію порожнин на поверхні білка з подальшою геометричною характеристикою кожної з них. Як геометричний метод для виявлення кишень, у цій роботі використано відкритий програмний пакет Focket, що є одним з найбільш поширених інструментів для вирішення даної задачі [77]. Підстави вибору такі: відкритий вихідний код і відсутність обмежень щодо використання, мінімальна залежність від параметрів, відсутність потреби у наперед підготовлених навчальних вибірках, що робить Focket універсальним щодо будь-якого білка незалежно від його родини чи доступності структурних гомологів.

Принцип роботи Focket базується на представленні форми порожнин у білковій структурі через альфа-сфери, сформульовані у роботі Liang *et al.* [78]. Альфа-сфера – це сфера, яка дотична до чотирьох атомів білка і не містить у своєму об'ємі жодного іншого атома. За визначенням, такі сфери заповнюють простір між атомами білка, а їхній радіус залежить від локальної геометрії: малі альфа-сфери описують щільно упаковані області всередині білка, дуже великі – зовнішню поверхню, тоді як альфа-сфери проміжного діапазону радіусів локалізовані саме у порожнинах, придатних для розміщення ліганда. Для обчислення альфа-сфер Focket застосовує тесселяцію Вороного за координатами важких атомів білка [77, 78]. Отримана діаграма задає топологічні відносини між атомами та визначає потенційні положення центрів альфа-сфер. Після фільтрації за радіусом альфа-сфери

кластеризуються за геометричною близькістю, і кожен утворений кластер трактується як окрема кишень-кандидат [14].

Результатом роботи геометричного модуля на одному білку є набір передбачених кишень, де кожна кишень представлена сукупністю альфа-сфер з відомими координатами центрів та радіусами. Як уже зазначалося у вступі до розділу, цей набір, як правило, містить значно більше елементів, включно з тими, які насправді не мають біологічного значення. Fprocket ідентифікує практично будь-яку геометричну порожнину, придатну за розміром для розміщення малої молекули, незалежно від того, чи відповідає вона експериментально підтвердженому сайту. Саме тому літературно-керований етап конвеєра виконує роль фільтра: з геометрично передбаченого набору відбираються лише ті кишені, для яких у наукових публікаціях є пряма експериментальна підтримка [14].

### **3.3. Отримання інформації про сайти зв'язування із застосуванням великих мовних моделей**

#### **3.3.1. Триетапний конвеєр: фільтрація, екстракція, уточнення**

Текстовий модуль конвеєра реалізовано як послідовність з трьох незалежних викликів однієї і тієї ж ВММ, кожен з яких відповідає за вузько окреслену підзадачу: оцінку релевантності вхідної публікації для цільової мішені, отримання описів сайтів зв'язування та їхніх залишків та уточнення і корекцію результатів екстракції. Загальну логіку модуля відображено на рис. 3.2.

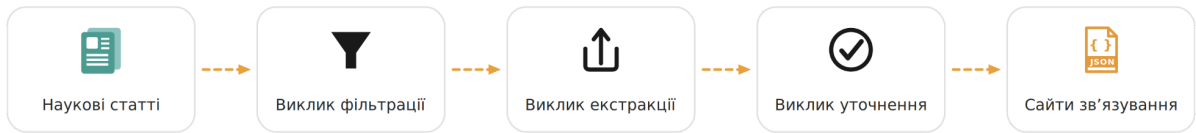


Рис. 3.2. Загальна схема BMM-модуля витягу інформації про сайти зв'язування з трьома послідовними етапами: фільтрацією, екстракцією та уточненням [14].

Свідомо відмовлено від використання архітектур з явним ланцюжком міркувань (CoT) чи деревом міркувань (ToT), попри їхню привабливість для складних багатокрокових задач [112, 113]. Обґрунтування цього вибору прагматичне: для застосування у високопродуктивних і обмежених по часу дослідженнях, критичними є час обробки однієї публікації і вартість запиту до моделі, а CoT- і ToT-архітектури суттєво підвищують обидва показники. Натомість застосовано пряме формування запитів з опорою на розгорнуті інструкції та структуровані приклади, що дозволяє отримувати стабільний за форматом результат при мінімальній кількості викликів моделі. Усі три етапи конвеєра виконуються однією і тією ж моделлю GPT-4o-mini. Повний текст статті пропускається через контекстне вікно (це максимальний обсяг тексту, який BMM може «бачити» і враховувати одночасно під час генерації відповіді) моделі цілком, оскільки об'єм сучасних публікацій у галузі структурної біології вміщується у контекстні вікна сучасних BMM [14].

На етапі фільтрації BMM отримує на вхід повний текст публікації і повинна дати бінарну відповідь щодо того, чи є ця публікація релевантною для пошуку сайтів зв'язування заданої мішені. Публікація вважається релевантною, якщо вона одночасно задовольняє дві умови: описує саме цільовий білок, а не його гомолог чи іншого представника тієї ж надродини; містить опис хоча б одного сайту зв'язування з роздільністю на рівні амінокислотних залишків, а не лише на рівні кишені чи домену. Позитивний результат фільтрації пропускає публікацію до наступного етапу, негативний –



виключає її з подальшої обробки і тим самим запобігає генерації хибнопозитивних кишень на підставі нерелевантного тексту [14].

Наступним етапом є екстракція під час якої ВММ здійснює автоматичне розпізнавання тексту публікації і формує структурований запис для кожного описаного сайту зв'язування, що включає назву кишені, її короткий функціональний опис і список залишків у заздалегідь визначеному форматі. Формат залишку обов'язково включає ідентифікатор ланцюга (chain ID), трибуквенне або однобуквенне позначення амінокислоти і номер залишку в послідовності. Саме на цьому кроці виконується інтерпретація стилістично різнорідних описів кишень з різних публікацій. Окремим завданням є узгодження нумерації залишків між текстом статті та відповідною структурою PDB [14].

На третьому етапі (уточнення) виконується постобробка результатів другого етапу з метою виправлення типових помилок екстракції. До них належать: пропуск окремих залишків, які згадано у тексті опосередковано, але які фактично належать до сайту; хибне розділення одного сайту на кілька записів, наприклад, коли опис сайту розподілений між кількома абзацами чи підзаголовками; некоректне включення до списку сайтів тих областей, що задіяні у білок-білкових взаємодіях (PPI) або у зв'язуванні ДНК чи РНК, які не є мішенями для малих молекул і тому мають бути виключені з результату [14]. Окрім змістових уточнень, на цьому етапі виконуються формальні перевірки: узгодженість ідентифікатора ланцюга, коду та номера залишку, а також коректність однобуквеного чи трибуквеного коду амінокислоти. Це забезпечує як змістову якість екстракції, так і придатність результатів до подальшого опрацювання.

### 3.3.2. Оптимізація системи запитів

Оптимізація запитів (промптів) для етапів фільтрації та екстракції проводилася ітеративно, шляхом послідовного введення окремих інструкцій і відстеження змін у метриках якості після кожного модифікованого варіанту. Зміна фіксувалася у фінальній версії промπτу лише за умови, що вона покращувала принаймні одну з цільових метрик або усувала системну помилку у форматуванні без погіршення інших показників [14]. Така послідовна стратегія дозволила зберегти прозорість внеску окремих інструкцій у підсумкову поведінку моделі і уникнути компромісних рішень, характерних для одночасної зміни кількох параметрів промπτу.

По-перше, до промπτу додано явні інструкції, що обмежують сферу зацікавленості моделі виключно сайтами зв'язування малих молекул, тим часом як області, задіяні у РРІ або у зв'язуванні ДНК чи РНК, повинні бути активно виключені з результату [14]. Ця інструкція виявилася найбільш ефективною для зниження частки хибнопозитивних кишень, оскільки у значній частині публікацій описи РРІ-інтерфейсів подаються у тому ж стилі, що й опис сайтів для малих молекул. По-друге, у промπτ введено обмеження, за яким до списку залишків кишені мають потрапляти лише ті з них, які у тексті статті прямо названі як такі, що беруть участь у формуванні кишені чи у взаємодії з лігандом. Це знизило частку залишків, які модель додавала за загальним контекстом абзацу, але не мали експериментальної прив'язки до сайту. По-третє, до запиту включено детальні вимоги до формату відповіді та явні інструкції щодо пріоритетів: модель орієнтується на повноту екстракції релевантних залишків і утримується від генерації неперевіраних або неточно записаних даних [14]. По-четверте, запроваджено суворі правила форматування запису залишку: обов'язкові поля ідентифікатора ланцюга, коду амінокислоти та номера залишку мають відповідати фіксованому шаблону. Будь-яке відхилення від цих правил виявляється на етапі валідації та виправляється окремим викликом моделі [14].

Повний текст фінальних версій промптів для етапів фільтрації, екстракції та уточнення наведено у додаткових матеріалах до публікації і через значний обсяг не дублюється у дисертації [14].

### **3.4. Алгоритмічне картування залишків та побудова об'ємного представлення**

#### **3.4.1. Картування залишків на структури PDB та кластеризація**

Списки залишків, витягнуті BMM із публікацій, не можуть бути безпосередньо картовані на конкретну структуру PDB без додаткового опрацювання. Списки залишків, витягнуті BMM із публікацій, не можуть бути безпосередньо зіставлені з конкретною структурою PDB без додаткового опрацювання – через низку добре відомих розбіжностей між науковою літературою та структурними даними. По-перше, структури у PDB нерідко містять інженерні мутації, вставки або делеції відносно послідовності, описаної у статті. По-друге, нумерація залишків у тексті може відповідати нумерації повнорозмірного білка, тоді як у PDB-файлі застосовується альтернативна нумерація (з перервами, константним зміщенням тощо). По-третє, у тексті статті часто не вказано, до якого ланцюга PDB-файлу належить залишок, особливо якщо мішень є гомомультимером. Окремим класом задач є ідентифікація міжсубодиничних кишень, що формуються на стиках кількох ланцюгів: їх автоматизоване розпізнавання засобами BMM-екстракції виявилось ненадійним і потребує окремого алгоритмічного кроку [14].

Для одночасного розв'язання всіх перелічених задач розроблено алгоритм картування залишків, загальну блок-схему якого подано на рис. 3.3.

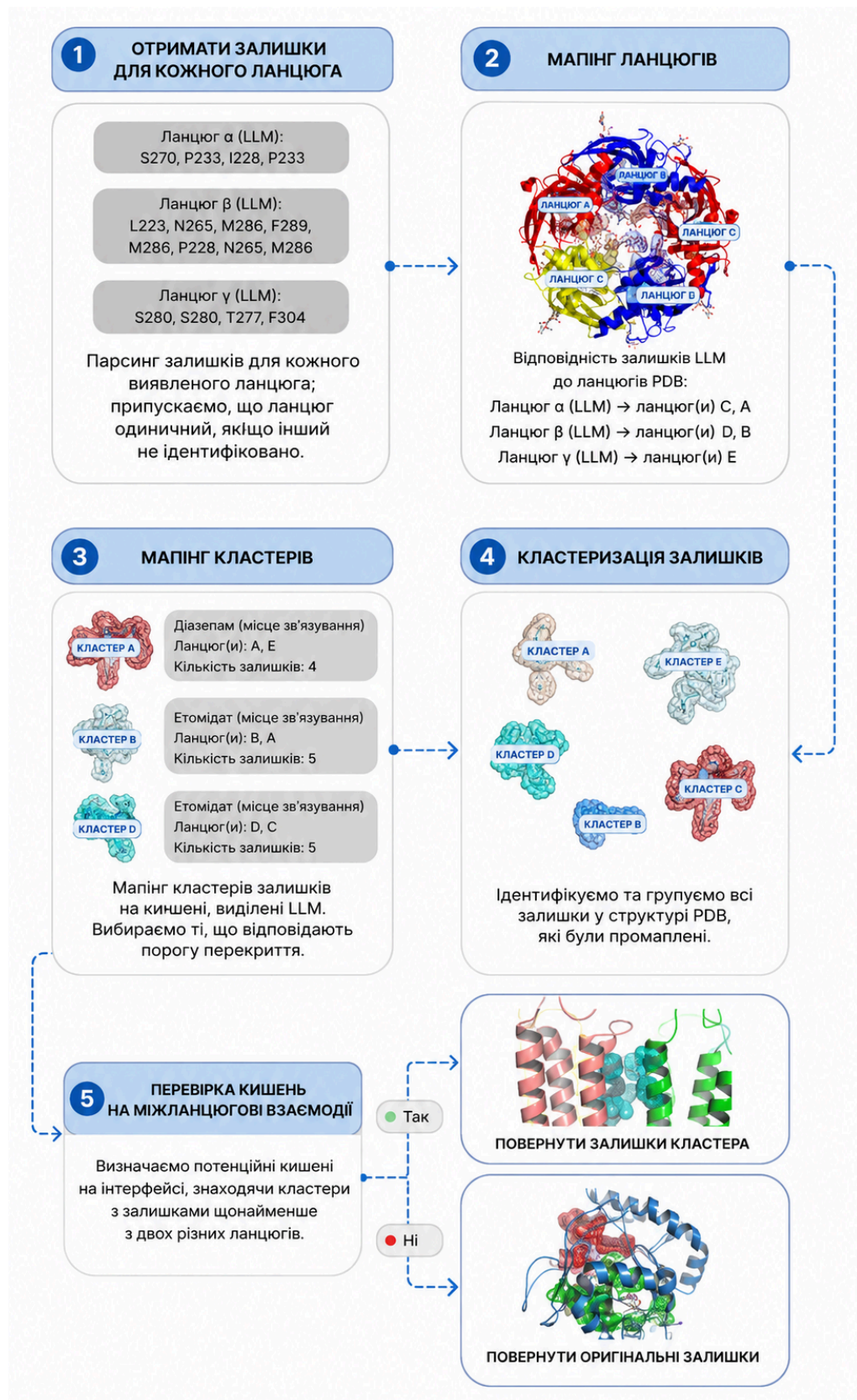


Рис. 3.3. Блок-схема алгоритму зіставлення залишків, отриманих за допомогою BMM, зі структурою PDB: визначення ланцюгової відповідності, просторова кластеризація залишків, зіставлення кластерів з BMM-кишнями

за критерієм перекриття та перевірка міжсубодиничної природи сайту зв'язування [14].

На першому кроці алгоритм зіставляє ланцюг, визначений BMM для групи залишків, з ланцюгами PDB-структури методом попарного вирівнювання послідовностей. Відповідність вважається встановленою, якщо індекс подібності послідовностей перевищує емпірично підібраний поріг 0,6 [14]. Такий помірний поріг обрано для забезпечення стійкості алгоритму до мутацій, делецій і розбіжностей у послідовності, характерних для інженерних конструктів у сучасних структурних публікаціях.

Другий крок реалізує просторову кластеризацію всіх залишків, успішно картованих на PDB-структуру, незалежно від того, до якої кишені їх віднесено під час екстракції за допомогою BMM. Для цього застосовано алгоритм MeanShift, ключова перевага якого це відсутність необхідності наперед задавати кількість кластерів [14]. Ця властивість є критичною, оскільки фактична кількість просторово відокремлених груп залишків у конкретній структурі невідома апріорі й варіює від білка до білка. Альтернативні алгоритми, зокрема DBSCAN, також здатні вирішити цю задачу, проте потребують значно тоншого налаштування параметрів, специфічних для конкретного набору даних [14].

На третьому кроці просторові кластери зіставляються зі списками залишків, визначеними BMM для кожної кишені. Критерієм зіставлення слугує індекс Жаккара – частка спільних залишків відносно об'єднання двох множин; поріг становить 0,7 [14]. Якщо жоден кластер не досягає цього порогу, алгоритм повертається до вихідного BMM-списку і картує його залишки на всі ланцюги PDB, що пройшли перший крок. Такий сценарій відповідає випадку, коли ідентична кишеня присутня на кількох субодинацях гомомультимера і будь-яка з них є придатною для подальшого докінгу.

Четвертий крок розмежовує внутрішньоланцюгові та міжсубодиничні кишені: кластер, узгоджений з списком залишків, що отримані BMM,

перевіряється на наявність залишків із більш ніж одного ланцюга, і за позитивним результатом класифікується як міжсубодинична кишень. Цей підхід є особливо ефективним для виявлення кишень на стиках субодиниць гомомультимерних мішеней, де BMM-екстракція найчастіше стикається з неоднозначністю у віднесенні залишків до конкретних ланцюгів [14].

Усі операції з атрибутами білкових структур, а саме вирівнювання послідовностей, обчислення відстаней між залишками та запис проміжних результатів у структурні формати виконувались засобами бібліотеки молекулярного моделювання MolAR [94]. MolAR реалізована на Rust із безпечним управлінням пам'яттю і підтримує як аналіз траєкторій симуляції молекулярної динаміки, так і статичне опрацювання структур PDB, що робить її придатною для використання на цьому етапі [94].

#### **3.4.2. Побудова сіткового представлення кишені**

Стандартне представлення кишень у Focket у вигляді набору альфа-сфер незручне для задач, що вимагають булевих операцій над об'ємами кишень, зокрема, злиття просторово суміжних кишень або розділення кишень, некоректно об'єднаних на етапі геометричного виявлення. Безпосереднє виконання таких операцій над альфа-сферами потребує нетривіальних геометричних обчислень – перетину сфер та інтегрування залишкових об'ємів і є обчислювально неефективним. Тому кожна відібрана кишенька перетворюється на сіткове представлення, що зводить всі подальші об'ємні операції до булевих операцій над множинами точок [14].

Сіткове представлення будується за такою процедурою. Спершу визначається прямокутний паралелепіпед, межі якого задано мінімальними й максимальними координатами важких атомів кишені вздовж кожної осі. Усередині цього паралелепіпеда генерується кубічна сітка з кроком  $1,5 \text{ \AA}$  – щільність, достатня для подальшого докінгу без надмірного обчислювального



навантаження. З отриманої сітки зберігаються лише точки, що розташовані всередині принаймні однієї альфа-сфери, визначеної Frocket для цієї кишені. Порівняння двох представлень для сайту зв'язування інгібіторів дигідрооротатдегідрогенази (PDB ID: 6OC0) наведено на рис. 3.4.

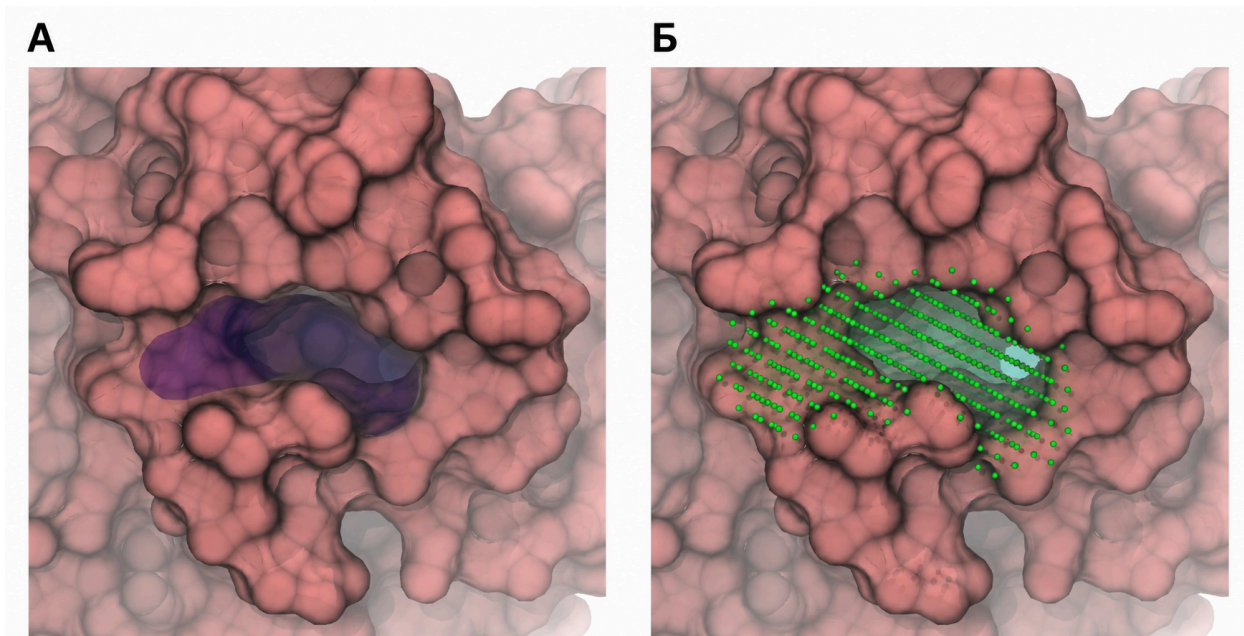


Рис. 3.4. Порівняння представлень одного й того ж інгібіторного сайту дигідрооротатдегідрогенази (PDB ID: 6OC0): А – представлення альфа-сферами Frocket, Б – об'ємне сіткове представлення [14].

Первинне сіткове представлення успадковує характерну ваду Frocket: альфа-сфери значної частини кишень не обмежуються порожниною білка, а частково виходять у розчинник, оскільки їхнє положення визначено лише локальною геометрією атомів без урахування форми молекулярної поверхні. Як наслідок, сіткові точки всередині таких «виступаючих» альфа-сфер формально належать до кишені, але фактично описують її артефактне розширення у бік розчинника. Для докінгу це є джерелом похибки: надлишковий об'єм допускає успішне розміщення сполук, більших за реальну порожнину, що спотворює оцінку їхньої придатності до зв'язування з мішенню [14].

Для усунення цього ефекту застосовано корекцію опуклою оболонкою. На основі важких атомів залишків кишені будується тривимірна опукла оболонка (convex hull), і з сіткового представлення видаляються всі точки, що виходять за її межі. Геометричну ілюстрацію наведено на рис. 3.5.

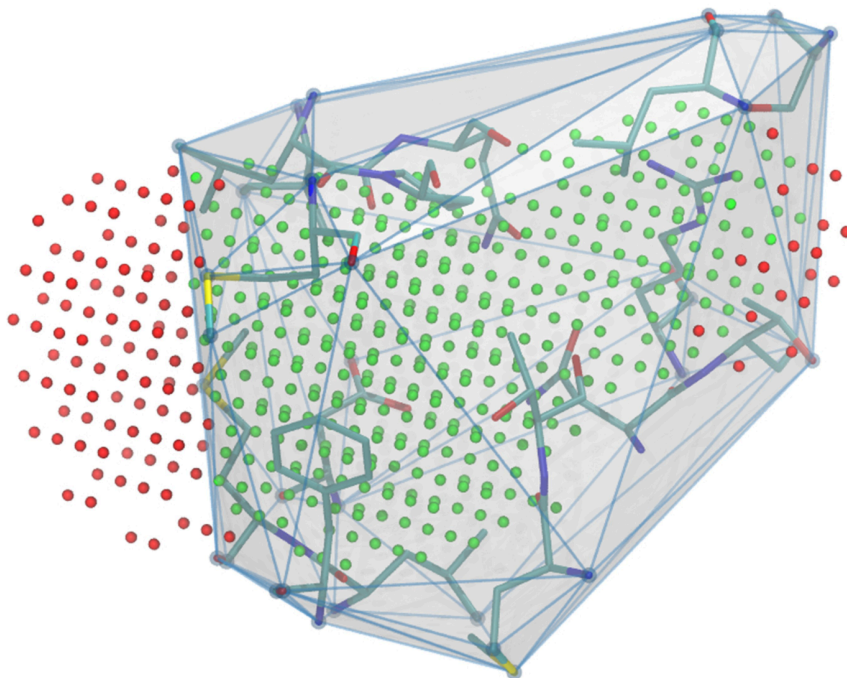


Рис. 3.5. Корекція сіткового представлення кишені опуклою оболонкою. Точки, позначені зеленим, розташовані всередині оболонки і зберігаються; точки, позначені червоним, виходять за її межі і відкидаються як артефактні [14].

Після застосування опуклої оболонки сіткове представлення додатково уточнюється шляхом видалення точок, які опинилися ближче до будь-якого атома білка, ніж ван-дер-ваальсів радіус цього атома. Цей крок не є строго необхідним для коректності докінгу, оскільки сучасні алгоритми докінгу самостійно не допускають стеричних конфліктів між лігандом і білком, проте він помітно поліпшує візуальну якість результату і прибирає «псевдоточки», що формально належать до кишені, але насправді перекриваються з атомами білкового каркасу [14].

Останнім елементом побудови сіткового представлення є злиття просторово суміжних кишень. Нерідко Focket розділяє єдиний



функціональний сайт, наприклад, АТР-зв'язувальний сайт протеїнкінази, на два чи більше сегменти, кожен з яких окремо не описує цілої біологічної кишені. Для відновлення цілісного представлення виконується попарне зіставлення сіткових представлень таких сегментів: якщо індекс Жаккара між ними перевищує емпірично обраний поріг, їхні точкові набори об'єднуються і трактуються як єдина кишеня. Цей самий механізм виконує й концептуально важливу додаткову функцію – агрегацію описів одного сайту, реконструйованих з різних публікацій, де зазвичай наведено різні або частково перекриті списки залишків. Відповідні кишені є просторово близькими і зливаються в єдине сіткове представлення, що узагальнює всю доступну літературну інформацію про цей сайт [14].

### **3.5. Оцінка ефективності розробленого методу**

Вибір методології оцінки ефективності розробленого конвеєра потребує окремого обґрунтування, оскільки у літературі відсутня усталена процедура, яка безпосередньо відповідала б поставленій задачі. Загальноприйняті метрики методів геометричного виявлення кишень – таких як DeepPocket, P2Rank чи PURESNet – не застосовні, бо ці методи розв'язують іншу за природою задачу: виявлення потенційних кишень за одною лише структурою білка [83, 84, 89]. Розроблений конвеєр приймає на вхід не лише структуру, а й сукупність наукових публікацій, і виконує дві послідовні операції – витяг описів кишень з літератури та пріоритизацію геометрично передбачених кишень за цими описами.

Для кількісної оцінки використано метрики на двох рівнях деталізації: на рівні кишень і на рівні амінокислотних залишків. Результати оцінки трьох конфігурацій конвеєра на бенчмарк-датасеті наведено на рис. 3.6.

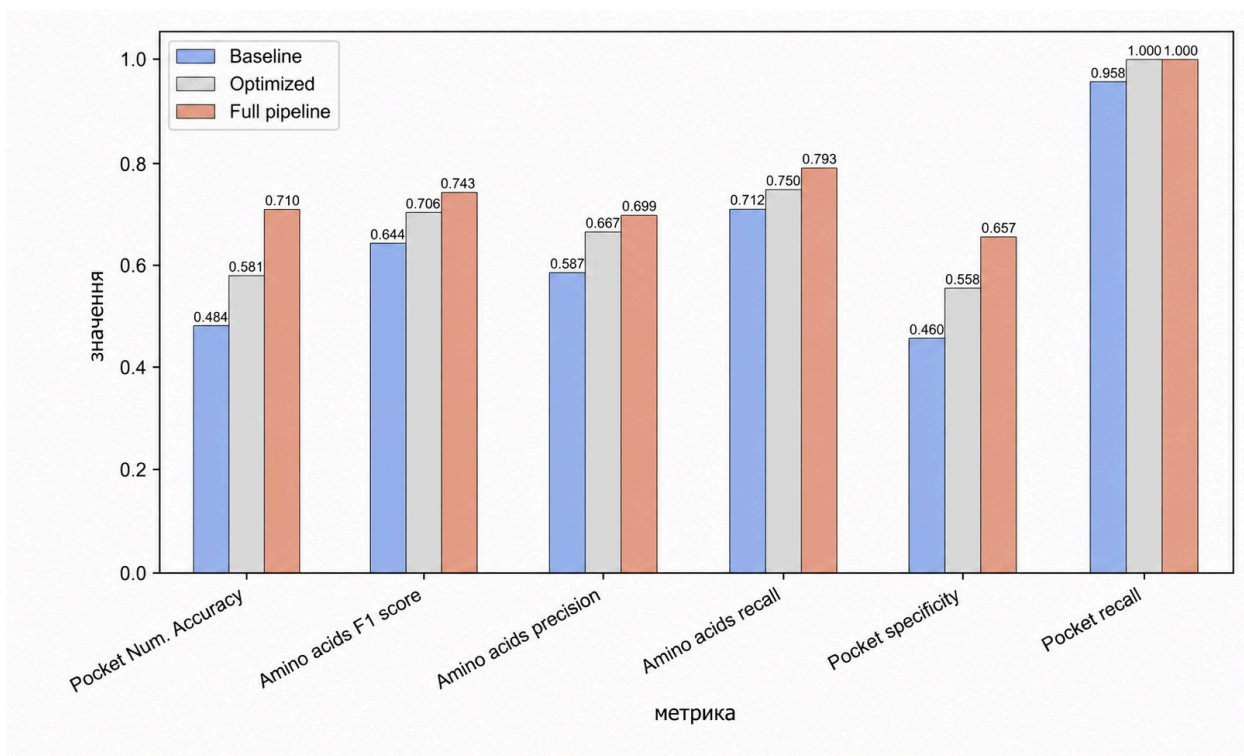


Рис. 3.6. Метрики ефективності моделі GPT-4o-mini для трьох конфігурацій конвеєра: Baseline (неоптимізовані промпти), Optimized (оптимізовані промпти без уточнення) та Full pipeline (повна конфігурація з уточненням) [14].

Для базової конфігурації Baseline Pocket Number Accuracy становить 0,484; Pocket Recall – 0,958; Pocket Specificity – 0,460. Оптимізація запитів на етапі екстракції при незмінному етапі фільтрації помітно покращує показники: Pocket Number Accuracy зростає до 0,581, Pocket Recall досягає 1,0, Pocket Specificity – 0,558. Водночас аналіз результатів конфігурації Optimized виявив характерну ваду: модель у частині випадків некоректно розділяє один сайт на кілька записів. Додавання етапу уточнення у повній конфігурації Full pipeline усуває цю ваду і додатково підвищує Pocket Number Accuracy до 0,710, Pocket Specificity – до 0,657, зі збереженням максимального значення Pocket Recall = 1,0. Повна конфігурація також покращує метрики на рівні залишків: F1-оцінка зростає до 0,743, precision – до 0,699, recall – до 0,793, що зумовлено доповненням списків кишень пропущеними амінокислотами на етапі уточнення [14].

Окремо оцінено якість етапу фільтрації як бінарної класифікації релевантних і нерелевантних публікацій. На цьому рівні модель демонструє accuracy 0,870, recall 1,0, precision 0,833 і FPR 0,363 за нульового значення FNR [14]. Нульовий FNR означає, що жодна релевантна публікація не втрачається на етапі фільтрації і вся наявна у літературі інформація про сайти зв'язування потенційно досяжна для подальшої екстракції. Водночас FPR 0,363 означає, що приблизно 36 % справді нерелевантних публікацій помилково проходять етап фільтрації. Ця частка визнана прийнятною, оскільки такі публікації, як правило, не здатні згенерувати хибну кишеню, яку було б складно виявити під час постобробки [14].

Окрім кількісної оцінки, розроблений метод перевірено на серії якісних прикладів, які демонструють його сильні й слабкі сторони на різнотипних мішенях.

Перший приклад – рецептор GABA<sub>A</sub> (PDB ID: 6X3T), де сайт зв'язування етомідату локалізовано на інтерфейсах між  $\alpha$ - і  $\beta$ -субодиницями. Розроблений алгоритм коректно ідентифікував дві еквівалентні кишені на інтерфейсах A:B і C:D, що підтверджує його здатність розпізнавати міжсубодиничні сайти у гомомультимерних зборках і однорідно розподіляти залишки між парами ланцюгів. Візуалізацію результату наведено на рис. 3.7.

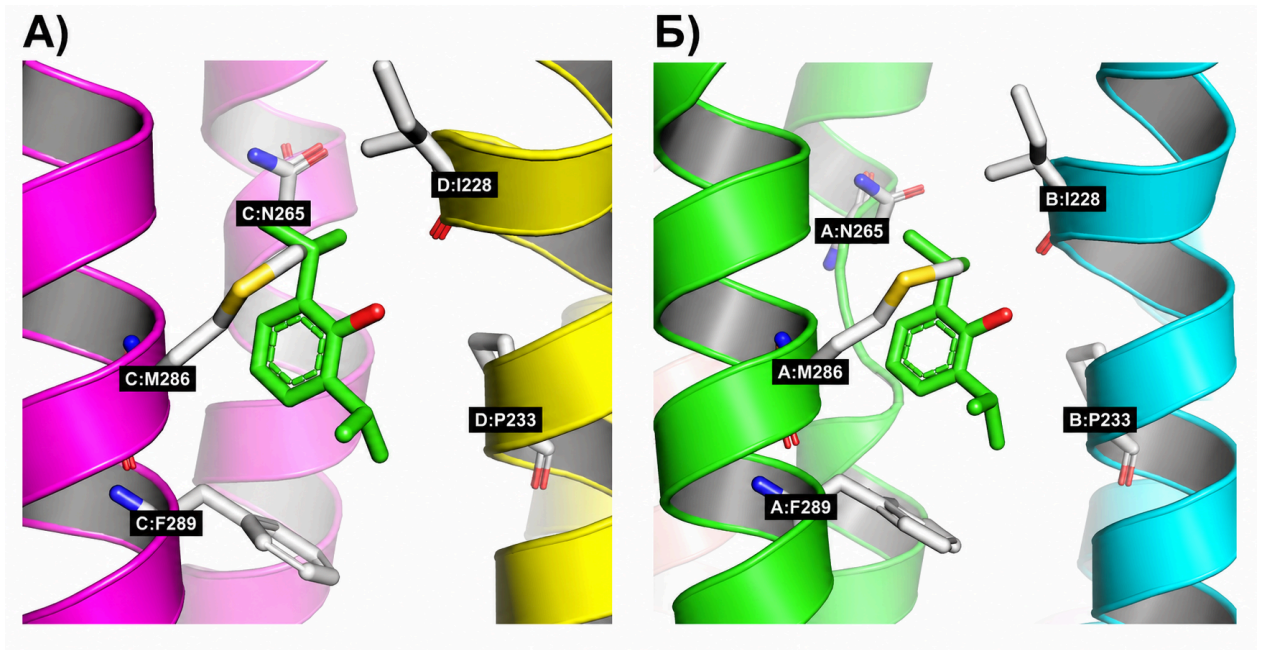


Рис. 3.7. Картування залишків, виконане розробленим методом, для двох еквівалентних сайтів зв'язування етомідату у рецепторі GABA<sub>A</sub> (PDB ID: 6X3T): А – інтерфейс ланцюгів А:В; Б – інтерфейс ланцюгів С:Д [14, 114].

Другий приклад – псевдокіназа MLKL (PDB ID: 7MON) ілюструє роботу етапу злиття просторово суміжних кишень. Focket з параметрами за замовчуванням розділив АТР-зв'язувальний сайт на два окремих сегменти (рис. 3.8 (А)). Крок злиття за індексом Жаккара усунув цю фрагментацію і сформував єдине цілісне сіткове представлення сайту (рис. 3.8 (Б)).

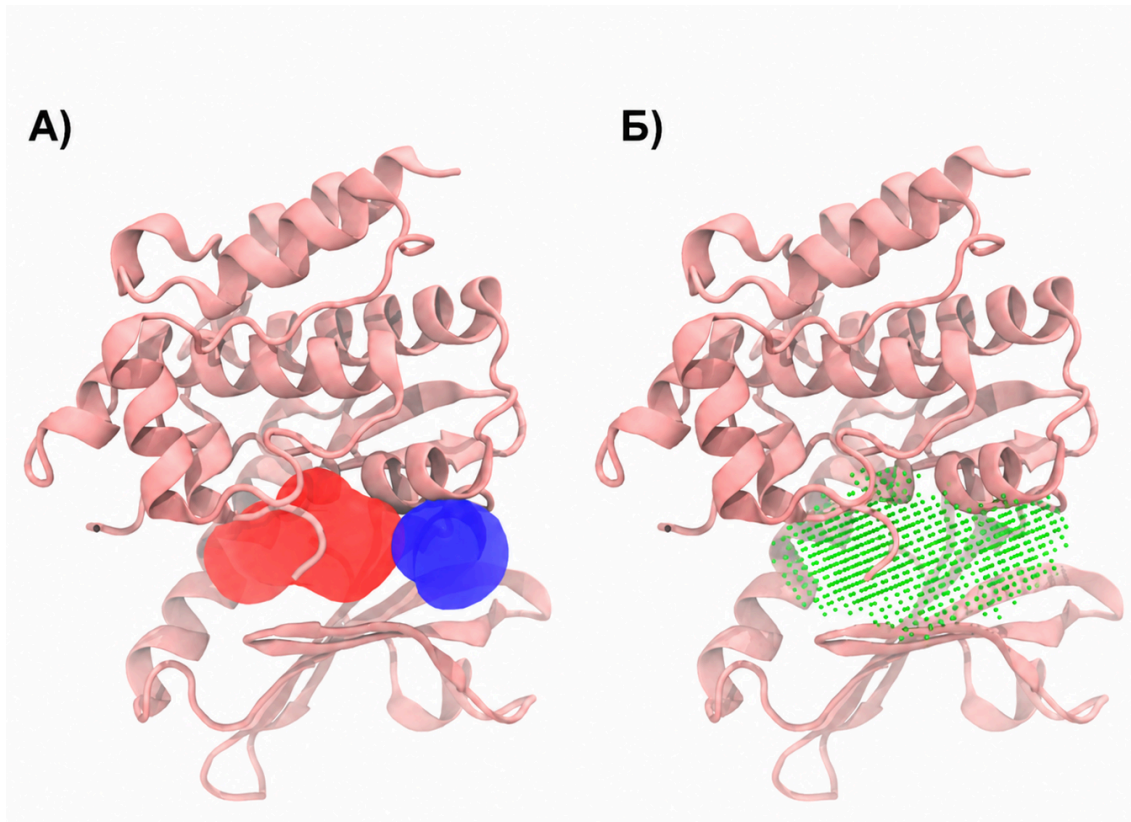


Рис. 3.8. Псевдокіназа MLKL (PDB ID: 7MON): А – вихідні фрагментовані кишені, отримані Focket, у вигляді альфа-сфер; Б – остаточна цілісна кишеня після злиття та постобробки [14, 115].

Такий тип фрагментації є типовим для протеїнкіназних АТР-сайтів і регулярно потребує злиття для адекватного опису каталітичної кишені [14, 115].

Третій приклад – мускариновий рецептор  $M_2$  (PDB ID: 4MQT) демонструє межі методу. Focket не зміг розділити дві відомі кишені – ортостеричний сайт і позаклітинний вестибюль і об'єднав їх у спільну порожнину вже на етапі геометричного виявлення (рис. 3.9).

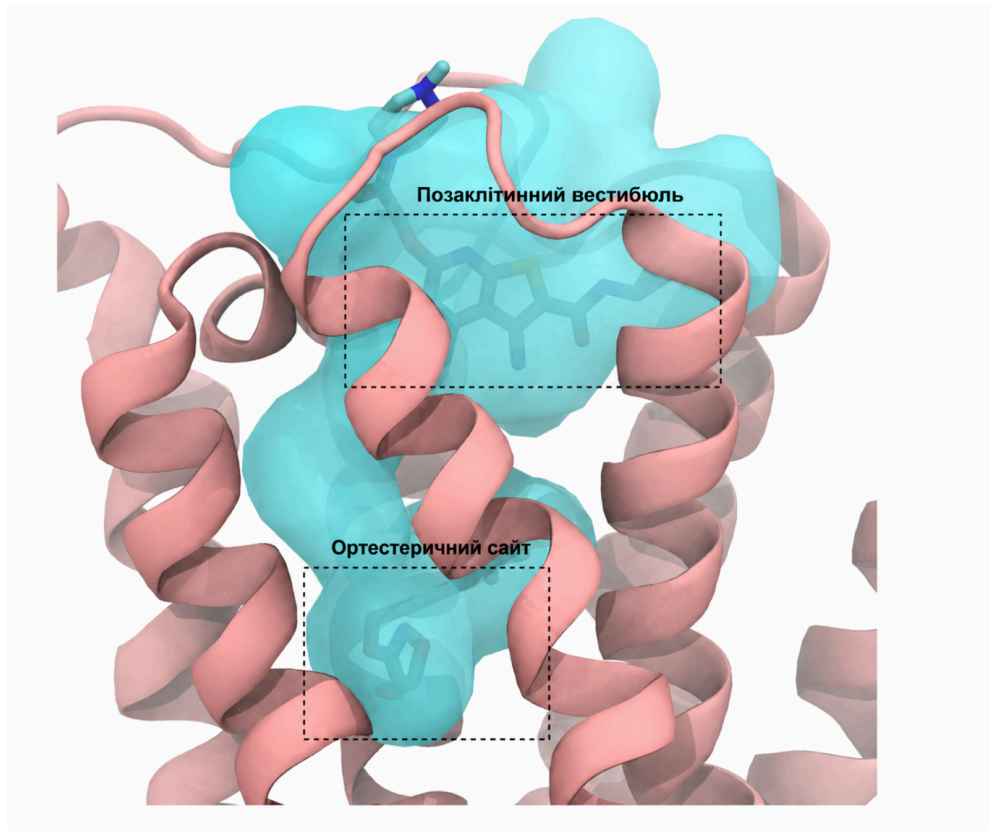


Рис. 3.9. Мускариновий рецептор  $M_2$  (PDB ID: 4MQT): кишеня, що об'єднує ортостеричний сайт і позаклітинний вестибюль [14, 116].

Механізм злиття за індексом Жаккара не вирішує задачі розділення виявленої кишені, яка насправді є двома різними сайтами, оскільки діє у протилежному напрямі – об'єднує фрагменти, а не розділяє ціле. Цей приклад ілюструє, що ефективність конвеєра обмежена точністю попереднього геометричного етапу і успадковує принаймні частину обмежень Fprocket.

Четвертий приклад – хімерна структура людського натрієвого каналу  $\text{Nav}_{1.7}$ -VSD4-NavAb (PDB ID: 5EK0) є тетрамерним комплексом, у якому кожен з чотирьох ланцюгів містить перенесений методами білкової інженерії домен сенсора напруги VSD4 людського  $\text{Nav}_{1.7}$ . Конвеєр правильно локалізував чотири ідентичні сайти зв'язування – по одному у межах кожного VSD4 – що відповідають відомому сайту зв'язування сполуки GX-936 в екстрацелюлярній щілині VSD4 (рис. 3.10).



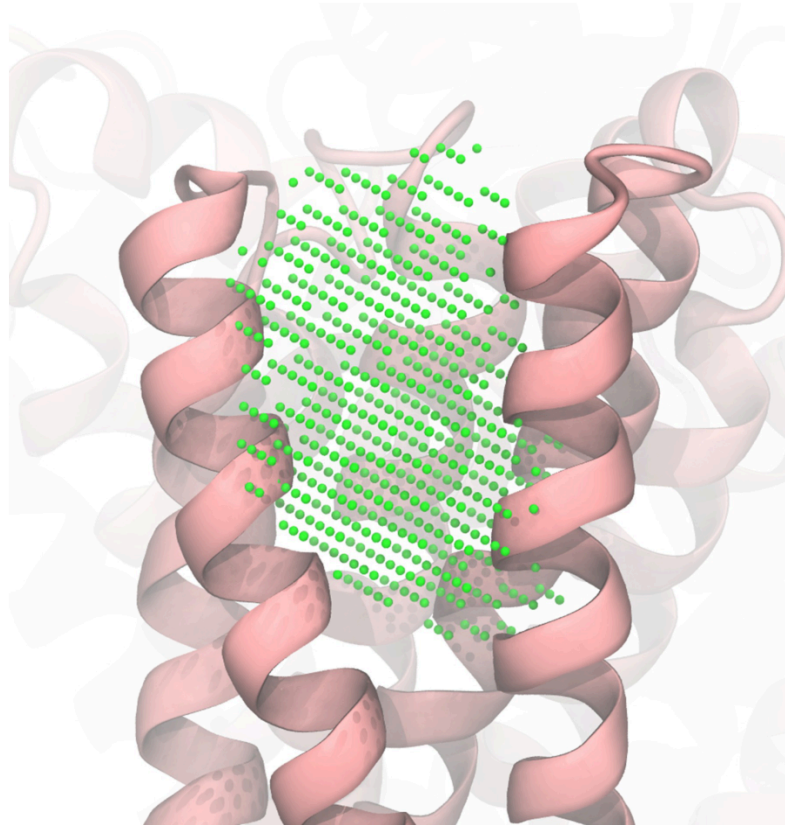


Рис. 3.10. Кишеня зв'язування сполуки GX-936 у домені VSD4 людського Nav1.7 (PDB ID: 5EK0) [14, 117].

Цей приклад підтверджує здатність алгоритму працювати з гомомерними молекулярними мішенями і виявляти в них набір еквівалентних сайтів.

Попри отримані результати, розробленому методу властиві певні обмеження. По-перше, результати отримано для конкретної моделі GPT-4o-mini і перенесення методу на інші моделі потребує повторної валідації. По-друге, бенчмарк-датасет охоплює обмежену кількість публікацій і мішеней. Хоча він включає основні типи складних випадків такі як, розрізнення між кількома сайтами зв'язування, коректне віднесення залишків до відповідної кишені та узгодження різномірної термінології для одного сайту, його розширення здатне виявити додаткові граничні випадки. По-третє, у цьому дослідженні не виконано прямого кількісного порівняння з ручною обробкою людиною-експертом; емпірично час роботи конвеєра (секунди–хвилини на одну публікацію) на порядки менший за типовий час

ручної обробки (години), проте точність конвеєра залишається нижчою за граничну точність експертного аналізу. По-четверте, конвеєр успадковує обмеження Frocket у випадках, коли дві фізично відокремлені кишені геометрично об'єднано у спільну порожнину, як проілюстровано на прикладі рецептора M<sub>2</sub>. По-п'яте, поточна реалізація не здатна безпосередньо витягати інформацію з табличних і графічних елементів публікацій, оскільки працює на рівні текстового вмісту, отриманого за допомогою GROBID.

Перспективними напрямками розвитку методу є: розширення бенчмарк-дасету за рахунок додаткових мішеней і типів сайтів; випробування альтернативних геометричних предикторів, що пом'якшують обмеження Frocket; застосування архітектур CoT або ToT для складних випадків, де пряме промптування не дає задовільної точності; а також інтеграція з базою BioLip2 для перехресної валідації витягнутих залишків і використання наявних у базі констант активності як додаткової інформації про кишені. Ці напрями не змінюють загальної архітектури конвеєра, але здатні розширити сферу його практичного застосування у високопродуктивних сценаріях обчислювального пошуку ліків.

Валідований на відкритих бенчмарках метод пріоритизації сайтів зв'язування у наступному підрозділі застосовано до TRPC4 миші (PDB ID: 5Z96), що дозволило ідентифікувати найбільш перспективні кишені для подальшого віртуального скринінгу низькомолекулярних модуляторів.

### **3.6. Застосування гібридного методу для аналізу сайтів зв'язування TRPC4**

Ідентифікація та структурно-функціональна характеристика сайтів зв'язування низькомолекулярних лігандів у TRPC4 є безпосередньою передумовою всіх подальших етапів обчислювального дизайну модуляторів цього каналу. У цьому підрозділі запропонований конвеєр застосовано до



конкретної мішені – мишачого TRPC4, а кріо-ЕМ структури 5Z96, з метою локалізації всіх сайтів зв'язування, підтверджених у літературі, їх структурного опису та підготовки об'ємних представлень, придатних для подальшого молекулярного докінгу.

Вибір структури 5Z96 обґрунтовано її роллю як однієї з перших і найкраще охарактеризованих кріо-ЕМ структур TRPC4 ссавців, що представляє повноцінний гомотетрамерний комплекс із чотирма ідентичними субодиницями (ланцюги A, B, C, D) і дозволяє одночасно візуалізувати як внутрішньосубодиничні, так і міжсубодиничні кишені [25]. Такий вибір узгоджується з особливостями розробленого, у підрозділі 3.4, алгоритму картування залишків, який розроблено саме для коректної роботи з гомомультимерними молекулярними мішенями.

### **3.6.1. Формування корпусу публікацій для ВММ-опосередкованої екстракції сайтів зв'язування TRPC4**

Для ідентифікації біологічно релевантних сайтів зв'язування у структурі TRPC4 первинний пул кандидатних публікацій було сформовано вручну і склали його чотири рецензовані джерела, відібрані за результатами пошуку рецензованої літератури за запитом "TRPC4 structure" та "TRPC4 binding site" і подальшою перевіркою на відповідність TRPC4 як прямій мішені (а не виключно його гомологу TRPC5 чи іншим представникам родини TRPC) [25–27, 54]. Такий порядок ручного попереднього відбору відповідає процедурі, прийнятій, де вхідні статті обираються дослідником із топу веб-пошуку без додаткових автоматизованих критеріїв [14].

Сформований пул із чотирьох публікацій було подано на вхід триетапного ВММ-конвеєра (фільтрація, екстракція, уточнення) у конфігурації, описаній у підрозділі 3.3.1. На етапі фільтрації одне з джерел – оглядова стаття, присвячена молекулярним механізмам активації TRPC4 як

детектора  $G_{i/o}$ - і  $G_{q/11}$ -сигналіngu, – було класифіковано як нерелевантне для задачі та виключено з подальшої обробки [54]. Ця оцінка є методологічно коректною: робота систематизує функціональні відомості про CIRB-мотив, ділянку зв'язування  $G\alpha_{12}$  і фосфоліпідні взаємодії, проте оперує посиленнями на первинні структурні джерела і не містить власних оригінальних переліків амінокислотних залишків, що формують конкретні кишені зв'язування малих молекул; описані в ній інтерфейси (CIRB, С-кінцевий G-білковий сайт) є інтерфейсами білок–білкової взаємодії, що не відповідає цільовому визначенню конвеєру. Таким чином, фінальний корпус склали три структурні роботи, які сумарно характеризують усі основні типи сайтів TRPC4, релевантних для подальшого віртуального скринінгу – від ліпідних кишень і регуляторних катіон-зв'язувальних сайтів до внутрішньомембранної фармакологічно адресованої порожнини VSL-домену.

Кріо-ЕМ структура апо-форми *Danio rerio* TRPC4 надає референтний опис іон-провідної пори – селективного фільтра G577, нижніх воріт I617/N621 з додатковою цитоплазматичною констрикцією Q625 та LFW-мотиву (L571–F572–W573), а також двох ліпід-зв'язувальних кишень: сайту холестерильного гемісукцинату у порожнині між pre-S1-elbow-доменом і спіралями S1/S4 (Y315, F366, L374, L492, L495, F496, N499) і кільцевої кишені PA на інтерфейсі pore-helix/S6 сусіднього протомеру (Q569, R553, F572, W573, T599, T603, V606, I607, L520, L527) [26]. Додатково описано позаклітинний дисульфідний місток C549–C554 та цитоплазматичний край rib-helix із залишками E648/E649, що формують інтерфейс взаємодії зі Stim1 [26].

Незалежна кріо-ЕМ структура апо-форми мишачого TRPC4 при роздільності 3,3 Å підтверджує архітектуру пори і селективного фільтра та виявляє додатковий катіон-зв'язувальний сайт у S1–S4-боксі VSL-домену, координований залишками E417 і Q420 (S2), N435 і D438 (S3), з Y373 (S1) та R491 (S4) у ролі "кришки" з боку позаклітинного простору; щільність у центрі сайту в цій роботі інтерпретовано як  $Na^+$  [25]. Це джерело також

підтверджує наявність РА в околі пори та CHS-сайтів на периферії трансмембранного домену.

Серія крио-ЕМ структур TRPC4 у комплексі з піридазиновими інгібіторами GFB-8438, GFB-9289 і GFB-8749 та з кальмодуліном надає атомарний опис трьох ключових для фармакологічної інтервенції сайтів [27]. По-перше, внутрішньомембранний інгібіторний сайт у порожнині VSL-домену, утворений залишками H369, T370, Y373 (S1), F413 (S2), D438, M441, N442 (S3), S488, R491, L495 (S4) і Y646 з TRP-helix; цей сайт є цільовим для всіх трьох GFB-сполук і визначає селективність до TRPC4/TRPC5 над TRPC6, зокрема, через заміни N442 на L534 та F413 на M505 у TRPC6. По-друге,  $\text{Ca}^{2+}$ -сайт у тому ж VSL-домені (E417, Q420 із S2; Y429, N435, D438 із S3), що геометрично співпадає з катіон-зв'язувальним сайтом, описаним у праці Duan *et al.*, але з альтернативною інтерпретацією щільності як  $\text{Ca}^{2+}$  та з додатковим залученням Y429 [25]. По-третє, CaM-зв'язувальний сайт на rib-helix, що в апо-формі частково неупорядкований і стабілізується при зв'язуванні CaM; до інтерфейсу залучаються залишки 273–277 та 666–703 TRPC4, зокрема F688, L691, R693, A696, N698, R700, H703 з боку rib-helix і M275, D274 з боку прилеглої петлі. У тій же роботі підтверджено наявність кільцевої РА в околі пори в усіх лігандних станах каналу.

### 3.6.2. Кишені-кандидати у структурі 5Z96 та їхня кластеризація

Застосування Focket до структури 5Z96 із параметрами за замовчуванням дало набір із 267 кишень-кандидатів. Кожна з цих кишень є кластером просторово близьких альфа-сфер, отриманим внутрішньою процедурою Focket: алгоритм будує тесселяцію Вороного на атомах протеїну, ідентифікує альфа-сфери у її вершинах, відфільтровує занадто малі та занадто великі сфери і групує решту у кластери, де один кластер

відповідає одній потенційній кишені [77]. Така надлишковість виходу є очікуваною поведінкою геометричних детекторів кишень для великих мембранних тетрамерних білків і відображає схильність Fpocket ідентифікувати кожен поверхневий заглибину, борозну та міжспіральний простір незалежно від її біологічної значущості [14]. Більшість із цих 267 кандидатів мають низькі значення Druggability Score та невеликий об'єм і відповідають шумовим геометричним особливостям поверхні, що не пов'язані із зв'язуванням малих молекул.

Послідовна фільтрація за двома критеріями: threshold residue match (мінімальний відсоток атомів залишку, що контактують з альфа-сферами кишені) та threshold pocket match (мінімальний Jaccard-індекс перекриття залишків кишені з BMM-екстрагованим набором), звузила множину кандидатів до 16 кишень, що пройшли збіг із літературно-валідованими описами сайтів. Враховуючи гомотетрамерну природу TRPC4, ця кількість закономірно відображає наявність груп із чотирьох еквівалентних сайтів, локалізованих симетрично на кожній із чотирьох субодиниць, а також додаткові інтерфейсні сайти, що формуються на стиках сусідніх субодиниць.

Алгоритмічне картування залишків, описане у підрозділі 3.4.1, разом з етапом просторової кластеризації BMM-екстрагованих залишків методом MeanShift і узгодженням кластерів з описами сайтів дали змогу звести цю множину до чотирьох біологічно унікальних сайтів зв'язування, кожен з яких представлений у структурі чотирма еквівалентними копіями. Слід підкреслити, що ця кластеризація принципово відрізняється від внутрішньої кластеризації альфа-сфер у Fpocket: вона оперує на рівні амінокислотних залишків, витягнутих BMM-конвеєром із літератури, і має на меті виявлення сайтів, що формуються на інтерфейсах між субодиницями [14]. Виявлено такі унікальні сайти:

1. Катіон-зв'язувальний сайт у межах VSL-домену кожної субодиниці (ланцюги A, B, C, D) TRPC4;

2. Сайт зв'язування PA на інтерфейсі порової спіралі та спіралі S6 сусідньої субодиниці (інтерфейси A:B, A:D, B:C, C:D);
3.  $\text{Ca}^{2+}$ -зв'язувальний сайт у VSL-домені (ланцюги A, B, C, D);
4. Сайт зв'язування інгібіторів у VSL-домені, у якому, за даними попередніх структурних і біохімічних досліджень, локалізовано зв'язування низькомолекулярних інгібіторів серії GFB – GFB-8438, GFB-9289 та GFB-8749 (ланцюги A, B, C, D).

Підсумкове число у 16 фінальних кишень (4 унікальні сайти  $\times$  4 еквівалентні копії) узгоджується з вихідною множиною з 16 кишень, відфільтрованих за збігом із BMM-екстрагованими залишками: етап об'єднання кишень, що перекриваються не зменшив їх кількості, оскільки еквівалентні копії одного сайту, розташовані на різних субодиницях, просторово не перекриваються і коректно тримаються алгоритмом як окремі записи. Подальша конвертація альфа-сфер Fprocket у сіткові представлення з обрізкою за convex hull залишкового оточення дала фінальні об'ємні описи кишень, придатні для подальшого віртуального скринінгу [14].

Важливо підкреслити, що катіон-зв'язувальний сайт (пункт 1) і  $\text{Ca}^{2+}$ -зв'язувальний сайт (пункт 3) обидва є складовими VSL-домену TRPC4. До того ж самого домену належить і сайт зв'язування інгібіторів (пункт 4), що робить VSL-домен TRPC4 ключовою фармакологічною областю цього каналу. PA-кишеня (пункт 2), натомість, розташована поза межами VSL-домену в області порового домену і визначає ліпідне оточення каналу у мембрані.

### **3.6.3. Ідентифікація сайтів зв'язування у структурі TRPC4 миші та їх структурно-функціональна характеристика**

Застосування описаного у попередніх підрозділах гібридного конвеєра до структури 5Z96 дозволило звузити вихідну множину з 267 кандидатних

кишень, отриманих Focket, до чотирьох біологічно унікальних сайтів зв'язування, кожен з яких представлений у тетрамері чотирма еквівалентними копіями. У цьому підрозділі для кожного з виявлених сайтів подано структурну характеристику, перелік амінокислотних залишків, що формують його у структурі 5Z96, та функціональну інтерпретацію з критичною оцінкою якості автоматизованої ВММ-екстракції відносно первинних літературних джерел [25–27].

**Катіон-зв'язувальний сайт у ділянці S1–S4 VSL домену TRPC4.** Цей сайт локалізовано у цитоплазматичній частині трансмембранного домену S1–S4, що у TRPC4 не виконує класичної функції сенсора напруги, а становить структурно-регуляторний VSL-домен. Кишеня є зануреним внутрішньосубодиничним сайтом: усі шість залишків, що формують її, належать одному ланцюгу (A, B, C або D), а гомотетрамерна симетрія каналу забезпечує існування чотирьох еквівалентних копій сайту, по одній на кожную субодиницю. За хімічним характером сайт є переважно полярним, що відображає його функцію координації катіона: чотири координуючі залишки (E417 і Q420 зі спіралі S2, N435 і D438 зі спіралі S3) формують гідрофільну порожнину з боку цитоплазми, а тирозин Y373 (спіраль S1) і аргінін R491 (спіраль S4) розташовані над сайтом з протилежного боку і відіграють роль «кришки», що обмежує вихід катіона у позаклітинний простір [25].

Функціонально сайт є компонентом катіон-зв'язувальної системи VSL-домену TRPC4 і бере участь в модуляції гейтингу каналу. У структурі Duan *et al.* щільність у центрі сайту інтерпретовано як іон  $\text{Na}^+$ , оскільки натрій був домінуючим катіоном у буфері очистки [25]. Геометричний тип координації (чотири полярні залишки на S2/S3 з негативно зарядженою «кишенею» E417/D438) характерний для лужноземельних і лужних катіонів, і збережений у межах підродини TRPC (за винятком TRPC1) [25]. Сайт не є мішенню зв'язування для низькомолекулярних модуляторів і тому не використовується у віртуальному скринінгу в межах цієї дисертації. Натомість він важливий як референтна точка для розуміння  $\text{Ca}^{2+}$ -залежної

регуляції TRPC4 і просторово співвідноситься з ліганд-зв'язувальною кишенею VSL-домену, що буде описана далі.

Звірка вмісту PDB-файлу цього сайту з первинним описом Duan *et al.* показує дослівний збіг 6 із 6 залишків: у тексті статті прямо названо чотири координуючі залишки на S2/S3 та два «кришкові» залишки на S1/S4 [25]. Перелік ідентифікованих залишків наведено в табл. 3.1.

Таблиця 3.1.

Амінокислотні залишки катіон-зв'язувального сайту S1–S4 у структурі TRPC4 миші (5Z96, ланцюг A).

| Ланцюг | Залишок       | Локалізація |
|--------|---------------|-------------|
| A      | Tyr373 (Y373) | спіраль S1  |
| A      | Glu417 (E417) | спіраль S2  |
| A      | Gln420 (Q420) | спіраль S2  |
| A      | Asn435 (N435) | спіраль S3  |
| A      | Asp438 (D438) | спіраль S3  |
| A      | Arg491 (R491) | спіраль S4  |

Жодного хибнопозитивного включення ВММ не зробила; жодного релевантного залишку з джерела не пропущено. Це чистий випадок без галюцинацій, що демонструє характерну сильну сторону ВММ-екстракції у конвеєрі: для прозо-багатих структурних публікацій, у яких автори очевидно перераховують амінокислоти, що координують ліганд або іон, автоматизоване витягання еквівалентне ручному прочитанню досвідченим дослідником [14]. Окремо слід відзначити, що ВММ коректно не злила цей сайт з  $\text{Ca}^{2+}$ -сайтом, описаним Vinayagam *et al.* (див. нижче), попри значне просторове перекриття (4 спільні залишки з 5) і альтернативну інтерпретацію щільності як  $\text{Ca}^{2+}$  замість  $\text{Na}^+$ [27]. Обидва описи збережено як окремі записи відповідно до семантики джерел, а просторова кластеризація не виконала їх злиття. Це є свідомим методологічним рішенням конвеєра, що пріоритизує авторську

інтерпретацію над геометричним об'єднанням; межі такого підходу обговорено далі у контексті  $\text{Ca}^{2+}$ -сайту (рис. 3.11).

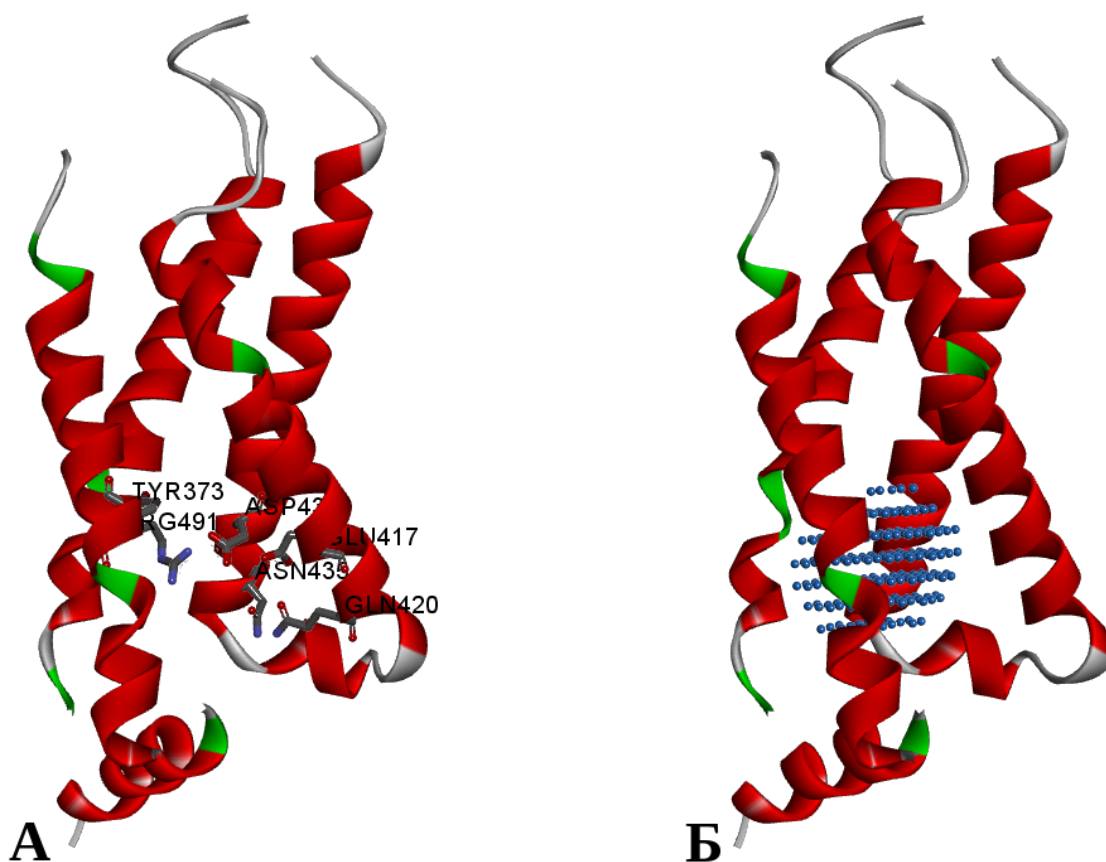


Рис. 3.11. Катіон-зв'язувальний сайт у ділянці S1–S4 VSL домену TRPC4 (структура 5Z96; ланцюг А): А – амінокислотне оточення кишені; Б – об'ємне сіткове представлення порожнини сайту (сині точки).

**Сайт зв'язування РА.** Даний сайт є міжсубодиничним сайтом, що формується на інтерфейсі між поровим хеліксом однієї субодиниці та спіраллю S6 сусідньої субодиниці [26]. Сайт розташований у трансмембранній частині каналу, на периферії порового домену, у площині мембрани й безпосередньо контактує з ліпідним бішаром. Гомотетрамерна симетрія TRPC4 забезпечує наявність чотирьох еквівалентних копій цього сайту, локалізованих на інтерфейсах А:В, А:Д, В:С та С:Д; усі чотири копії були коректно ідентифіковані як інтерфейсні, розробленим у підрозділі 3.4.1 алгоритмом картування, завдяки виявленню залишків з двох різних ланцюгів



у одному просторовому кластері. У хімічному відношенні сайт є амфіпатичним: полярна головка РА утворює специфічні контакти з амід- і амінозалишками порового хеліксу (Q569, W573) та з гідроксильною групою бічного ланцюга T599 спіралі S6, тоді як жирнокислотні хвости стабілізуються гідрофобним оточенням залишків спіралей S5, S6 і порової спіралі прилеглих субодиниць [26]. Перелік ідентифікованих залишків наведено в табл. 3.2.

Таблиця 3.2.

Амінокислотні залишки кільцевої РА-кишені у структурі TRPC4 миші (5Z96, інтерфейс A:B як представницький приклад)

| Ланцюг | Залишок       | Локалізація    |
|--------|---------------|----------------|
| A      | Thr599 (T599) | спіраль S6     |
| B      | Gln569 (Q569) | порова спіраль |
| B      | Trp573 (W573) | порова спіраль |

Інші три інтерфейси (A:D, B:C, C:D) мають ідентичну структурну організацію за гомотетрамерною симетрією: на кожному з них одна субодиниця надає залишок T599 зі спіралі S6, а сусідня – залишки Q569 і W573 з порового хеліксу.

РА-кишеня є не лише ліпід-зв'язувальним сайтом, але й фармакологічно валідованою мішенню для ксантинового класу інгібіторів TRPC1/4/5. Vinayagam *et al.* зафіксували щільність РА у цьому сайті як в апо-структурі TRPC4, так і в усіх лігандних станах і у комплексі з кальмодуліном, що свідчить про коеволюційно збережений характер взаємодії [26, 27]. У гомологічному каналі TRPC6 показано, що зв'язування DAG у структурно відповідній області активує канал, додатково підтримуючи функціональну значущість цього ліпідного сайту для гейтингу TRPC-каналів [118, 119]. Принципове відкриття щодо фармакологічного потенціалу сайту зроблено у роботі Wright *et al.* (PDB ID: 6YSN) і паралельно у роботі Song *et al.*: кріо-ЕМ структури TRPC5 у комплексі з ксантиновими інгібіторами

Pico145 (HC-608) та HC-070 продемонстрували, що ці сполуки зв'язуються саме у консервативній ліпід-зв'язувальній кишені, витісняючи з неї ендогенний фосфоліпід [11, 32]. Pico145 на сьогодні є одним з найпотужніших і найселективніших інгібіторів TRPC1/4/5, з  $IC_{50}$  у пікомолярному – нижньому наномолярному діапазоні і вираженою активністю проти гетеромерних TRPC1/4 і TRPC1/5 каналів [29]. Структурно сайт зв'язування Pico145 утворений тими самими залишками порового хеліксу (Q573 і W577 у нумерації TRPC5, гомологічні до Q569 і W573 TRPC4) з додатковими S6-залишками сусідньої субодиниці (T603, G606, V610, V614 у нумерації TRPC5, що відповідають T599, G602, V606, V610 у TRPC4) [11]. До однієї тетрамерної молекули TRPC5 одночасно зв'язується до чотирьох молекул інгібітора, по одній на кожен субодиничний інтерфейс – стехіометрія, що відповідає виявленню у структурі 5Z96 чотирьом еквівалентним копіям RA-кишені TRPC4 [11, 32].

РА-сайт є показовим прикладом водночас двох категорій обмежень ВММ-екстракції, прямо обговорених у попередніх підрозділах. По-перше, GROBID-парсер, на якому базується вхідний шар конвеєра, надійно витягує текст і підписи до рисунків, проте не здатний інтерпретувати маркери, нанесені безпосередньо на тривимірні структури або графічні елементи. У статті Vinayagam *et al.* у прозі прямо названо лише три залишки, що контактують з полярною головкою РА: T599 (S6), W573 та Q569 (поровий хелікс) – саме ці три залишки наявні у PDB-файлі сайту у складі мінімального ядра [26]. Решта залишків ліпідного оточення (R553, L520, L527, F572, T603, V606, I607), які видно на Figure 4c статті, у фінальному PDB-представленні відсутні, оскільки вони присутні лише як мітки на рисунку без явного перерахування у тексті [26]. Це не є помилкою ВММ-екстракції у вузькому розумінні (модель витягла все, що було у текстовому потоці), однак ілюструє системну межу текст-орієнтованого конвеєра для структурних публікацій. По-друге, кейс РА-кишені висвітлює окреме обмеження, пов'язане зі складом вхідного корпусу публікацій.

Сформований у підрозділі 3.2.1 пул із трьох рецензованих джерел не охоплює структурно-фармакологічно ключові роботи Wright *et al.* і Song *et al.*, у яких безпосередньо описано зв'язування Pico145 у цьому сайті та додаткові S6-залишки (T603, G606, V610, V614 у нумерації TRPC5), залучені до контактів з інгібітором [11, 25–27, 32]. Стаття Vinayagam *et al.* лише згадує ці роботи у вступі, без перерахування залишків, тому конвеєр, обмежений вхідним корпусом, об'єктивно не мав змоги витягти ці залишки [27]. Включення додаткових робіт у розширений корпус для повторного запуску конвеєра є наступним кроком при поглибленому аналізі сайту, що ілюструє методологічну рекомендацію щодо ітеративного нарощування корпусу публікацій по мірі поглиблення дослідження конкретної мішені. Алгоритмічна постобробка частково компенсує редукцію текст-витягнутого ядра при побудові сіткового представлення: точкова хмара охоплює простір, обмежений усіма альфа-сферами Frosket навколо мінімального ядра, тож фінальний об'єм кишені відображає реальні стеричні межі сайту, включно з областю S6, що зв'язує гідрофобну частину Pico145 (Рис. 3.12).

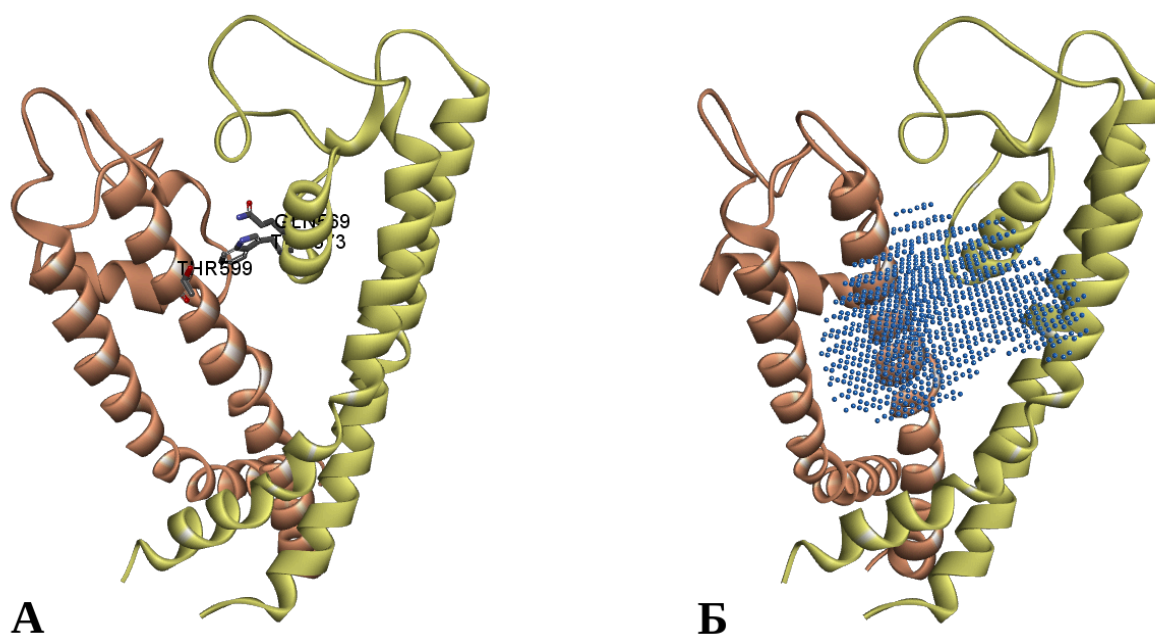


Рис. 3.12. Сайт зв'язування фосфатидної кислоти на інтерфейсі порової спіралі однієї субодиниці та спіралі S6 іншої субодиниці TRPC4 (інтерфейс

А:В ланцюгів структури 5Z96): А – амінокислотне оточення кишені; В – об'ємне сіткове представлення порожнини сайту (сині точки) .

Окремо слід відзначити коректну ідентифікацію міжсубодиничної природи сайту: чотири окремі PDB-файли для чотирьох інтерфейсів з правильним розподілом залишків між парами ланцюгів є прямим наслідком роботи кроку «check interface binding pockets» алгоритму картування, що засвідчує його ефективність на конкретній мішені TRPC4 і відповідає експериментальній стехіометрії «до чотирьох молекул Pico145 на один тетрамер TRPC5», встановлений Wright *et al.* [11].

У межах цієї дисертації РА-кишеня використана у Розділі 4 як цільовий сайт для молекулярного докінгу Pico145, що є одним з основних завдань дослідження.

**Ca<sup>2+</sup>-зв'язувальний сайт у VSL-доміні.** Даний сайт локалізовано у межах VSL-домену кожної субодиниці у безпосередній просторовій близькості до сайту зв'язування низькомолекулярних інгібіторів (див. нижче). Сайт є внутрішньосубодиничним і присутній у структурі 5Z96 у чотирьох еквівалентних копіях (по одній на ланцюги А, В, С і D), що відповідає гомотетрамерній організації каналу. За геометрією і хімічним характером сайт є полярною координаційною кишенею: п'ять залишків – E417 і Q420 зі спіралі S2, Y429 з петлі S2–S3, N435 і D438 зі спіралі S3 – формують октаедричне (з участю містка води) оточення для катіона Ca<sup>2+</sup> у конформації, типовій для зв'язування лужноземельних іонів [27]. Сайт частково перекривається з катіон-зв'язувальним сайтом, описаним Duan *et al.*, розділяючи з ним чотири з п'яти координуючих залишків (E417, Q420, N435, D438) і відрізняючись наявністю Y429 та відсутністю «кришкових» Y373 і R491 [25].

Vinayagam *et al.* виявили щільність у цьому сайті у всіх отриманих крио-ЕМ структурах і тлумачать її як іон Ca<sup>2+</sup> на підставі двох спостережень: (1) геометрія координації полярних залишків E417, Q420, N435, D438

відповідає координаційному оточенню, типовому для лужноземельних катіонів за критеріями CheckMyMetal; (2) гідроксильна група бічного ланцюга Y429 та оксо-група GFB-лігандів, що сидять у суміжній інгібіторній кишені, разом з містковою молекулою води можуть завершувати октаедричну координацію іона [27]. Активація TRPC4 і TRPC5 описана у літературі як  $\text{Ca}^{2+}$ -залежна, що додатково підтримує функціональну значущість цього сайту у механізмі гейтингу [49]. З фармакологічного погляду просторова суміжність  $\text{Ca}^{2+}$ -сайту з інгібіторною кишенею GFB-сполук створює потенційний механізм взаємної стабілізації: зв'язування  $\text{Ca}^{2+}$  може впливати на афінність зв'язування піридазинонів у сусідньому сайті і навпаки, що становить інтерес для подальшої експериментальної валідації.

Звірка вмісту PDB-файлу з первинним описом у тексті Vinayagam *et al.* показує дослівний збіг усіх 5 із 5 залишків: автори прямо згадують у тексті чотири координуючі залишки на S2/S3 («coordinated by the carbonyl oxygen of D438 and N435 of the S3 helix, along with E417 and Q420 of the S2 helix») та додатково тирозин Y429 («a hydroxyl group of Y429 close to the density») [27]. Перелік ідентифікованих залишків наведено в табл. 3.3.

Таблиця 3.3.

Амінокислотні залишки  $\text{Ca}^{2+}$ -зв'язувального сайту у структурі TRPC4 миші (5Z96, ланцюг A).

| Ланцюг | Залишок       | Локалізація |
|--------|---------------|-------------|
| A      | Glu417 (E417) | спіраль S2  |
| A      | Gln420 (Q420) | спіраль S2  |
| A      | Tyr429 (Y429) | петля S2–S3 |
| A      | Asn435 (N435) | спіраль S3  |
| A      | Asp438 (D438) | спіраль S3  |

Жодного хибнопозитивного включення ВММ не зробила, жодного релевантного залишку не пропущено. Цей кейс ілюструє ту саму сильну

сторону конвеєру, що й попередній (катион-зв'язувальний сайт S1–S4): для публікацій з явним перерахунком координуючих залишків автоматизоване витягання дає результат, ідентичний експертному. Принципова методологічна особливість цього сайту в іншому: він геометрично майже збігається з катион-зв'язувальним сайтом Duan *et al.* (4 спільні залишки з 5), і ВММ-конвеєр зберіг обидва описи як окремі сайти, оскільки джерела пропонують різні інтерпретації щільності ( $\text{Na}^+$  vs  $\text{Ca}^{2+}$ ) та різний склад «допоміжних» залишків (Y373 і R491 у Duan, Y429 у Vinayagam) [25, 27]. Алгоритмічне злиття кишень за індексом Жаккара не виконало об'єднання, оскільки індекс перекриття між двома описами хоч і високий, але не досягає порогу 0,7. У фізико-хімічному сенсі ймовірно йдеться про один і той самий катион-зв'язувальний регуляторний модуль VSL-домену, представлений у різних структурах і станах різними наборами координуючих залишків, тож збереження їх як окремих записів відображає семантику публікацій. Це межа методу, який зберігає авторські інтерпретації замість їх уніфікації, і вона свідомо допущена розробниками конвеєра, як компроміс на користь мінімізації частоти хибнопозитивних злиттів (рис. 3.13) [14].

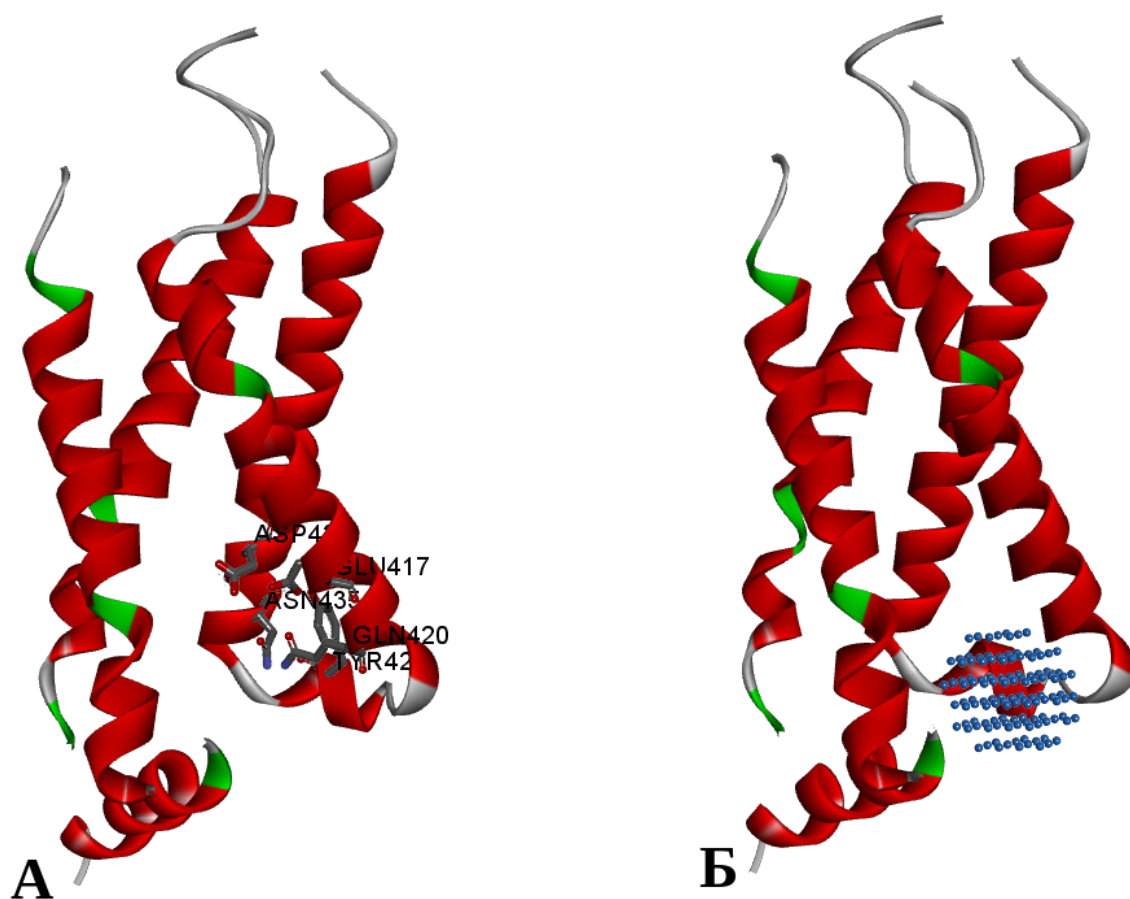


Рис. 3.13.  $\text{Ca}^{2+}$ -зв'язувальний сайт у VSL-домені субодиниці А TRPC4 (структура 5Z96; ланцюг А): А – амінокислотне оточення кишені; Б – об'ємне сіткове представлення порожнини сайту (сині точки).

**Сайт зв'язування інгібіторів у VSL-домені.** Даний сайт є внутрішньосубодиничним сайтом, локалізованим у центральній порожнині VSL-домену TRPC4 і обмеженим спіралями S1–S4 та TRP спіраллю. Структурно сайт розташований у трансмембранній частині каналу, що робить його доступним для зв'язування з мембрани, а не з цитоплазми чи позаклітинного простору. Гомотетрамерна симетрія каналу забезпечує наявність чотирьох еквівалентних копій кишені, по одній на субодиницю. У хімічному відношенні кишеня є переважно гідрофобною з полярним «дном», утвореним залишками S3 та S4: піридазиновий фрагмент лігандів формує  $\pi$ - $\pi$  стекінг з F413 (S2) та M441 (S3), водневі зв'язки з N442 (S3), Y373 (S1) і

S488 (S4), а також водневий зв'язок між оксо-групою центрального піперазинового кільця та R491 (S4); термінальні гідрофобні групи (трифторметилбензильна у GFB-8438, циклогексильна у GFB-9289 та GFB-8749) стабілізуються L495 (S4) і Y646 (TRP-хелікс), причому Y646 додатково утворює водневий зв'язок із дифторовим замісником у GFB-8749 [27].

Інгібіторна кишеня VSL-домену є центральною фармакологічною мішенню для піридазинового класу модуляторів TRPC4/5. У роботі Vinayagam *et al.* показано, що всі три досліджені сполуки – GFB-8438, GFB-9289 і GFB-8749 – зв'язуються у тій самій кишені і блокують активацію каналу енглерином А у дозозалежний спосіб ( $IC_{50}$  1,8 мкмоль для GFB-8438, 25,2 мкмоль для GFB-9289, 0,16 мкмоль для GFB-8749), причому інгібітори з нижчими  $IC_{50}$  додатково контактують з TRP-хеліксом через Y646 [27]. Це підтверджує запропоновану авторами модель алостеричного інгібування, згідно з якою зв'язування ліганду в периферійному VSL-домені через TRP-хелікс впливає на центральну пору, стабілізуючи канал у закритому стані [27]. Селективність GFB-сполук до TRPC4/TRPC5 над TRPC6 на 1–2 порядки пояснюється двома ключовими замінами у послідовності TRPC6: N442 на L534 (втрата водневого зв'язку з атомом азоту піридазинону) та F413 на M505 (втрата  $\pi$ – $\pi$  стекінгу з піридазиноновим кільцем) [27]. У межах нашого дослідження саме ця кишеня використана у Розділі 4, як цільовий сайт для молекулярного докінгу інгібітора GFB-887, що є одним з основних завдань дослідження.

Серед чотирьох сайтів TRPC4 інгібіторна кишеня має найбільш докладний літературний опис – зіставлення PDB-файлу з текстом Vinayagam *et al.* дозволяє верифікувати залишки з найвищою точністю [27]. У тексті статті явно перераховано 10 залишків, що контактують щонайменше з однією із трьох GFB-сполук – H369, Y373, F413, M441, N442, D438, S488, R491, L495, Y646 – і ще один залишок (F366) згадано у прозовому описі GFB-9289 як такий, що реорганізується при зв'язуванні циклогексильної групи;



додатково на панелях Figure 2D–F всіх трьох GFB-структур марковано залишок T370. PDB-файл містить 10 із цих залишків і не містить T370 та F366. Перелік ідентифікованих залишків наведено в табл. 3.4.

Таблиця 3.4.

Амінокислотні залишки інгібіторної кишені GFB-сполук у структурі TRPC4 миші (5Z96, ланцюг A).

| Ланцюг | Залишок       | Локалізація |
|--------|---------------|-------------|
| A      | His369 (H369) | спіраль S1  |
| A      | Tyr373 (Y373) | спіраль S1  |
| A      | Phe413 (F413) | спіраль S2  |
| A      | Asp438 (D438) | спіраль S3  |
| A      | Met441 (M441) | спіраль S3  |
| A      | Asn442 (N442) | спіраль S3  |
| A      | Ser488 (S488) | спіраль S4  |
| A      | Arg491 (R491) | спіраль S4  |
| A      | Leu495 (L495) | спіраль S4  |
| A      | Tyr646 (Y646) | TRP-спіраль |

T370 у тексті статті не описано як такий, що формує контакт з лігандом, тож його відсутність у PDB є коректним наслідком обмеження GROBID-парсера до текстового шару – це той самий клас обмежень, що й у випадку PA-кишені. F366 згадано у тексті лише побіжно у контексті опису GFB-9289, що, ймовірно, не пройшло порогу частоти згадування для включення у фінальний консенсусний список залишків кишені. Окремо цікавим є випадок D438: у переважній частині статті цей залишок коректно позначено як аспартат (зокрема, у Figure 2D–F та у описі  $\text{Ca}^{2+}$ -сайту), однак у одному реченні з опису апо-стану лігандної кишені трапляється помилкове написання «E438»; PDB містить саме D438, що свідчить про коректне

опрацювання конвеєром: або BMM зробила контекстне виправлення на основі переважного варіанту написання у статті, або алгоритм просторової кластеризації зіставив залишок із сусіднім  $\text{Ca}^{2+}$ -сайтом, де D438 ідентифіковано однозначно. Принципово важливим є і те, що BMM коректно не імпортувала у склад TRPC4-кишені залишки TRPC6 (L534 і M505), які у статті обговорюються лише як селективні заміни у гомологічному каналі, а не як частина TRPC4-сайту; жодного перенесення фактів з гомолога на цільовий білок без підстав у цьому кейсі не відбулося. Нарешті, агрегація інформації з трьох структур (TRPC4 з інгібіторами GFB-8438, GFB-9289, GFB-8749) у єдиний запис кишені відбулася за рахунок кроку злиття за індексом Жаккара, що дозволило сформувати уніфікований опис сайту, який інтегрує дещо різні набори контактів для кожної з трьох сполук – функціональність, явно винесена розробниками конвеєра, як перевага над підходами, що ґрунтуються на одиничній ко-кристалізованій структурі [14]. Таким чином, інгібіторна кишеня VSL-домену є прикладом успішної автоматизованої екстракції з помірною редукцією за рахунок втрати figure-only залишків – редукцією, яка не впливає на придатність сайту до подальших задач докінгу, оскільки відсутні залишки розташовано на периферії кишені і не задіяні у ключових специфічних взаємодіях з лігандами (рис. 3.14).

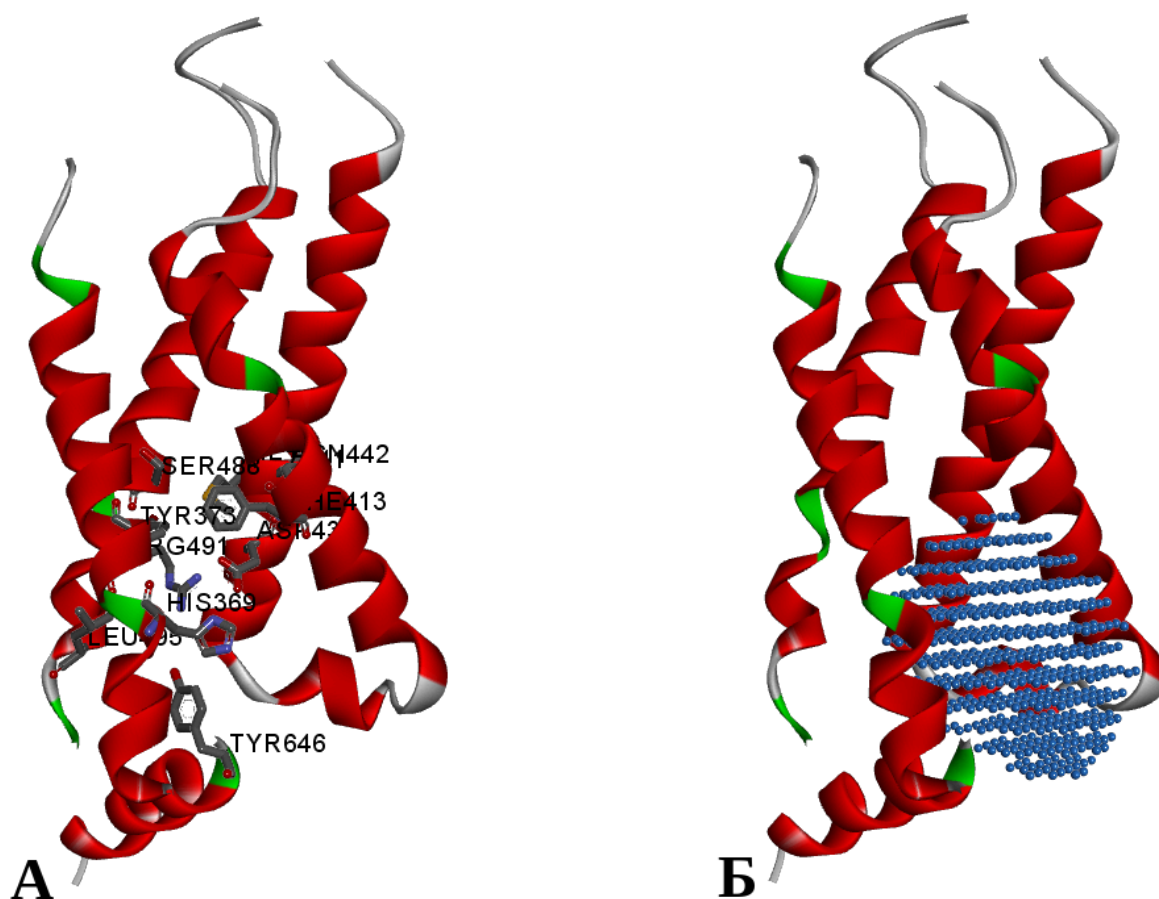


Рис. 3.14. Сайт зв'язування інгібіторів у VSL-домені TRPC4 (структура 5Z96; ланцюг А): А – амінокислотне оточення кишені; Б – об'ємне сіткове представлення порожнини сайту (сині точки).

За допомогою розробленого метода у структурі 5Z96 ідентифіковано чотири унікальні сайти зв'язування TRPC4 миші, два з яких (катион-зв'язувальний і  $\text{Ca}^{2+}$ -зв'язувальний) виконують регуляторну функцію через координацію катіонів у VSL-домені, один (інгібіторна кишеня GFB-сполук) є фармакологічно валідованим сайтом для піридазинонових модуляторів і четвертий (РА-кишеня) визначає ліпідне оточення каналу у мембрані і є сайтом зв'язування для інгібіторів інших хімічних класів. Сайт зв'язування низькомолекулярних інгібіторів у VSL-домені є цільовим сайтом для молекулярного докінгу GFB-887, а РА-сайт – для молекулярного докінгу Pico145.

## РОЗДІЛ 4

### МОЛЕКУЛЯРНЕ МОДЕЛЮВАННЯ ВЗАЄМОДІЇ PICO145 ТА GFB-887 З TRPC4

Незважаючи на те, що TRPC4 розглядається як перспективна терапевтична мішень при низці патологічних станів, структурні основи його взаємодії з фармакологічними модуляторами залишаються переважно невизначеними. Для Pico145 експериментальні структурні дані обмежені кріо-ЕМ структурами гомомерного людського TRPC5 та гетеромерного людського TRPC1/TRPC4, тоді як для гомомерного TRPC4, структура комплексу досі не отримана [10, 11]. Єдина опублікована спроба обчислювальної реконструкції зв'язування Pico145 з мишачим TRPC4 (докінг у структуру 5Z96) виконана лише як допоміжний експеримент і обмежена якісною констатацією подібності пози до TRPC5 без розгорнутого аналізу ліганд-білкових взаємодій, енергетики зв'язування чи додаткової валідації отриманої пози методами молекулярної динаміки [11]. Для GFB-887 структурні дані у комплексі з будь-яким представником підродини TRPC1/4/5 наразі відсутні взагалі, що робить обчислювальне моделювання підходящим інструментом для встановлення його сайту зв'язування та особливостей взаємодії з TRPC4.

#### 4.1. Молекулярний докінг Pico145 до TRPC4

Подібний тип сайту вперше був експериментально ідентифікований у роботі Wright *et al.* [11] при дослідженні комплексу TRPC5 з Pico145, де за допомогою кріо-ЕМ було показано, що ліганд зв'язується у міжсубодиничній гідрофобній кишені трансмембранного домену каналу. У тій самій роботі було висунуто гіпотезу про консервативність цього сайту для представників підродини TRPC1/4/5 на підставі структурної та послідовнісної подібності

відповідних ділянок білка. Подальше підтвердження цієї гіпотези було отримано у структурі гетеротетрамерного комплексу TRPC1/TRPC4 людини [10], де показано, що аналогічний міжсубодиничний сайт зберігається у цьому комплексі та відповідає сайту зв'язування Pico145.

У роботі Wright *et al.* [11] також було здійснено докінг Pico145 у структуру мишачого TRPC4 (PDB ID: 5Z96) [25] з метою перевірки гіпотези про консервативність механізму зв'язування; однак ця частина дослідження була описана стисло, без детального аналізу просторових відповідностей, енергетичних характеристик і структурно-функціональних наслідків взаємодії ліганду з каналом. У зв'язку з цим у даній роботі виконано розширене молекулярне моделювання особливостей зв'язування Pico145 з мишачим TRPC4 [120], результати якого детально розглянуто у цьому підрозділі; відповідні структурні передумови при цьому отримують функціональне та молекулярне підтвердження для мишачого TRPC4.

Першим етапом було проведено молекулярний докінг Pico145 у структуру 8WPM [10] після попереднього видалення ліганду з експериментального комплексу. Метою цього етапу було відтворення експериментально спостереженої пози Pico145 та оцінка здатності обраного докінг-протоколу коректно реконструювати реальну позу зв'язування. У межах молекулярного докінгу було отримано п'ять конформацій Pico145 з різною енергією зв'язування. Фінальний вибір оптимальної моделі здійснювався на основі комбінованої оцінки, що включала мінімальне значення передбаченої енергії зв'язування та візуальну перевірку просторової відповідності отриманої пози ліганду положенню Pico145 у референтній структурі 8WPM. Енергія зв'язування найкращого конформера становить  $-10,3$  ккал/моль. Отримані результати демонструють, що змодельована в процесі докінгу конформація Pico145 практично повністю відтворює експериментально визначену позу ліганду, встановлену за даними кріо-ЕМ для комплексу TRPC1/TRPC4 (PDB ID: 8WPM). Як показано на рис. 4.1, вдокована молекула Pico145 характеризується високим ступенем

просторового накладання з оригінальною позою ліганду, зберігаючи як орієнтацію ксантинового ядра, так і загальну геометрію розміщення у межах зв'язувальної кишені.

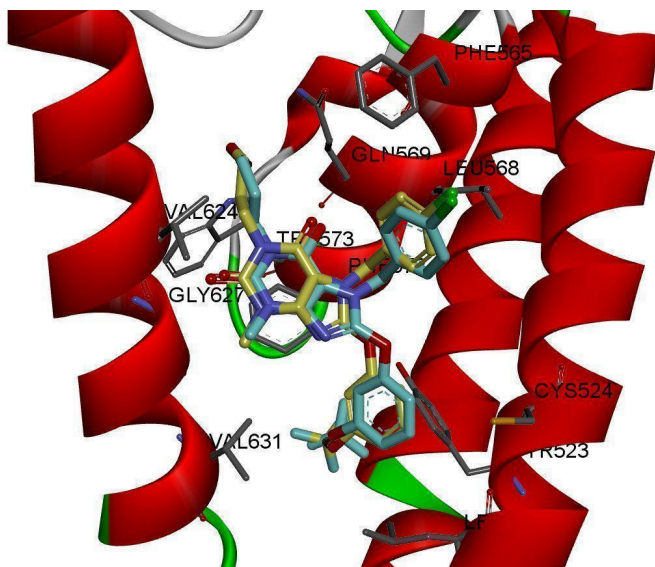


Рис. 4.1. Порівняння поз вдокованого Pico145 та оригінального ліганду з 8WPM у TRPC1/TRPC4. Кольорові позначення: жовтий – вдокований Pico145; світло-блакитний – оригінальна поза Pico145 (PDB ID: 8WPM) [120].

Така відповідність підтверджує коректність застосованого докінг-підходу та свідчить про високу достовірність відтворення механізму взаємодії Pico145 з TRPC-каналом, що є принципово важливим для подальшого перенесення результатів на модель мишачого TRPC4.

Після валідації докінг-протоколу на референтній структурі TRPC1/TRPC4:Pico145 (8WPM) було виконано докінг Pico145 в апо-структуру мишачого каналу TRPC4 (PDB ID: 5Z96) [25]. Вибір саме цієї структури як рецепторної моделі зумовлений її високою роздільністю, повною наявністю трансмембранного домену та відсутністю зв'язаного ліганду, що робить її оптимальною для дослідження механізму інгібування Pico145. Після проведення докінг-розрахунків для структури 5Z96 було згенеровано п'ять найкращих за енергією зв'язування конформацій Pico145. Фінальний вибір оптимальної моделі здійснювався за таким же принципом, як і на попередньому етапі дослідження. Обрана конформація демонструвала

найнижчу енергію зв'язування ( $-9,9$  ккал/моль) та найкраще накладання на експериментальну позу ліганду в кріо-ЕМ комплексі.

Для порівняльного аналізу просторової організації сайту зв'язування було виконано суперпозиціонування структур 5Z96 та 8WPM. Як показано на рис. 4.2, сайт, у якому зв'язується Pico145 у гетеротетрамерному комплексі 8WPM, просторово повністю узгоджується з відповідною областю у мишачій структурі TRPC4.

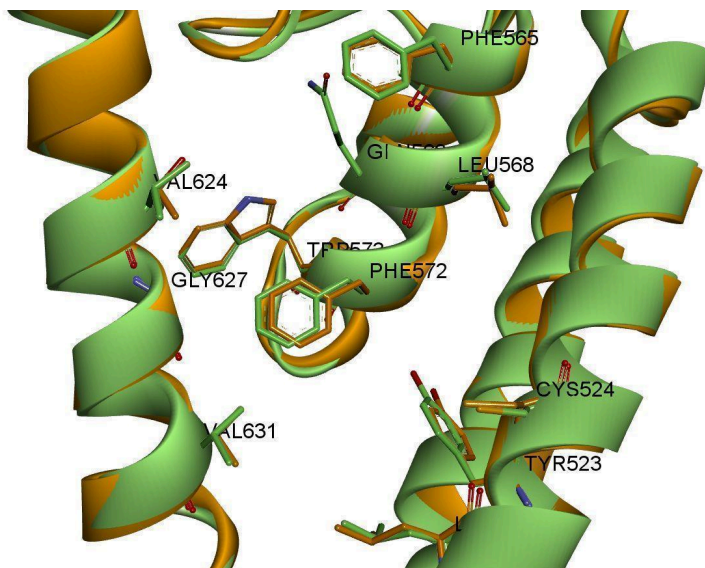


Рис. 4.2. Суперпозиціонування структур TRPC1/TRPC4 (колір: зелений; PDB ID: 8WPM) та TRPC4 (колір: помаранчевий; PDB ID: 5Z96) [120].

Геометрія кишені, орієнтація трансмембранних спіралей та локалізація ключових контактних залишків свідчать про високий ступінь структурної консервативності даного сайту між людськими та мишачими каналами.

Подальший докінг Pico145 у структуру 5Z96 продемонстрував стабільне зв'язування ліганду в тій самій міжсубодиничній гідрофобній кишені, яка була експериментально ідентифікована у референтній структурі 8WPM. Найкраща отримана модель відтворює характерну орієнтацію ксантинового ядра Pico145 та основні взаємодії з амінокислотними залишками сайту. На рис. 4.3 показано отриманий комплекс TRPC4:Pico145 з візуалізацією взаємодій з амінокислотними залишками.

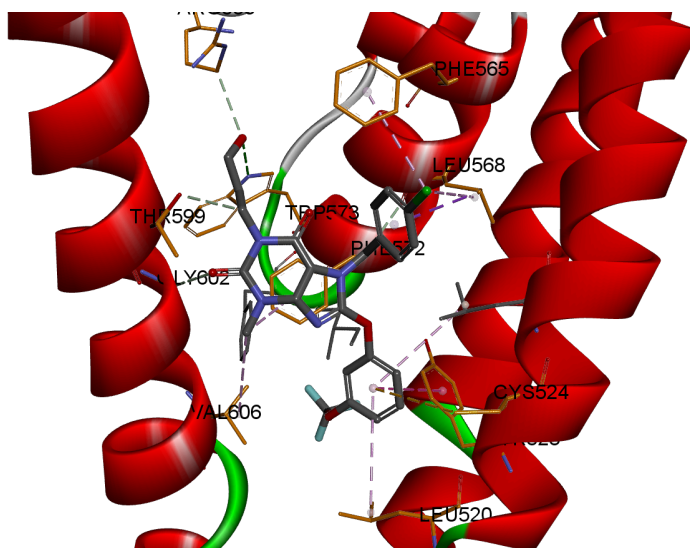


Рис. 4.3. Молекулярна модель комплексу Pico145 з мишачим TRPC4 [120].

Молекулярна модель комплексу Pico145 з мишачим TRPC4 (PDB ID: 5Z96), отримана в результаті докінгу, демонструє стабілізацію ліганду в міжсубодиничній гідрофобній кишеньці за рахунок поєднання полярних та гідрофобно-ароматичних взаємодій. Зокрема, полярний фрагмент Pico145 формує конвенційний водневий зв'язок з Trp573, який задає орієнтацію молекули в сайті. Ароматичний фрагмент ліганду додатково фіксується через  $\pi$ - $\pi$  (T-shaped) взаємодію з Tyr523 та  $\pi$ -sulfur контакт з Cys524, тоді як гідрофобне оточення кишеньці формується залишками Leu520, Leu527, Leu568, Phe565 та Phe572 (Alkyl/ $\pi$ -Alkyl/ $\pi$ - $\sigma$  взаємодії). Додаткові слабкі контакти (van der Waals) з Thr599, Gly602, Thr603 та Val606 підтримують щільне заповнення кишеньці лігандом.

Подальший аналіз відповідностей між амінокислотними залишками, що формують оточення кишеньці, у моделях 5Z96 та 8WPM дозволяє детальніше оцінити ступінь консервативності сайту зв'язування Pico145. Гідрофобні залишки Gly602 та Val606 мишачої структури 5Z96 повністю консервативні – їм у референтній структурі 8WPM відповідають ідентичні Gly627 та Val631. Натомість позиція Thr599 у 5Z96 відповідає залишку Val624 у 8WPM, що є напівконсервативною заміною полярного треоніну на гідрофобний валін.



Thr599 безпосередньо контактує з лігандом: у докінговій позі він входить до оточення, що щільно заповнює кишеню, а його гідроксильна група належить до полярних донорів/акцепторів, здатних утворювати водневі зв'язки з Pico145. Тому заміна Thr599 на Val624 відображає видоспецифічну відмінність у полярній складовій контактної мережі: у мишачому каналі в цій позиції можливе утворення водневого зв'язку, тоді як у людській структурі відповідний залишок гідрофобний і такої взаємодії не забезпечує. Водночас, з огляду на те, що Pico145 зв'язується і з людським каналом, зі структури якого отримано 8WPM, попри цю заміну, зазначений контакт доцільно розглядати як додатковий, що підсилює зв'язування, а не як визначальний. Загальна архітектура сайту та його ключові контакти між видами зберігаються, а сама заміна може модулювати локальну полярність середовища та динаміку бічних ланцюгів у цій ділянці кишені.

Таким чином, аналіз показує, що механізм зв'язування Pico145 є загально консервативним: гідрофобне ядро кишені та ключові контакти ліганду зберігаються між структурами 5Z96 і 8WPM. Виявлені відмінності стосуються не визначальних для зв'язування елементів, а додаткової полярної складової контактної мережі (позиція Thr599 на Val624), і мають характер допустимих міжвидових варіацій, які не порушують загальної архітектури та функціональної цілісності сайту.

## **4.2. Молекулярна динаміка комплексу TRPC4:Pico145**

Після визначення геометрії зв'язування Pico145 у міжсубодиничній кишені TRPC4 методом молекулярного докінгу, подальшим кроком була незалежна перевірка стабільності одержаної моделі в умовах, що наближають симуляцію до фізіологічного оточення іонного каналу. Для цього було проведено розрахунок МД комплексу TRPC4:Pico145 у мембранному середовищі. Мета цього етапу була двоскладовою: по-перше, оцінити

енергетичну стабільність системи та конформаційну стабільність білкової основи впродовж тривалої траєкторії, і по-друге, перевірити, чи зберігається докінгова поза ліганду під дією теплових флуктуацій та колективних рухів backbone білка.

Побудована модельна система відтворює гомотетрамерну організацію TRPC4 (PDB ID: 5Z96), у якій чотири еквівалентні міжсубодиничні гідрофобні кишені симетрично зайняті чотирма копіями Pico145 (для стислості далі – PICO 1 (ланцюг А), PICO 2 (ланцюг В), PICO 3 (ланцюг С) та PICO 4 (ланцюг D)). Така конфігурація використана для того, щоб здійснити аналіз чотирьох формально еквівалентних траєкторій одного й того самого типу зв'язування в межах єдиної симуляції, що істотно підвищує статистичну вагу результатів. Деталі протоколу – вибір силового поля, підготовка бішару POPC, сольватація, схема врівноваження, тривалість і параметри продукційного етапу – викладено у Розділі 2; тут наведено лише ті з них, що необхідні для інтерпретації результатів. Продукційний етап симуляції тривав  $\approx 215$  нс у NPT-ансамблі.

Першим індикатором валідності підготовленої моделі є поведінка повної та потенційної енергій упродовж продукційного етапу. Як видно з рис. 4.4, усі ключові термодинамічні параметри – потенційна, кінетична та повна енергії, температура і тиск від початку траєкторії флуктуюють навколо сталих середніх значень без помітних дрейфів чи систематичних змін.

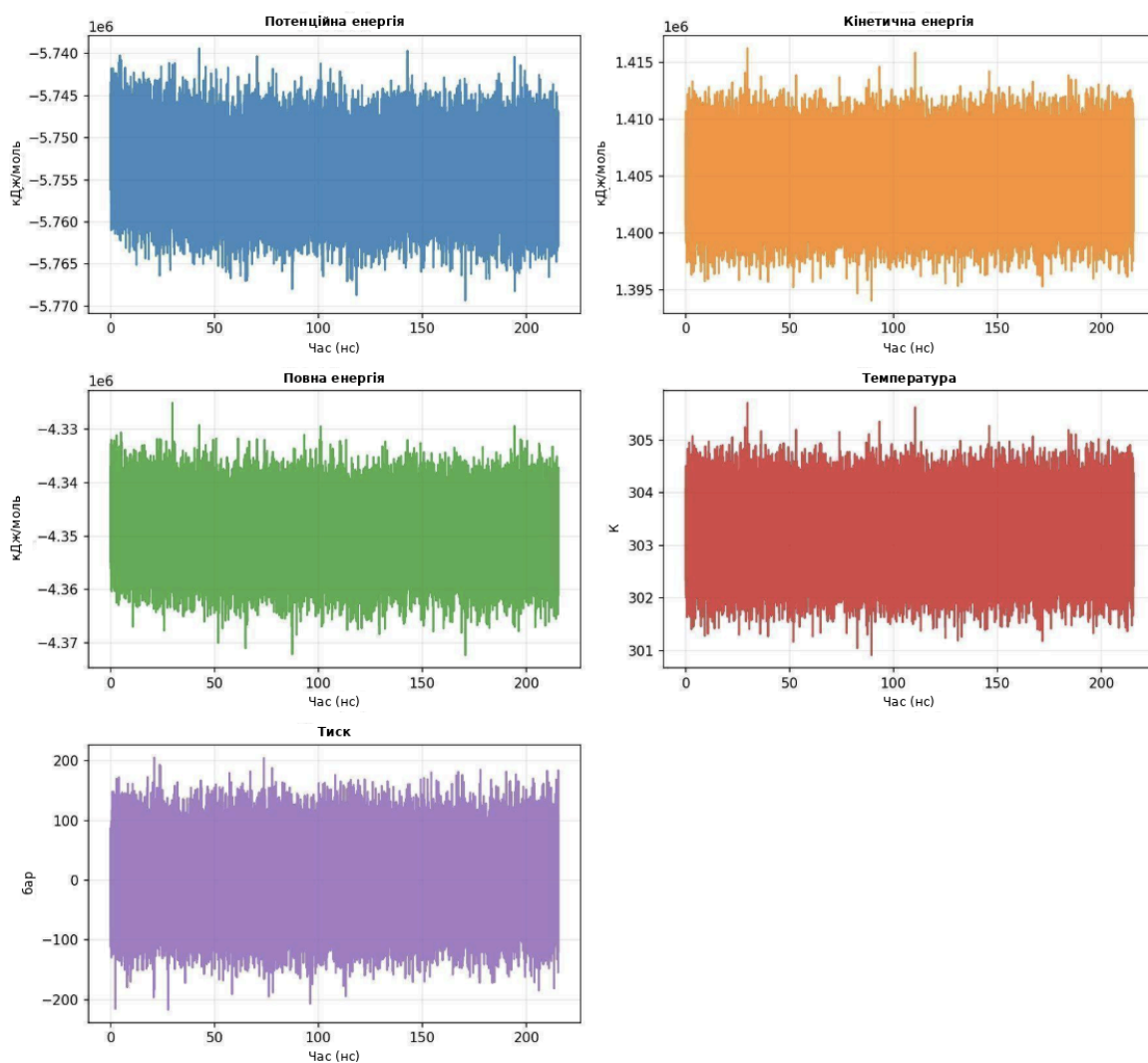


Рис. 4.4. Енергетичні та термодинамічні параметри системи TRPC4:Рісо145 впродовж продукційного етапу молекулярної динаміки: потенційна, кінетична та повна енергії (кДж/моль), температура (К) і тиск (бар).

Потенційна енергія коливається в діапазоні приблизно  $-5,77 \dots -5,74 \times 10^6$  кДж/моль, кінетична енергія близько  $1,40 \times 10^6$  кДж/моль, повна енергія у діапазоні  $-4,37 \dots -4,33 \times 10^6$  кДж/моль. Температура стабільно утримується поблизу заданої точки з короткохвильовими флуктуаціями в межах кількох кельвінів, а тиск коливається навколо нуля з амплітудою  $\approx 200$  бар, що є типовою картиною для баростатованих симуляцій у NPT-ансамблі в мембранних системах такого розміру. Відсутність монотонних дрейфів у будь-якій з панелей і рівномірна дисперсія сигналу вздовж усієї осі часу свідчать про те, що система до моменту початку продукційного етапу

повністю врівноважена, а її термодинамічний стан зберігається протягом усієї траєкторії. Це створює необхідну передумову для коректної структурної інтерпретації подальших метрик.

Структурна стабільність конформації ліганду оцінена за RMSD важких атомів ліганду, обчисленим окремо для кожної з чотирьох копій Pico145 після структурного вирівнювання на основний ланцюг (backbone) відповідної субодиниці. Результати представлено на рис. 4.5 як композит із чотирьох панелей (А–Г), що відповідають PICO 1 – PICO 4.

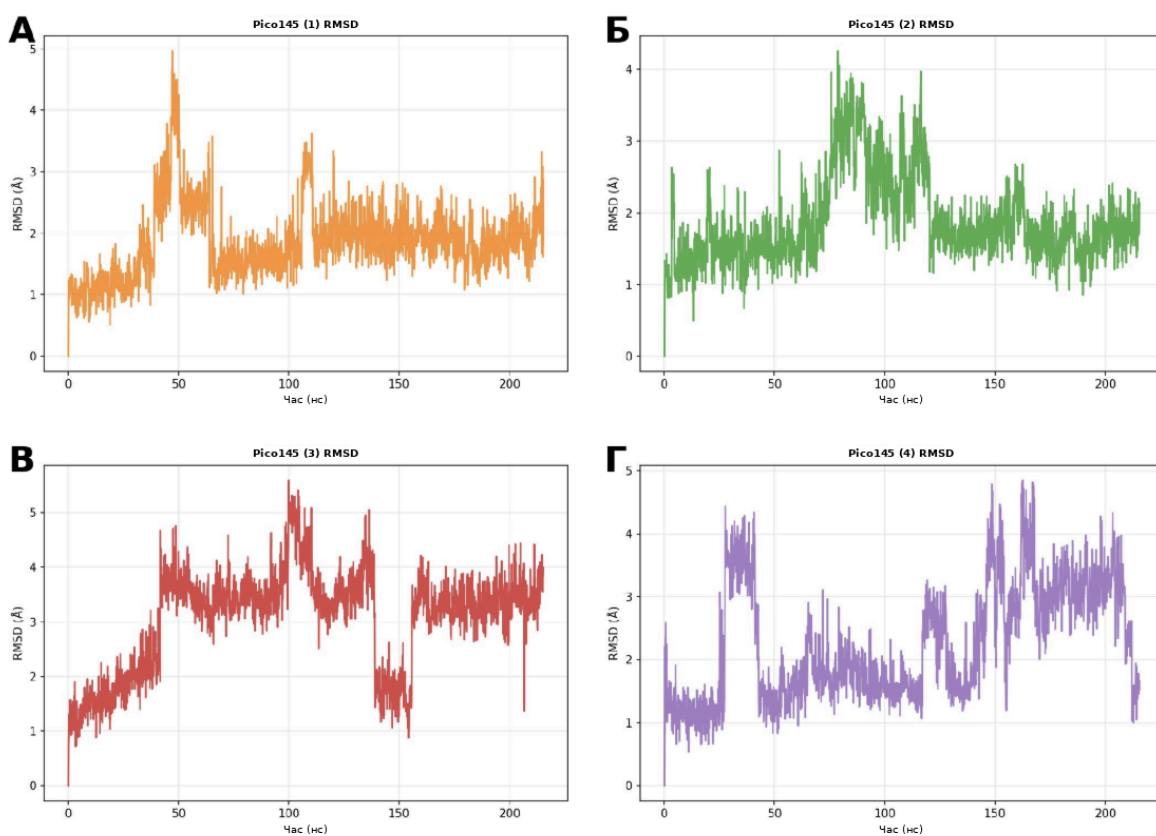


Рис. 4.5. Середньоквадратичне відхилення важких атомів чотирьох копій Pico145 у структурі TRPC4 протягом продукційного етапу МД: А – PICO 1; Б – PICO 2; В – PICO 3; Г – PICO 4.

PICO 1 (панель А) та PICO 2 (панель Б) демонструють якісно подібну поведінку: після короткої початкової стадії перебудови з вихідної докінгової пози впродовж перших  $\approx 5$  нс RMSD виходить на стаціонарний рівень у діапазоні приблизно 1,5...2,5 Å і утримується в цьому діапазоні до кінця

симуляції. Спостерігаються короточасні епізоди підвищення RMSD: для PICO 1 пік до  $\approx 5$  Å поблизу 50 нс, для PICO 2 ділянка підвищених значень у діапазоні  $\approx 75\ldots 100$  нс з максимумами близько 4 Å та поверненням до базового рівня приблизно до 120 нс. Після цих відрізків, траєкторії повертаються до базового рівня, тобто ліганд відтворює вихідний режим зв'язування без виходу з кишені. Отриманий діапазон  $1,5\ldots 2,5$  Å і повернення траєкторії у нього після відхилень свідчать про узгодження з отриманим просторовим розміщенням внаслідок докінгу.

PICO 3 (панель В) та PICO 4 (панель Г) демонструють суттєво вищу конформаційну рухливість. У випадку PICO 3 RMSD уже через  $\approx 40$  нс виходить на плато в діапазоні  $3\ldots 4$  Å, яке утримується майже до кінця траєкторії, з локальним максимумом  $\approx 5,5$  Å поблизу 100 нс та короточасним поверненням до рівня  $1,5\ldots 2$  Å між 140 і 155 нс. PICO 4 характеризується ще більш нерегулярним профілем: ділянки зі стабільним RMSD у діапазоні  $1,2\ldots 1,8$  Å чергуються з епізодами різкого підйому до  $3\ldots 4,5$  Å (поблизу 30–40 нс, 120 нс і особливо 150–200 нс). Форма цих стрибків є характерною: підвищення RMSD відбувається дискретно, у вигляді «сходинок», а не у вигляді плавного дрейфу. Такий патерн у термінах молекулярної динаміки типово відповідає дискретним конформаційним переходам окремих фрагментів ліганду, наприклад поворотам навколо обертальних зв'язків або фліпам ароматичних кілець, що відбуваються на тлі збереженого положення в кишені.

Ця інтерпретація узгоджується з будовою Pico145, який містить трифторметоксифенільний фрагмент із низьким бар'єром обертання навколо зв'язку C–C (ароматичний); такий поворот на  $180^\circ$  у симетричному за розмахом оточенні кишені еквівалентний, по суті, ізоенергетичному переходу, однак у координатах важких атомів дає значний вклад у миттєве значення RMSD. Відповідно, продемонстрований на панелі В, профіль PICO 3, добре пояснюється чергуванням двох переважних орієнтацій цього фрагмента. Для PICO 4 до цього механізму додатково накладаються рідші фліпи

ароматичного кільця, що зумовлюють наявність ділянок траєкторії з високим значенням RMSD, нетипових для суто обертального руху. У жодному з цих випадків ліганд не залишає кишені: після кожного конформаційного переходу RMSD повертається у характерний діапазон, а не зростає монотонно, як це було б у випадку дисоціації.

Слід зазначити, що всі чотири копії Pico145 стартували з ідентичних координат у еквівалентних за симетрією кишенях, отже за достатньої тривалості траєкторії їхня статистична поведінка має зійтися до спільного розподілу конформацій. Систематична розбіжність профілів RMSD між парами PICO 1/PICO 2 та PICO 3/PICO 4 вказує на те, що на масштабі  $\approx 215$  нс симетрія поведінки ліганду між чотирма еквівалентними сайтами ще не досягнута. Для виходу окремих копій на узгоджений стаціонарний режим доцільне продовження траєкторії.

Додатковою характеристикою якості зв'язування є кількісний і геометричний аналіз водневих зв'язків між Pico145 і залишками кишені, проведений за всіма чотирма лігандами з усередненням по копіях. Результати подано на рис. 4.6.

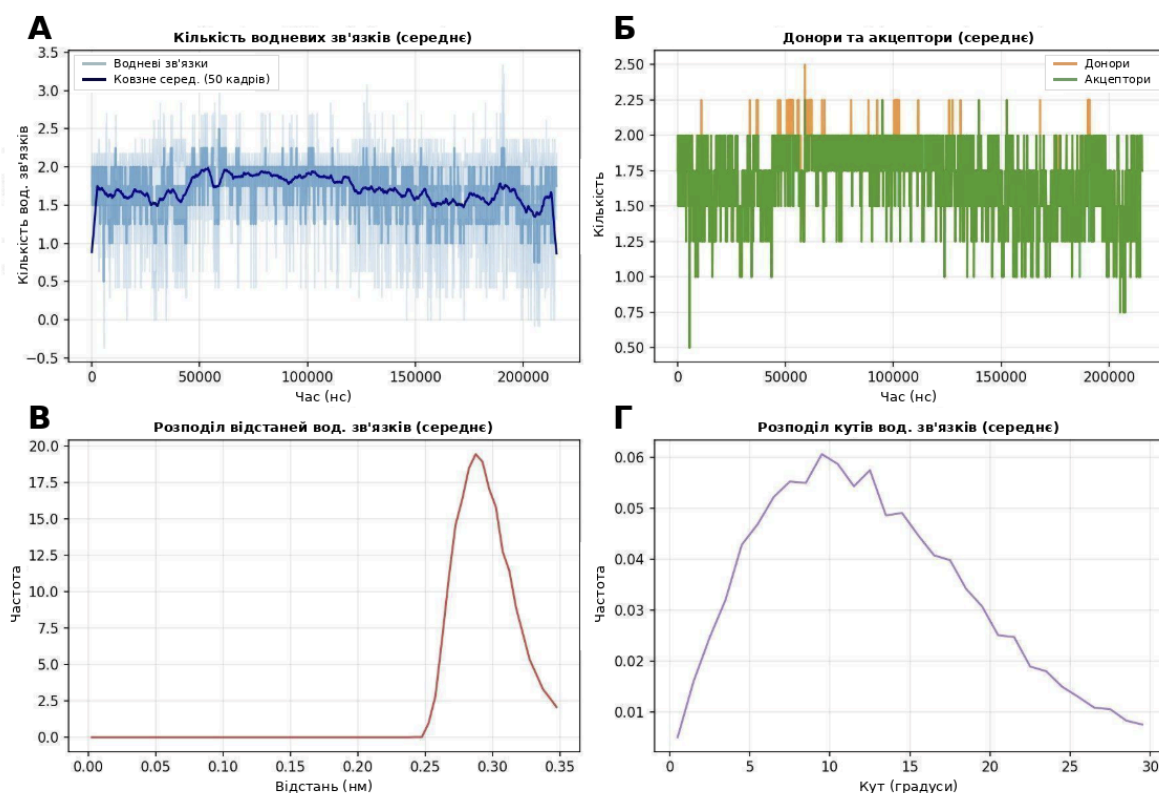


Рис. 4.6. Аналіз водневих зв'язків у комплексі TRPC4:Рісо145 протягом 215 нс МД-симуляції, усереднений по чотирьох копіях ліганду: А – середня кількість водневих зв'язків у часі (світло-синій – миттєві значення; темно-синій – ковзне середнє по 50 кадрах); Б – середня кількість донорних та акцепторних атомів у часі; В – розподіл відстаней донор–акцептор; Г – розподіл відхилень водневих зв'язків від лінійності. Шкала часу на панелях А та Б наведена в пікосекундах (215 000 пс = 215 нс).

Часовий профіль середньої кількості водневих зв'язків на ліганд (панель А) утримується на рівні  $\approx 1,5 \dots 2,0$  впродовж усієї траєкторії; згладжена ковзним вікном крива (50 кадрів) демонструє лише дрібні флуктуації навколо цього рівня без ознак спадного тренду. Кількість донорних і акцепторних водневих зв'язків (панель Б) змінюється синхронно, зі схожим характером часових профілів, що вказує на скорельовану динаміку полярних контактів Рісо145 і узгоджену участь ліганду у водневих зв'язках як донора та акцептора протона – картина, характерна для ліганду, що утримується в кишені стійкою мережею полярних взаємодій.

Геометричний аналіз підсилює цей висновок. Розподіл відстаней донор–акцептор (панель В) утворює вузький пік в області 0,27–0,29 нм; переважна частина розподілу лежить нижче 0,35 нм, у межах геометричного критерію ідентифікації водневого зв'язку, прийнятого у цьому дослідженні (див. Розділ 2). Розподіл кутів відхилення донор–H $\cdots$ акцептор (панель Г) має максимум в області 8–10° і неширокий хвіст до  $\approx$ 30°, що відповідає геометрично сприятливій, близькій до колінеарної орієнтації зв'язків. Сукупно ці розподіли свідчать, що водневі зв'язки, які утворює Pico145 у кишені TRPC4, мають сприятливу геометрію і не є артефактом порогу відсікання: положення піків помітно відступає від граничних значень критерію та відтворює геометрію, характерну для повноцінних водневих зв'язків помірної сили.

Цей результат варто зіставити з переліком контактів, встановлених у докінговій моделі. Серед амінокислотних залишків, що формують сайт зв'язування, частина має полярні функціональні групи, здатні виступати донорами та/або акцепторами протона, зокрема Trp573 (індольний NH, донор), Tyr523 (фенольний OH), Thr599 і Thr603 (гідроксильні групи). Середня кількість водневих зв'язків (1,5–2,0 на траєкторію) добре узгоджується з наявністю у кишені кількох потенційних полярних партнерів і підтверджує, що контакти, виявлені у докінговій моделі, мають не виключно гідрофобну природу.

Комплексний аналіз  $\approx$ 215-наносекундної траєкторії МД системи TRPC4:Pico145 показує, що побудована модель є врівноваженою та структурно стабільною. Повна і потенційна енергії, температура й тиск залишаються на плато впродовж усього продукційного етапу і не демонструють дрейфів, що підтверджує коректність схеми врівноваження та придатність траєкторії для кількісного аналізу. Усі чотири копії ліганду зберігаються у відповідних кишенях, що свідчить про відтворюваність докінгової пози в умовах повноатомної МД. Принаймні у двох копіях (PICO 1 і PICO 2) RMSD стабілізується на рівні 1,5–2,5 Å без ознак виходу ліганду з



кишені, що вказує на повну конформаційну стабілізацію. У двох інших копіях (PICO 3 і PICO 4) ліганд також залишається в кишені, однак демонструє дискретні конформаційні переходи обертальних та ароматичних фрагментів, що свідчить про недостатність часу симуляції для повного відсемплювання внутрішніх конформаційних станів ліганду; цей аспект потребує продовження траєкторії у подальших дослідженнях. Профіль водневих зв'язків – стабільний у часі, зі сприятливою геометрією та типовою кількістю  $\approx 1,5\text{--}2,0$  контактів на ліганд, додатково підтверджує, що докінгова поза Pico145, одержана у підрозділі 4.1, є фізично релевантною.

### 4.3. Молекулярний докінг GFB-887 до TRPC4

Наступним етапом обчислювального дослідження взаємодії GFB-887 з каналом TRPC4 став молекулярний докінг цієї сполуки до структури TRPC4 миші (PDB ID: 5Z96). Принципова відмінність від докінгу Pico145 полягає у виборі сайту зв'язування: для GFB-887 як цільовий регіон обрано не міжсубодиничну гідрофобну кишеню, а трансмембранний сайт у межах voltage-sensor-like домену кожної субодиниці, сформований сегментами S1–S4.

Зокрема, для подальшого молекулярного докінгу було обрано трансмембранний сайт зв'язування, локалізований у регіоні між сегментами S1 та S4, який демонстрував найвищу узгодженість між геометричним прогнозуванням і літературними даними (рис. 4.7).

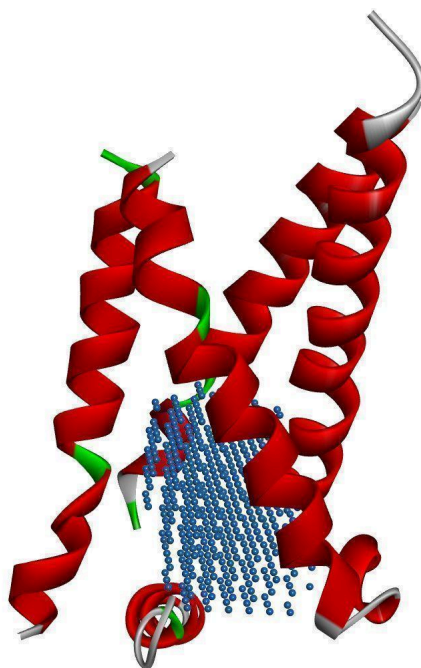


Рис. 4.7. Об'єм ідентифікованого сайту зв'язування в регіоні S1–S4 каналу TRPC4 (PDB ID: 5Z96), представлений у вигляді сіткового представлення (позначено синім кольором) [121].

Набір амінокислотних залишків, що визначають цей сайт, був витягнутий з літературних джерел та картований на структуру TRPC4 (PDB ID: 5Z96) і включає: His369, Tyr373, Phe413, Asp438, Met441, Asn442, Ser488, Arg491, Leu495 та Tyr646 (рис. 4.8).

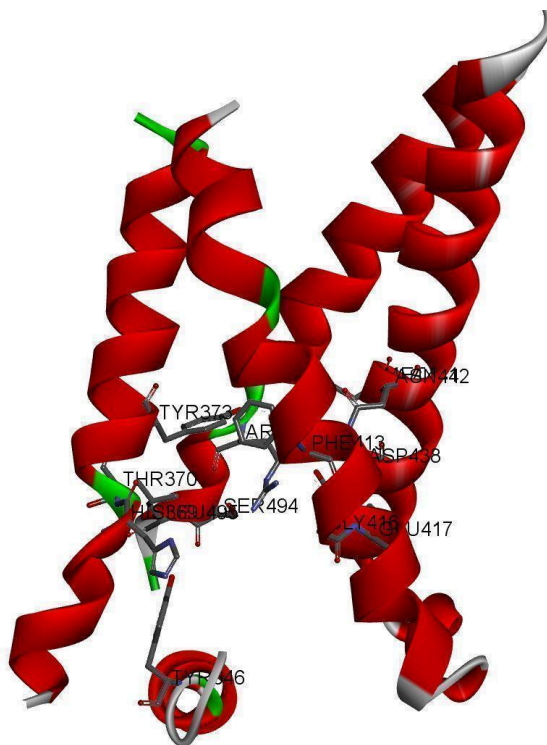


Рис. 4.8. Ідентифікований сайт зв'язування в регіоні S1–S4 каналу TRPC4 (PDB ID: 5Z96), використаний для проведення молекулярного докінгу [121].

Просторова локалізація ідентифікованого сайту зв'язування та його амінокислотний склад узгоджуються з даними крио-електронної мікроскопії TRPC4 у комплексі з малими молекулами-інгібіторами, опублікованими раніше. Зокрема, у крио-ЕМ структурах TRPC4 у комплексі з інгібіторами GFB-8438, GFB-9289 та GFB-8749 (PDB ID: 7B0S, 7B16 та 7B05 відповідно) показано, що зв'язування лігандів відбувається у трансмембранній області каналу, сформованій сегментами S1–S4, яка відповідає VSL домену [27]. У цих структурах густина ліганду локалізується на внутрішній поверхні трансмембранного домену, де взаємодії з бічними ланцюгами амінокислот стабілізують неактивну конформацію каналу та зумовлюють алостеричний механізм інгібування.

Для молекулярного докінгу сполуки GFB-887 як рецепторну модель було обрано апо-структуру мишачого каналу TRPC4 (PDB ID: 5Z96). Вибір

саме цієї структури зумовлений її високою роздільною здатністю, повною реконструкцією трансмембранного домену та відсутністю зв'язаного ліганду, що мінімізує структурні упередження та забезпечує коректний аналіз потенційного сайту зв'язування. Крім того, ця структура широко використовується як репрезентативна модель TRPC4 у структурно-функціональних дослідженнях.

У результаті докінгу GFB-887 у структуру 5Z96 було згенеровано п'ять найбільш енергетично сприятливих поз ліганду. Фінальний комплекс TRPC4:GFB-887 було обрано на основі сукупності критеріїв, що включали мінімальну енергію зв'язування, просторову узгодженість із попередньо ідентифікованим сайтом зв'язування та відповідність відомим експериментальним даним щодо локалізації інгібіторів TRPC4 у трансмембранному домені. Обрана докінгова поза GFB-887 характеризувалася вільною енергією зв'язування  $-8,11$  ккал/моль та демонструвала стабільне розміщення ліганду в запропонованому сайті зв'язування (рис. 4.9).

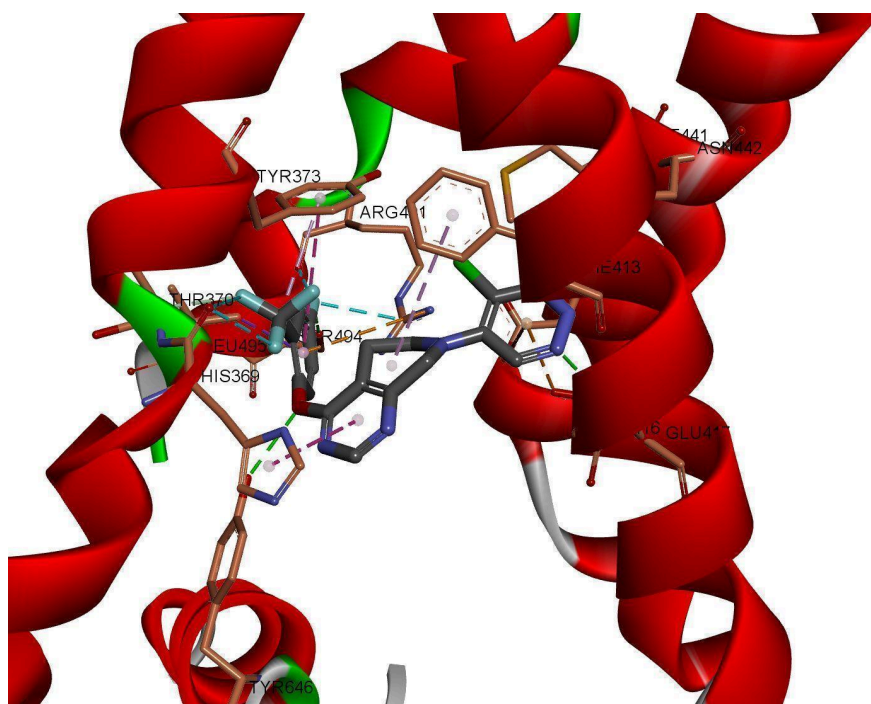


Рис. 4.9. Модель зв'язування GFB-887 з TRPC4 миші у регіоні S1–S4 (PDB ID: 5Z96) з позначеними ключовими амінокислотними залишками та міжмолекулярними взаємодіями [121].

У сформованому комплексі TRPC4:GFB-887 ліганд розташовується між спіралями трансмембранного домену, займаючи конформацію, що забезпечує щільне заповнення кишені. Гетероциклічне ядро GFB-887 орієнтоване у бік залишків Phe413, Tyr373 та Tyr646, формуючи стабілізуюче гідрофобно-ароматичне оточення. Фіксація ліганду в цій області забезпечується поєднанням  $\pi$ - $\pi$  stacking,  $\pi$ - $\pi$  T-shaped,  $\pi$ -alkyl та  $\pi$ - $\sigma$  взаємодій з ароматичними й аліфатичними фрагментами бічних ланцюгів, а також ван-дер-ваальсовими контактами.

Додаткову стабілізацію та орієнтацію молекули у кишені забезпечує система полярних і електростатичних контактів із залишками His369, Glu417, Asp438, Asn442, Arg491 та Ser494. Залежно від локальної геометрії та протонаційного стану бічних ланцюгів ці взаємодії реалізуються у вигляді конвенційних водневих зв'язків, а також  $\pi$ -cation та  $\pi$ -anion взаємодій за участі заряджених залишків. Окрім цього, у стабілізації комплексу відіграють роль гідрофобні залишки Met441, Leu492 та Leu495, які формують неполярне мікросередовище навколо відповідних фрагментів GFB-887 і беруть участь переважно в утворенні ван-дер-ваальсових та  $\pi$ -alkyl контактів. Dodatkowo, наявність атомів флуору в структурі ліганду забезпечує додаткові ван-дер-ваальсові контакти із залишками кишені, що може давати внесок у локальну стабілізацію комплексу в трансмембранній кишені.

Отримані дані узгоджуються з кріо-ЕМ структурами комплексів TRPC4 *Danio rerio* із інгібіторами серії GFB (PDB ID: 7B05, 7B0S, 7B16), які взаємодіють у сайті в межах VSL-домену, що просторово межує з TRP-спіраллю [27]. У цих структурах показано, що інгібітори даного хемотипу займають замкнену порожнину сайту та контактують із залишками Tyr373, Phe413, Met441, Asn442, Arg491, Leu492, Leu495 та Tyr646. Локалізація GFB-887 у цій ділянці є просторово сумісною з описаним у літературі положенням інгібіторів GFB-серії у структурах TRPC4. Додатковим підтвердженням отриманих результатів є також дослідження, в якому було представлено кріо-ЕМ структуру TRPC5 людини в комплексі з

інгібітором клемізол: встановлено, що клемізол взаємодіє із внутрішньою порожниною VSL-домену кожної субодиниці та стабілізує закриту конформацію каналу при фіксації у цій кишені [32].

Порівняння отриманої докінгової моделі TRPC4:GFB-887 з моделлю TRPC4:Pico145 демонструє, що обидва інгібітори локалізуються у трансмембранному домені каналу, однак займають структурно різні кишені. Pico145 зв'язується в міжсубодиничній гідрофобній порожнині, утвореній переважно неполярними залишками сусідніх субодиниць, тоді як GFB-887 розміщується у внутрішньосубодиничному сайті VSL-домену одиначної субодиниці. Така відмінність у просторовій локалізації сайтів зв'язування відповідає двом різним молекулярним механізмам інгібування TRPC4 і пояснює розбіжності в профілях амінокислотних контактів двох інгібіторів.

#### **4.4. Молекулярна динаміка комплексу TRPC4:GFB-887**

Стабільність отриманого докінгового комплексу TRPC4:GFB-887 та збереження ключових взаємодій між лігандом і рецептором в реалістичному мембранному оточенні оцінювали методом симуляції МД. Протокол підготовки системи, параметризації, врівноваження та продукційного етапу МД детально описаний у Розділі 2 й аналогічний тому, що був застосований для комплексу TRPC4:Pico145. Симуляцію виконували для повної тетрамерної конфігурації каналу, вбудованої у POPC-бішар, з чотирма копіями GFB-887, попередньо розміщеними у чотирьох симетричних VSL-сайтах відповідно до результатів докінгу. Тривалість продукційного етапу NPT-симуляції становила 200 нс.

Динаміку основних термодинамічних параметрів системи протягом продукційного етапу подано на рисунку 4.10.

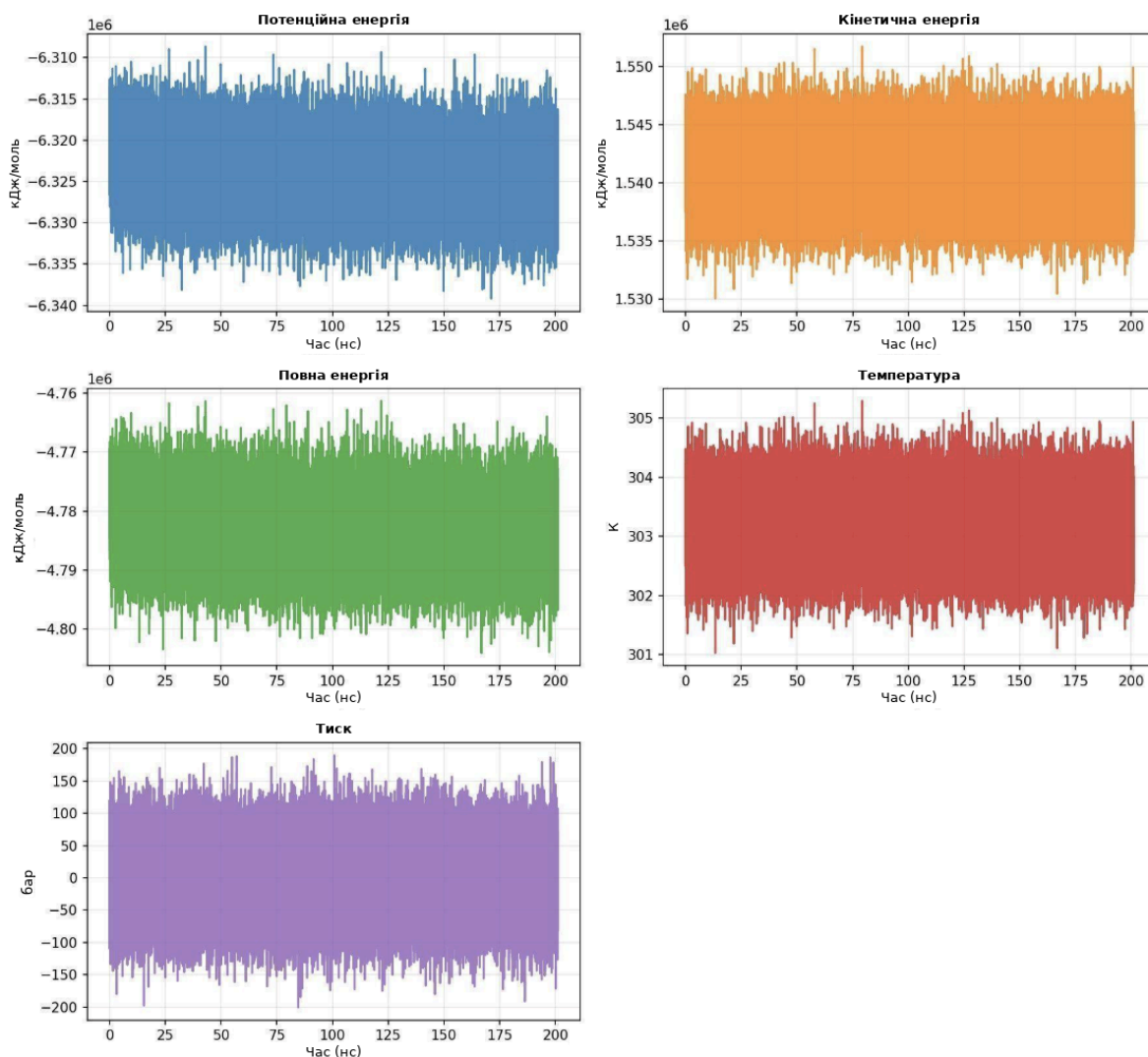


Рис. 4.10. Динаміка термодинамічних параметрів системи TRPC4:GFB-887 протягом 200 нс продукційного етапу NPT-симуляції: потенційна енергія, кінетична енергія, повна енергія, температура та тиск.

Повна енергія системи зберігає стабільний рівень упродовж усієї траєкторії (діапазон  $-4,76 \dots -4,80 \times 10^6$  кДж/моль) з вузькою амплітудою термічних флуктуацій без помітного дрейфу, що відповідає виходу системи на плато та її врівноваженню. Потенційна та кінетична енергії демонструють аналогічну картину: їх середні значення не змінюються в межах траєкторії, а флуктуації мають характер випадкових відхилень від стаціонарного середнього. Температура утримується в діапазоні 301–305 К з середнім значенням, близьким до цільових 303 К, а тиск коливається навколо нуля в

діапазоні  $\pm 200$  бар, що є типовим для миттєвих значень тиску у NPT-ансамблі за умови ефективного контролю з боку баростату. Сукупно це свідчить про коректне функціонування термостата й баростата та про термодинамічну стаціонарність системи протягом усієї симуляції.

Структурну стабільність чотирьох копій ліганду оцінювали за профілем RMSD важких атомів кожної копії GFB-887 відносно її стартового положення після структурного вирівнювання (рис. 4.11).

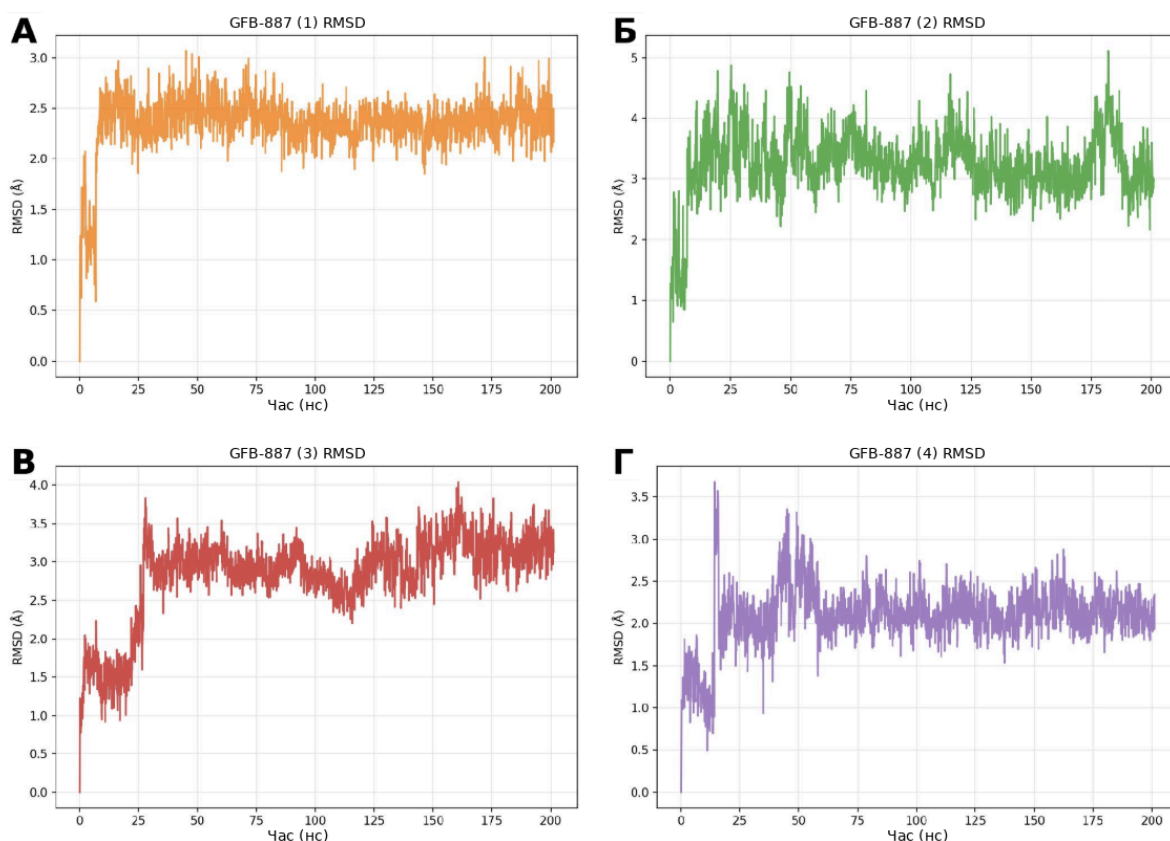


Рис. 4.11. Середньоквадратичне відхилення чотирьох копій GFB-887, зв'язаних у симетричних VSL-сайтах тетрамера TRPC4, упродовж 200 нс МД-симуляції: А – копія 1; Б – копія 2; В – копія 3; Г – копія 4.

Упродовж перших 5–10 нс усі чотири копії виходять на плато, що відповідає етапу релаксації ліганду в кишені після початкового розміщення, отриманого внаслідок докінгу.

Після виходу на плато копії 1 (панель А) та 4 (панель Г) демонструють найстабільнішу поведінку: значення RMSD коливаються навколо 2,0–2,5 Å та



2,0–2,2 Å відповідно, з вузькою амплітудою флуктуацій протягом решти траєкторії. Копії 2 та 3 (панелі Б і В) характеризуються дещо вищими значеннями RMSD у діапазоні 3,0–3,5 Å та більшою амплітудою флуктуацій, що вказує на більшу конформаційну рухливість ліганду у відповідних сайтах. Для копії 3 спостерігається ступеневе зростання RMSD приблизно на 25 нс симуляції (панель В), яке відповідає локальній конформаційній перебудові ліганду в кишені з переходом до нового стабільного рівня без подальшого дрейфу. У всьому досліджуваному часовому вікні жодна з чотирьох копій GFB-887 не залишає свого сайту зв'язування, а значення RMSD залишаються у діапазоні, що типовий для стабільно зв'язаних лігандів у гідрофобних трансмембранних кишенях.

Додатково було проаналізовано поведінку водневих зв'язків між GFB-887 та оточення сайту зв'язування протягом траєкторії (рис. 4.12).

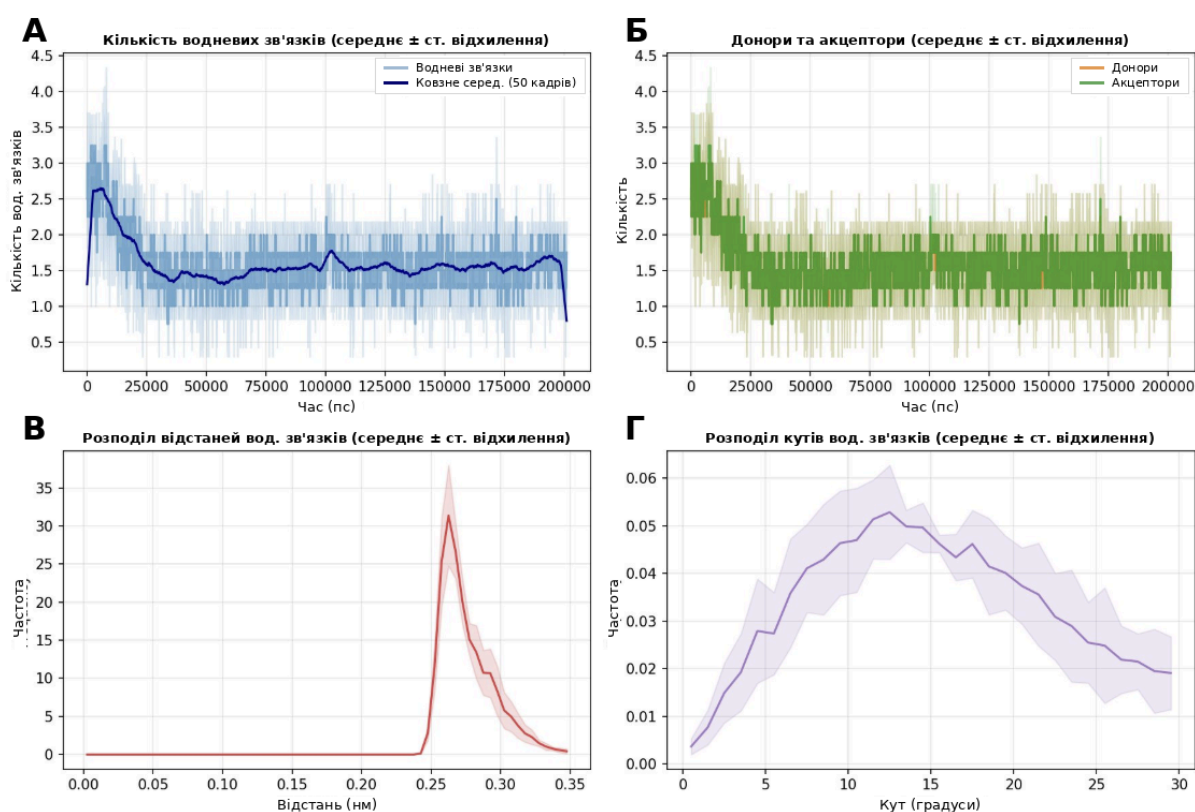


Рис. 4.12. Аналіз водневих зв'язків у комплексі TRPC4:GFB-887 протягом 200 нс МД-симуляції, усереднений по чотирьох копіях ліганду: А – середня кількість водневих зв'язків у часі (світло-синій – миттєві значення;

темно-синій – ковзне середнє по 50 кадрах); Б – середня кількість донорних та акцепторних атомів у часі; В – розподіл відстаней донор–акцептор; Г – розподіл відхилень водневих зв'язків від лінійності. Шкала часу на панелях А та Б наведена в пікосекундах ( $200\,000\text{ ps} = 200\text{ ns}$ ).

Після початкової релаксаційної фази (перші  $\sim 25\text{ ns}$ ), під час якої середня кількість водневих зв'язків транзитно досягає значення близько 3, профіль виходить на стабільний рівень у діапазоні 1,5–2,0 зв'язків на один комплекс і зберігається таким до кінця симуляції (панель А). Кількість залучених донорних та акцепторних атомів підтримується в аналогічному діапазоні й не демонструє дрейфу (панель Б).

Геометричні характеристики цих водневих зв'язків відповідають критеріям сильних контактів. Розподіл відстаней донор–акцептор має виражений максимум близько 0,27 нм (панель В), що лежить у характерному для стабільних водневих зв'язків діапазоні й суттєво менше за геометричний поріг відсікання 0,35 нм. Розподіл відхилень від лінійності має моду в діапазоні  $12\text{--}15^\circ$  (панель Г), що відповідає геометрії, характерній для повноцінних водневих зв'язків, і вказує на переважно сприятливу орієнтацію донорно-акцепторних пар. Висока частота водневих зв'язків протягом усієї траєкторії свідчить про збереження стійких полярних контактів між GFB-887 і TRPC4, незважаючи на локальні позиційні флуктуації ліганду.

Порівняння отриманих під час МД-симуляції взаємодій із докінговою моделлю демонструє, що полярні контакти, ідентифіковані на етапі докінгу, зберігаються й у мембранному оточенні. Залишки His369, Asn442, Arg491 та Ser494, яким у докінговій моделі приписувалася роль полярних «якорів» зв'язування GFB-887, продовжують формувати водневі зв'язки з полярними фрагментами ліганду протягом усієї траєкторії. Утримання цих контактів разом із збереженням гідрофобно-ароматичного оточення, забезпеченого залишками Phe413, Tyr373, Tyr646 та аліфатичними Leu492/Leu495,

підтверджує фізичну обґрунтованість докінгової пози та свідчить про те, що отриманий механізм зв'язування не є артефактом статичної оптимізації.

Узагальнюючи результати МД-моделювання, можна констатувати, що комплекс TRPC4:GFB-887, побудований на основі докінгового розміщення, зберігає структурну стабільність протягом 200 нс симуляції у POPC-бішарі. Про це свідчать стаціонарність термодинамічних параметрів системи, обмежений рівень RMSD чотирьох копій ліганду, стабільний профіль водневих зв'язків та збереження ключових полярних і гідрофобних контактів, ідентифікованих на етапі докінгу. Отримані дані підтверджують, що запропонований сайт зв'язування у VSL-доміні є енергетично й геометрично сумісним зі сполукою GFB-887 і може розглядатися як структурна основа для подальшого аналізу молекулярних детермінант інгібування TRPC4.

## РОЗДІЛ 5

### ОЦІНКА ADMET-ПАРАМЕТРІВ

Попередні розділи присвячені структурним аспектам взаємодії ліганду з TRPC4 – ідентифікації сайтів зв'язування та оцінці стабільності комплексу методами докінгу й молекулярної динаміки. Проте ефективність сполуки як лікарського кандидата визначається не лише спорідненістю до мішені: значна частка активних молекул відсіюється на доклінічних і клінічних етапах через неоптимальні ADMET-параметри. Експериментальна оцінка цих властивостей є ресурсо- та часовитратною і непридатною для скринінгу великих бібліотек, що зумовило розвиток *in silico* методів ADMET-прогнозування на основі машинного навчання – від класичних дескрипторів і фінгерпринтів до навчених векторних ембедингів та графових нейронних мереж.

У цьому розділі описано розроблений метод прогнозування ADMET-параметрів на основі descriptor-augmented Mol2Vec-ембедингів, який поєднує контекстні векторні представлення молекулярних субструктур із класичними фізико-хімічними дескрипторами [17]. Наведено архітектуру методу, протокол навчання та результати порівняльного бенчмаркінгу на наборах TDC.

Серед існуючих підходів до молекулярних представлень окремий клас утворюють методи некерованого навчання, зокрема Mol2Vec, який генерує представлення молекулярних підструктур за аналогією з Word2Vec в обробці природної мови [99]. Такий підхід відображає хімічний контекст і водночас уникає характерних для традиційних фінгерпринтів проблем високої розмірності та колізій бітів. Велика кількість ознак, притаманна комбінованим представленням, ускладнює навчання моделей на невеликих датасетах через ризик перенавчання; для пом'якшення цієї проблеми зазвичай застосовують методи відбору ознак – відбір за порогом дисперсії, кореляційну фільтрацію та відбір на основі оцінки важливості Random Forest.

Однак систематична оцінка Mol2Vec у широкому спектрі ADMET-задач на стандартизованих бенчмарках TDC досі не проводилася [100].

### 5.1. Загальна схема методу

Розроблений метод є конвеєром послідовно з'єднаних модулів, загальна архітектура якого наведена на рисунку 5.1.

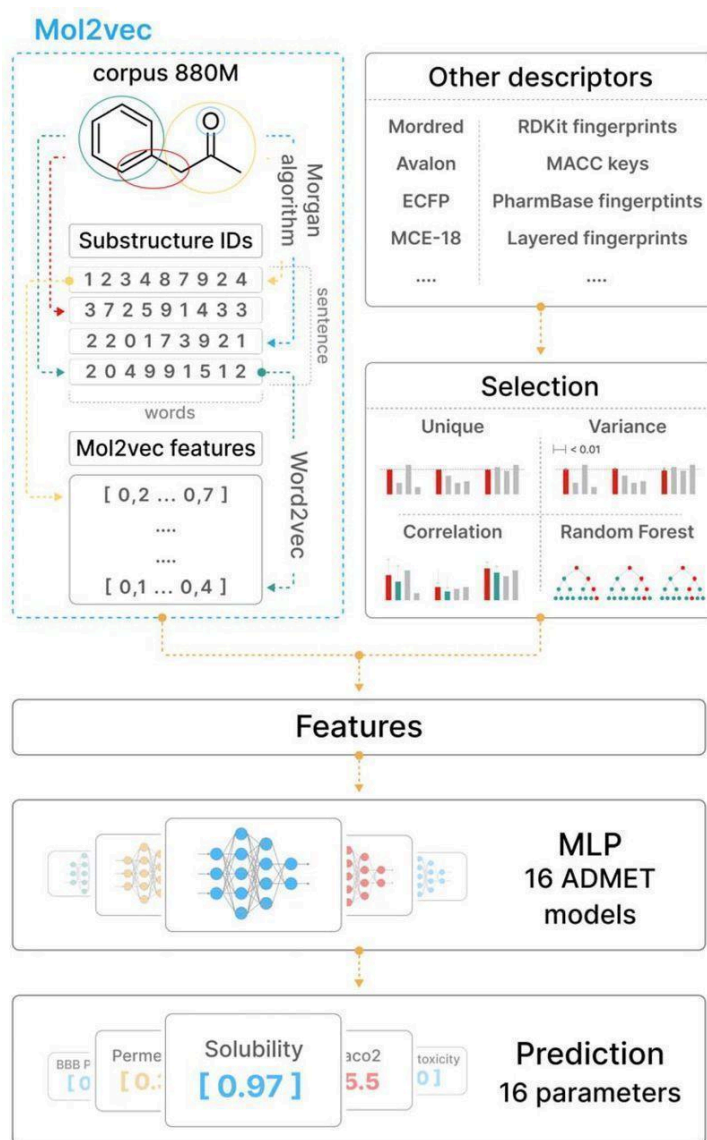


Рис. 5.1. Загальна схема методу прогнозування ADMET-параметрів на основі descriptor-augmented Mol2Vec-ембедингів [17]. Конвеєр починається з виокремлення молекулярних підструктур алгоритмом Morgan, застосованим до корпусу ZINC20 обсягом 880 млн сполук; так генеруються послідовності

ідентифікаторів підструктур (аналоги «речень»), на основі яких навчаються Mol2Vec-ембединги моделлю Word2Vec Skip-gram. Отримані ембединги поєднуються з традиційними молекулярними дескрипторами у спільний простір ознак. Модуль відбору ознак (фільтрація за унікальністю, відсікання за дисперсією, видалення корельованих ознак та оцінка важливості за Random Forest) скорочує цей високорозмірний простір. Відібрані ознаки подаються на вхід 16 задач-специфічних MLP-моделей для прогнозування ключових ADMET-властивостей на 16 бенчмарках TDC.

На першому етапі з корпусу ZINC20 виокремлюються молекулярні підструктури алгоритмом Morgan; ці підструктури інтерпретуються як «слова», а цілі молекули, як «речення», сформовані послідовностями «слів». На цьому корпусі навчається модель Mol2Vec за архітектурою Word2Vec Skip-gram, яка породжує 512-вимірні вектори-ембединги підструктур. Для кожної молекули ембединги її підструктур агрегуються у єдиний 512-вимірний вектор молекулярного представлення.

На другому етапі Mol2Vec-ембедінг конкатенується з розширеним набором класичних молекулярних дескрипторів і фінгерпринтів: Mordred, MCE-18, Morgan, RDKit, Layered, Atom, Topological, Avalon, PharmBase та MACCS keys. До отриманого високорозмірного простору ознак застосовується чотириступеневий конвеєр відбору ознак: послідовно видаляються ознаки з єдиним унікальним значенням, ознаки з низькою дисперсією, високо корельовані пари ознак, а також ознаки з низькою важливістю за оцінкою Random Forest. Відібраний набір ознак конкатенується з фіксованим 512-вимірним Mol2Vec-вектором і подається на вхід 16 окремих багатосарових перцептронів (MLP) – по одній моделі на кожну з 16 обраних задач TDC, гіперпараметри яких оптимізуються незалежно. Таким чином, архітектура методу дає змогу поєднати переваги неконтрольованого підструктурного ембедингу, експертного опису молекул класичними дескрипторами та задач-специфічної оптимізації MLP-моделей.

## 5.2. Підготовка корпусу та генерація Mol2Vec-ембедингів

Для формування репрезентативного навчального корпусу використано базу ZINC20 із застосуванням послідовності фільтрів, які забезпечують хімічну релевантність і якість даних [122]. Зокрема, відбиралися лише сполуки з характеристикою «2D representation», позначені як «Standard reactivity» та зі статусом комерційної доступності «Wait OK». Внаслідок такої фільтрації сформовано набір із 883 897 271 малих молекул.

На етапі попередньої обробки з молекулярних структур видалялися фрагменти розчинників, після чого для кожної сполуки генерувалися канонічні SMILES засобами RDKit, що забезпечує однозначність та унікальність подання. Молекулярні підструктури виокремлювалися алгоритмом Morgan з двома значеннями радіуса: радіусом 0 (ідентичність атома) та радіусом 1 (атом із безпосереднім оточенням) [17, 96, 123]. Отримані підструктури трактувалися як «слова», а окрема молекула – як «речення», сформоване послідовністю таких «слів».

Підсумковий корпус охоплює словник із 13278 унікальних молекулярних «слів». Щоб пом'якшити ризик перенавчання і покращити узагальнювальну здатність ембедингу, усі підструктурні токени з частотою появи менше 4 у корпусі замінювалися спеціальним токеном UNK – такий прийом є стандартною практикою обробки рідкісних токенів у машинному навчанні.

Для генерації молекулярних ембедингів застосовувалися дві усталені архітектури, адаптовані з обробки природної мови, – Continuous Bag of Words (CBOW) та Skip-gram. Архітектура CBOW спрямована на прогнозування цільового токена за його контекстним оточенням, тоді як Skip-gram виконує обернену задачу, а саме прогнозує контекстні токени за цільовим. Попереднє порівняння обох варіантів показало, що найбільш ефективні молекулярні представлення формує модель Skip-gram із розміром контекстного вікна 10 та мінімальним порогом частоти підструктури (unseen cut-off), що дорівнює 4;

саме цю конфігурацію зафіксовано для всіх подальших експериментів та навчання фінальних задач-специфічних моделей.

Кодування підструктурних токенів у 512-вимірні вектори виконувалося засобами бібліотеки Gensim. Окрема молекула представлялася одним вектором, який отримувався агрегуванням ембедингів усіх її підструктур. Підвищення розмірності ембедингу з 300, використаних в оригінальному Mol2Vec [99], до 512 у поєднанні з істотно розширеним корпусом ZINC20 спрямоване на підвищення здатності представлення відображати хімічну різноманітність реального простору лікоподібних сполук.

### **5.3. Стратегії конструювання ознак та їх відбір**

Для формування вхідних ознак ADMET-моделей оцінювалися чотири стратегії конструювання ознак. Перша, Mol2Vec Only, спирається лише на 512-вимірний Mol2Vec-ембедінг і слугує базовою точкою відліку для оцінки самостійної корисності запропонованого неконтрольованого представлення. Друга стратегія доповнює Mol2Vec фізико-хімічними властивостями – молекулярною масою, logP, числом донорів і акцепторів водневого зв'язку, числом обертальних зв'язків, загальною полярною площею поверхні, формальним зарядом та іншими дескрипторами. Третя стратегія розширює Mol2Vec 2D-дескрипторами Mordred, що охоплюють широкий набір структурних та фізико-хімічних характеристик. Четверта стратегія, Mol2Vec Best, комбінує Mol2Vec із розгалуженим набором додаткових молекулярних дескрипторів і фінгерпринтів, після чого застосовується автоматизований відбір ознак. Саме ця композитна конфігурація забезпечує найкращу якість прогнозування і виступає основною у цьому розділі. Повний перелік додаткових молекулярних дескрипторів та фінгерпринтів, задіяних у конфігурації Mol2Vec Best, подано в таблиці 5.1.



Таблиця 5.1.

Додаткові молекулярні дескриптори, використані у конфігурації Mol2Vec Best [17].

| Дескриптори              | Кількість | Параметри                                                                    |
|--------------------------|-----------|------------------------------------------------------------------------------|
| Mordred descriptors      | 1613      | лише 2D                                                                      |
| MCE-18 descriptor        | 1         | –                                                                            |
| Morgan fingerprints      | 28672     | bits: [512, 1024, 2048];<br>radius: [1, 2, 3, 4];<br>branches: [True, False] |
| RDKit fingerprints       | 25088     | bits: [512, 1024, 2048];<br>max path: [1, 2, 3, 4, 5, 6, 7]                  |
| Layered fingerprints     | 25088     | bits: [512, 1024, 2048];<br>max path: [1, 2, 3, 4, 5, 6, 7]                  |
| Atom fingerprints        | 3584      | bits: [512, 1024, 2048]                                                      |
| Topological fingerprints | 3584      | bits: [512, 1024, 2048]                                                      |
| Avalon fingerprints      | 3584      | bits: [512, 1024, 2048]                                                      |
| PharmBase fingerprints   | 3584      | –                                                                            |
| MACCS keys               | 167       | –                                                                            |

Початковий простір ознак у конфігурації Mol2Vec Best охоплює 94 729 кандидатних ознак на кожну ADMET-задачу. Для систематичного скорочення розмірності застосовується чотириступеневий конвеєр відбору ознак, спрямований на усунення неінформативних, малозмінних та сильно корельованих змінних. На першому кроці задіюється селектор унікальності: ознаки, що набувають у навчальній вибірці єдиного унікального значення, виключаються. На другому кроці застосовується селектор за дисперсією: з розгляду видаляються ознаки, у яких дисперсія у навчальній вибірці є нижчою за 0,01, як такі, що демонструють мінімальну варіативність. На третьому кроці діє кореляційний селектор: з пар ознак, які мають коефіцієнт

кореляції Пірсона понад 0,95, одна ознака виключається. На четвертому кроці застосовується селектор важливості ознак на основі Random Forest: регресором Random Forest обчислюється важливість кожної ознаки, після чого ознаки з найнижчою важливістю виключаються, а найбільш інформативні зберігаються. Ці кроки застосовуються послідовно до кожного бенчмарку окремо.

Підсумкова підмножина відібраних ознак конкатенується з фіксованим 512-вимірним набором Mol2Vec-ознак і формує остаточний вхідний вектор для навчання задач-специфічної моделі. Розмірність підсумкового простору ознак залежить від задачі і змінюється у широких межах: від понад 600 ознак для компактних датасетів до понад 8000 для найбільших (зокрема, інгібування CYP2D6). У таблиці 5.2 наведено підсумкову розмірність для всіх 16 датасетів: від найменшого (HIA) до найбільшого (CYP2D6).

Таблиця 5.2.

Загальна кількість ознак, використаних на вході моделей Mol2Vec Best після відбору, для всіх 16 ADMET-датасетів [17].

| Датасет           | Відібрані<br>ознаки | Mol2Vec-ознак<br>и | Усього |
|-------------------|---------------------|--------------------|--------|
| Caco2             | 1021                | 512                | 1533   |
| HIA               | 151                 | 512                | 663    |
| Pgp Inhibition    | 688                 | 512                | 1200   |
| Bioavailability   | 1171                | 512                | 1683   |
| BBB               | 1329                | 512                | 1841   |
| PPBR              | 1997                | 512                | 2509   |
| VDss              | 1613                | 512                | 2125   |
| CYP2C9 Inhibition | 6795                | 512                | 7307   |

| Датасет           | Відібрані<br>ознаки | Mol2Vec-ознак<br>и | Усього |
|-------------------|---------------------|--------------------|--------|
| CYP2D6 Inhibition | 7588                | 512                | 8100   |
| CYP3A4 Inhibition | 7065                | 512                | 7577   |
| CYP2C9 Substrate  | 1124                | 512                | 1636   |
| CYP2D6 Substrate  | 7588                | 512                | 8100   |
| CYP3A4 Substrate  | 7065                | 512                | 7577   |
| hERG              | 794                 | 512                | 1306   |
| AMES              | 3130                | 512                | 3642   |
| DILI              | 639                 | 512                | 1151   |

#### 5.4. Навчання MLP-моделей та оптимізація гіперпараметрів

Для кожної з 16 ADMET-задач окремо навчалася індивідуальна MLP-модель з власною задач-специфічною конфігурацією гіперпараметрів. Архітектура моделі параметризувалася фіксованим набором кандидатних значень для всіх налаштовуваних параметрів, які оптимізувалися незалежно для кожного датасету.

Оптимізація гіперпараметрів виконувалася у фреймворку Optuna версії 3.3.0 з використанням алгоритму Tree-structured Parzen Estimator (TPE) зі стандартними параметрами [124]. Процес оптимізації запускався окремо для кожного з 16 бенчмарків TDC, що давало окремий набір оптимальних гіперпараметрів на кожну задачу. До оптимізації залучалися такі параметри: кількість прихованих шарів у діапазоні від 1 до 5; кількість нейронів у кожному шарі; коефіцієнт dropout у діапазоні від 0,0 до 0,5; функція активації

(на вибір – ELU, ReLU, PReLU або Leaky ReLU); розмір батчу; значення weight decay; швидкість навчання (learning rate).

Для навчання на всіх бенчмарках застосовувався оптимізатор Adam; швидкість навчання підбиралася Optuna разом з іншими гіперпараметрами. Кожна модель тренувалася упродовж 100 епох, а кількість ітерацій (trials) процесу підбору для кожної задачі становила 200. Для розбиття на навчальну та тестову підвибірki використовувалися scaffold-based splits, надані TDC. Відповідно до стандартів оцінювання TDC, показники якості для кожного параметру подаються у форматі «середнє  $\pm$  стандартне відхилення».

### **5.5. Порівняння з існуючими підходами на бенчмарках TDC**

Для оцінювання методу, з платформи TDC відібрано 16 бенчмарків, найбільш релевантних до задач реального drug discovery. Ці бенчмарки згруповано у чотири категорії ADMET. Перша категорія – абсорбція, що охоплює чотири задачі: проникність через Caco-2, HIA, інгібування Р-глікопротеїну та пероральну біодоступність [125–128]. Друга категорія, розподіл, охоплює три задачі, що описують поширення сполуки в організмі: проникнення через BBB, PPB та VDss [129–131]. Третя категорія – метаболізм, що охоплює шість задач, пов'язаних із метаболічним перетворенням сполук та його впливом на експозицію й ефективність препарату: інгібування CYP2D6, CYP3A4 і CYP2C9, а також класифікацію субстратів CYP2D6, CYP3A4 і CYP2C9 [132].

Четверта категорія, токсичність, включає три задачі, що оцінюють потенційні побічні ефекти: Ames mutagenicity, DILI та інгібування hERG [133–135]. Серед відібраних задач, 3 сформульовані як регресійні, а решта 13, як задачі бінарної класифікації; обсяги датасетів варіюються від малих (HIA – 578 сполук) до великих (інгібування CYP2D6 – 13 130 сполук).

Загальне порівняння моделей на лідерборді TDC станом на липень 2025 року подано в таблиці 5.3, що містить ранги за кожною з 16 задач та підрахунок кількості перших місць серед інших учасників лідерборду.

Таблиця 5.3.

Ранги моделей на лідерборді TDC (станом на липень 2025) та моделей на основі Mol2Vec за 16 ADMET-задачами [17]. Нижче значення рангу відповідає кращій моделі. Найкращі результати на кожній задачі виділено жирним

| ADMET параметр  | MapLig ht | MapLig ht + GNN | ZairaChem | DeepPurPOSE (Morgan FP) | Mol2Vec Only | Mol2Vec Best |
|-----------------|-----------|-----------------|-----------|-------------------------|--------------|--------------|
| Caco-2          | <b>1</b>  | 4               | –         | 17                      | 10           | 7            |
| HIA             | 3         | 2               | 4         | 13                      | 10           | <b>1</b>     |
| Pgp-inh         | 3         | <b>1</b>        | 10        | 16                      | 10           | <b>1</b>     |
| Bioavailability | 3         | <b>1</b>        | 2         | 13                      | 8            | <b>1</b>     |
| BBB             | <b>1</b>  | 3               | 5         | –                       | 2            | <b>1</b>     |
| PPBR            | 2         | <b>1</b>        | –         | 15                      | 14           | <b>1</b>     |
| VDss            | 2         | <b>1</b>        | –         | 8                       | 13           | 4            |
| CYP2D6-inh      | 5         | <b>1</b>        | 4         | 13                      | 9            | 3            |
| CYP3A4-inh      | 3         | <b>1</b>        | 8         | 12                      | 11           | 4            |
| CYP2C9-inh      | 4         | <b>1</b>        | 6         | 12                      | 13           | 4            |
| CYP2D6-sub      | 5         | 2               | <b>1</b>  | 10                      | 7            | <b>1</b>     |

|                                       |   |          |          |    |    |           |
|---------------------------------------|---|----------|----------|----|----|-----------|
| CYP3A4-s<br>ub                        | 3 | 2        | 6        | 8  | 12 | <b>1</b>  |
| CYP2C9-s<br>ub                        | 2 | 3        | 7        | 6  | 16 | <b>1</b>  |
| hERG                                  | 3 | <b>1</b> | 4        | 13 | 11 | <b>1</b>  |
| AMES                                  | 3 | 2        | <b>1</b> | 12 | 12 | <b>1</b>  |
| DILI                                  | 6 | 3        | <b>1</b> | 14 | 11 | 4         |
| <b>Кількість<br/>перших<br/>місць</b> | 2 | 8        | 3        | 0  | 0  | <b>10</b> |

Для порівняння відібрано репрезентативні моделі з принципово різними стратегіями кодування молекул. MapLight – провідна модель для прогнозування різноманітних ADMET-параметрів, яка поєднує класичні фінгерпринти та графові представлення. Її розширений варіант MapLight + GNN додає ембединги графових нейронних мереж та стабільно досягає топ-1 позицій на низці параметрів [100]. ZairaChem побудована як ансамбль різних методів молекулярного кодування, включно з представленнями на основі мовних моделей, та демонструє сильні результати, однак обмежена лише задачами класифікації [97]. DeepPurpose ґрунтується на архітектурі «енкодер-декодер», де Morgan-фінгерпринти подаються на вхід MLP-енкодера, за яким йде декодер для прогнозування біологічної активності; DeepPurpose включено до порівняння саме через архітектурну близькість до базової моделі Mol2Vec Only [136].

У категорії абсорбції метод Mol2Vec Best досягнув першого місця у трьох із чотирьох задач: НІА, інгібування P-gr та пероральна біодоступність. Для цих параметрів додавання молекулярних дескрипторів з відбором ознак до Mol2Vec Only дало суттєвий приріст якості, достатній для виходу на провідну позицію в рейтингу. Винятком є задача Caco-2: тут приріст від доповнення дескрипторами мінімальний, і Mol2Vec Best сягає лише 7-го місця. Такий результат свідчить, що додаткові дескриптори не дають істотної

прогностичної цінності для цього параметру, ймовірно через обмежену релевантність доданих ознак або притаманний датасету шум. Характерно, що найкращий результат у цій задачі показує MapLight – модель, яка поєднує класичні фінгерпринти з графовими представленнями; сильні позиції має й SimGCN [137]. Водночас додавання окремого GNN-модуля (MapLight + GNN) не покращило результат саме на Caco-2 (ранг 4 проти 1 у базового MapLight). Це свідчить, що вирішальною є наявність структурно-графової інформації про молекулу, а не конкретна GNN-архітектура. Підструктурні ембединги Mol2Vec Best, позбавлені явного графового компонента, досягають тут лише 7-го місця..

У категорії розподілу метод досягнув провідних позицій у двох із трьох задач. Примітно, що вже на базовій конфігурації Mol2Vec Only модель виходить на 2-ге місце у задачі BBB без додаткових дескрипторів. Це свідчить, що Mol2Vec-ембединги самі по собі добре відображають ключові структурні та фізико-хімічні ознаки, релевантні для проникності гематоенцефалічного бар'єру. Для решти задач, а саме PPB та VDss, конфігурація Mol2Vec Best забезпечує подальший приріст та виводить метод на 1-ше і 4-те місця відповідно.

У категорії метаболізму, моделі Mol2Vec Best демонструють провідну ефективність на більшості задач класифікації субстратів цитохромів P450: для субстратів CYP2D6, CYP3A4 та CYP2C9 досягнуто перших місць. На задачах інгібування цитохромів Mol2Vec Only поступається сучасним моделям, однак у конфігурації Mol2Vec Best додавання дескрипторів істотно покращує якість і виводить результати в кращі позиції (3–4-те місце для CYP2D6-inh, CYP3A4-inh та CYP2C9-inh). Серед конкурентів у цій категорії високу ефективність демонструють MapLight GNN, особливо на задачах інгібування CYP, та ZairaChem, зокрема на класифікації субстратів CYP2C9. Ці результати підкреслюють важливість збагачених молекулярних представлень для відображення складних біохімічних детермінант метаболізму препаратів.

У категорії токсичності метод Mol2Vec Best досяг найкращого рангу серед порівнюваних моделей в двох з трьох задач – інгібування hERG, Ames mutagenicity та DILI. Натомість Mol2Vec Only демонструє тут субоптимальну ефективність, що свідчить про недостатність самих лише підструктурних ембедингів для відображення складних біологічних сигналів, пов'язаних із токсичністю. Значний приріст якості підкреслює важливість інтеграції різноманітних молекулярних дескрипторів із цілеспрямованим відбором ознак, зокрема, на фоні сильних базових моделей, таких як MapLight GNN (особливо ефективний для hERG) та ZairaChem (добре виступає на Ames і DILI), що додатково підтверджує ефективність запропонованої композитної стратегії формування ознак.

Підсумовуючи дані таблиці 5.3, Mol2Vec Best посідає перше місце у 10 з 16 бенчмарків – на момент проведення цього етапу дослідження, це найвищий показник серед усіх моделей лідерборда TDC. Наприклад, для MapLight + GNN це значення становить 8, для ZairaChem – 3, для MapLight – 2, а для DeepPurpose та Mol2Vec Only – 0. Водночас базова модель Mol2Vec Only показує результати, конкурентні з обраними для порівняння моделями, та перевершує свого найближчого архітектурного аналога DeepPurpose з Morgan-фінгерпринтами та MLP-енкодером.

Отримані результати демонструють, що збагачені дескрипторами ембединги, оптимізовані завдяки принциповому відбору ознак, здатні конкурувати з більш складними архітектурами такими, як GNN-моделі та підходи на основі мовних моделей, але при істотно менших вимогах до обчислювальних ресурсів і архітектурної складності. Розроблений метод застосований для оцінки ADMET-профілів досліджуваних сполук.



## 5.6. Оцінка ADMET-параметрів для Pico145 та GFB-887

Для доповнення результатів молекулярного моделювання, наведених у попередньому розділі, проведено оцінку прогнозованих ADMET-параметрів для сполук Pico145 та GFB-887. Отримані прогнози дозволяють інтерпретувати ймовірні фармакокінетичні та безпекові характеристики досліджуваних молекул з позиції експериментальних ADME-показників, які є стандартними у галузі розробки лікарських засобів, та окреслити пріоритетні напрями подальшої експериментальної перевірки. Прогнози отримано за допомогою моделі, описаної в попередніх підрозділах цього розділу та опублікованої у вигляді препринта, натренованої на наборах експериментальних даних платформи TDC [17, 100]. Використано два типи моделей – класифікаційні, які повертають ймовірність належності сполуки до певного класу, сформованого за емпіричним порогом на етапі навчання, та регресійні, які прогнозують числове значення параметра у фізичних одиницях.

Оцінку здійснено за такими параметрами: Caco-2, HIA, інгібування P-глікопротеїну, BBB, PPB VD<sub>ss</sub>, зв'язування у якості субстрату та інгібування ізоформ цитохрому P450 (CYP2C9, CYP2D6, CYP3A4), тест Ames mutagenicity, DILI та блокування hERG.

### 5.6.1. ADMET-профіль Pico145

Прогнозовані ADMET-параметри Pico145 наведено у таблиці 5.4. Отримані значення далі інтерпретуються за групами параметрів: абсорбції, розподілу, метаболізму та токсичності у співставленні з доступними експериментальними даними для цієї сполуки. Слід враховувати, що у відкритих джерелах фармакокінетичних даних безпосередньо для Pico145

небагато, тоді як для його структурного аналога НС-070 наведено повніший профіль [138].

Таблиця 5.4.

Прогнозовані ADMET-параметри сполуки Pico145.

| Параметр                                     | Прогнозоване значення | Інтерпретація                                     |
|----------------------------------------------|-----------------------|---------------------------------------------------|
| <b>Абсорбція</b>                             |                       |                                                   |
| С <sub>асо-2</sub> , P <sub>app</sub> (см/с) | $3,47 \times 10^{-6}$ | помірна проникність, обмежена пасивна дифузія     |
| НІА (ймов.)                                  | 0,9799                | ефективна кишкова абсорбція                       |
| Інгібування Р-глікопротеїна (ймов.)          | 0,8498                | висока ймовірність інгібування                    |
| <b>Розподіл</b>                              |                       |                                                   |
| ВВВ (ймов.)                                  | 0,4648                | помірно-низька проникність через ВВВ              |
| РРВ (%)                                      | 97,6                  | дуже високий ступінь зв'язування з білками плазми |
| VD <sub>ss</sub> (л/кг)                      | 3,44                  | широкий об'єм розподілу                           |
| <b>Метаболізм – субстратність CYP</b>        |                       |                                                   |
| CYP3A4 (субстрат, ймов.)                     | 0,9999                | ймовірний субстрат                                |
| CYP2C9 (субстрат, ймов.)                     | 0,0715                | малоймовірний субстрат                            |

| Параметр                            | Прогнозоване значення | Інтерпретація                        |
|-------------------------------------|-----------------------|--------------------------------------|
| CYP2D6 (субстрат, ймов.)            | 0,0453                | малоймовірний субстрат               |
| <b>Метаболізм – інгібування CYP</b> |                       |                                      |
| CYP3A4 (інгібування, ймов.)         | 0,9379                | ймовірний інгібітор                  |
| CYP2C9 (інгібування, ймов.)         | 0,9905                | ймовірний інгібітор                  |
| CYP2D6 (інгібування, ймов.)         | 0,9988                | ймовірний інгібітор                  |
| <b>Токсичність</b>                  |                       |                                      |
| Ames mutagenicity (ймов.)           | 0,2369                | низька ймовірність генотоксичності   |
| DILI (ймов.)                        | 0,9989                | висока ймовірність гепатотоксичності |
| Блокування hERG (ймов.)             | 0,3237                | низька ймовірність блокади каналу    |

Серед прогнозованих параметрів першочергову увагу привертають показники абсорбції. Коефіцієнт  $P_{app}$  через моношар Caco-2 становив  $3,47 \times 10^{-6}$  см/с (помірна проникність), тоді як НІА дорівнював 0,9799, відносячи Pico145 до сполук з ефективною кишковою абсорбцією. Цей прогноз отримує пряме експериментальне підтвердження у роботі Dryn *et al.*: пероральне введення 1 мг/кг мишам викликало пригнічення моторики тонкої кишки після прийому їжі приблизно на 24 %, що свідчить про достатню системну експозицію [20]. Сполука також систематично описується як перорально біодоступний інгібітор, тоді як кількісні параметри абсорбції у відкритих джерелах не наведено [20, 139].

Показники розподілу мали такі прогнозні значення: BBB – 0,4648, PPB – 97,6 %, VD<sub>ss</sub> – 3,44 л/кг. Прогнозована ймовірність проникнення через BBB

(0,4648) є пограничною – близькою до порогу класифікації. Для структурно близької сполуки НС-070 співвідношення концентрацій у мозку та плазмі після введення мишам становило  $\approx 0,9$ , що радше вказує на відчутне проникнення сполук цього класу в ЦНС [138]; з огляду на це прогноз для Pico145 доцільно розглядати як орієнтовний, що потребує експериментального підтвердження. Прогнозоване значення PPV добре узгоджується з експериментально визначеним для НС-070 ступенем зв'язування з білками плазми 99,6 %, а прогнозний VDss якісно узгоджується з експериментальним Vss для НС-070, що перевищував об'єм води в організмі ( $\approx 0,7$  л/кг) і характеризував сполуку як таку, що активно розподіляється у тканини [138].

Класифікаційна модель субстратності CYP передбачила з дуже високою ймовірністю належність Pico145 до субстратів CYP3A4 (0,9999) і з низькою до CYP2C9 (0,0715) та CYP2D6 (0,0453), що є очікуваним для ліпофільної молекули з об'ємними арильними замісниками, на відміну від простих метилксантинів, які метаболізуються переважно через CYP1A2. Натомість моделі, що прогнозують інгібування, дали високі значення для всіх трьох ізоформ: CYP3A4 – 0,9379, CYP2C9 – 0,9905, CYP2D6 – 0,9988. Експериментальних *in vitro* даних щодо CYP-профілю Pico145 у відкритих джерелах не опубліковано.

Класифікаційні прогнози токсикологічних параметрів становили: Ames mutagenicity – 0,2369, DILI – 0,9989, hERG – 0,3237. Низьке значення тесту Еймса свідчить про низьку ймовірність генотоксичної активності сполуки.. Прогноз hERG прямо узгоджується з експериментальними даними Just та співавторів: НС-608 (Pico145) у складі панелі селективності демонстрував профіль, аналогічний НС-070, який інгібував струм через hERG менш ніж на 50 % за 1,9 мкмоль/л, забезпечуючи щонайменше 400-кратну селективність на користь TRPC4/5-каналів [138]. Натомість високе значення DILI залишається непідтвердженим: прямих експериментальних даних щодо

гепатотоксичності у публічному доступі немає, а наявні *in vivo* дослідження не передбачали оцінки печінкових біомаркерів [20, 138].

Підсумовуючи, прогнозований ADMET-профіль Pico145 окреслює сполуку з загалом сприятливими фармакокінетичними характеристиками, але варто зазначити, що є і певні застереження. Часткове підтвердження прогнозу експериментальними даними щодо структурно близького НС-070 свідчить про придатність застосованого моделювання для сполук цього хімічного класу, водночас більшість параметрів залишаються без прямого експериментального референсу для самого Pico145 [138]. Це визначає необхідність подальших цілеспрямованих експериментальних досліджень для поглибленого розуміння фармакокінетичного профілю сполуки.

### 5.6.2. ADMET-профіль GFB-887

Прогнозовані ADMET-параметри сполуки GFB-887 наведено у таблиці 5.5. Отримані значення далі інтерпретуються за групами параметрів: абсорбції, розподілу, метаболізму та токсичності у співставленні з доступними експериментальними та клінічними даними для цієї сполуки.

Таблиця 5.5.

Прогнозовані ADMET-параметри сполуки GFB-887.

| Параметр                 | Прогнозоване значення | Інтерпретація                                                          |
|--------------------------|-----------------------|------------------------------------------------------------------------|
| <b>Абсорбція</b>         |                       |                                                                        |
| Caco-2, $P_{app}$ (см/с) | $1,46 \times 10^{-5}$ | висока проникність, сприятлива пасивна дифузія через кишковий епітелій |
| НІА (ймов.)              | 0,9735                | ефективна кишкова абсорбція                                            |

| Параметр                              | Прогнозоване значення | Інтерпретація                                  |
|---------------------------------------|-----------------------|------------------------------------------------|
| Інгібування Р-глікопротеїну (ймов.)   | 0,4474                | помірна ймовірність інгібування                |
| <b>Розподіл</b>                       |                       |                                                |
| BBB (ймов.)                           | 0,6004                | помірна проникність через BBB                  |
| PPB (%)                               | 95,8                  | високий ступінь зв'язування з білками плазми   |
| VD <sub>ss</sub> (л/кг)               | 0,95                  | помірний об'єм розподілу у стаціонарному стані |
| <b>Метаболізм – субстратність CYP</b> |                       |                                                |
| CYP3A4 (субстрат, ймов.)              | 0,9982                | імовірний субстрат                             |
| CYP2C9 (субстрат, ймов.)              | 0,8018                | імовірний субстрат                             |
| CYP2D6 (субстрат, ймов.)              | 0,1809                | малоймовірний субстрат                         |
| <b>Метаболізм –інгібування CYP</b>    |                       |                                                |
| CYP3A4 (інгібування, ймов.)           | 0,1784                | малоймовірний інгібітор                        |
| CYP2C9 (інгібування, ймов.)           | 0,1178                | малоймовірний інгібітор                        |
| CYP2D6 (інгібування, ймов.)           | 0,0198                | малоймовірний інгібітор                        |
| <b>Токсичність</b>                    |                       |                                                |

| Параметр                  | Прогнозоване значення | Інтерпретація                        |
|---------------------------|-----------------------|--------------------------------------|
| Ames mutagenicity (ймов.) | 0,2113                | низька ймовірність генотоксичності   |
| DILI (ймов.)              | 0,8861                | висока ймовірність гепатотоксичності |
| Блокування hERG (ймов.)   | 0,5244                | ймовірна блокада каналу              |

Серед прогнозованих параметрів першочергову увагу привертають показники, що характеризують абсорбцію сполуки. Коефіцієнт проникності ( $P_{app}$ ) через моношар клітин Caco-2 становив  $1,46 \times 10^{-5}$  см/с, що вказує на високу проникність і свідчить про сприятливий потенціал пасивної дифузії через кишковий епітелій. Узгоджено з цим, класифікаційний показник НІА дорівнював 0,9735, тобто модель з високою ймовірністю відносить GFB-887 до класу сполук з ефективною кишковою абсорбцією. У доступних публікаціях щодо GFB-887 не наведено числових результатів *in vitro* тестів проникності Caco-2 або інших кількісних оцінок кишкової абсорбції, придатних для прямого порівняння. Водночас доклінічні дослідження фармакокінетики показали, що пероральне введення GFB-887 у щурів із трансплантованими людськими органами забезпечує вимірювану концентрацію препарату як у плазмі, так і в тканинах органів, що опосередковано підтверджує здатність сполуки надходити у системний кровотік після перорального застосування [140]. Крім того, у публікації про перше клінічне дослідження за участю людини зазначено, що сполука характеризувалася фармакокінетичним профілем, сумісним із режимом одноразового добового дозування та сприятливим профілем безпеки [12, 140]. Хоча наведені клінічні дані не є прямим еквівалентом експериментальних ADME-параметрів, вони опосередковано підтверджують відсутність обмежень з боку абсорбції, що узгоджується з високими прогнозними значеннями Caco-2 та НІА. Додатково, для параметра P-gp inhibition отримано

значення 0,4474, що відповідає помірній ймовірності інгібування P-gr і не дозволяє зробити однозначний висновок без експериментальної перевірки. У літературі для GFB-887 також відсутні дані щодо взаємодії з транспортерами, тому цей параметр набуває особливої ваги як напрям для подальшої валідації.

Показники, що характеризують розподіл сполуки в організмі, мали такі прогнозні значення: BBV – 0,6004, PPV – 95,8 %, VDss – 0,95 л/кг. Публічні клінічні матеріали не містять детальних даних щодо співвідношення концентрацій у мозку та плазмі, фракції незв’язаної речовини або параметрів розподілу в термінах VDss, тому отримані значення доцільно розглядати як прогнозні оцінки. З точки зору фармакокінетики, поєднання високого зв’язування з білками плазми та помірного об’єму розподілу є типовим для сполук, які тривалий час утримуються у системній циркуляції та характеризуються відносно повільним зниженням концентрації у плазмі. Така конфігурація параметрів не суперечить клінічно заявленому режиму одноразового добового дозування для GFB-887, хоча сама ця узгодженість не може розглядатися як підтвердження коректності числових прогнозів [12].

Класифікаційна модель субстратності ізоформ CYP показала високу ймовірність належності GFB-887 до субстратів CYP3A4 (0,9982) і CYP2C9 (0,8018) та низьку для CYP2D6 (0,1809). Натомість модель інгібування CYP дала низькі значення для всіх трьох ізоформ: CYP2C9 – 0,1178, CYP2D6 – 0,0198, CYP3A4 – 0,1784. Таке поєднання результатів означає, що за прогнозом GFB-887 може метаболізуватися переважно за участю CYP3A4 та частково CYP2C9, водночас малоімовірно виступаючи інгібітором цих ферментів. Тобто модель вказує на участь системи CYP у біотрансформації сполуки без ознак значущого пригнічення активності відповідних ізоформ. У доступних клінічних публікаціях щодо GFB-887 відсутні експериментальні *in vitro* та *in vivo* дані, які б характеризували CYP-субстратність чи CYP-інгібування у форматі, придатному для прямого порівняння [12, 140]. Отримані результати слід розглядати як прогнозні оцінки, що визначають



доцільність подальшої експериментальної перевірки участі CYP3A4 та CYP2C9 у метаболізмі сполуки.

Для GFB-887 класифікаційні прогнози токсикологічних параметрів становили: ймовірність мутагенності за тестом Еймса – 0,2113, індекс DILI – 0,8861, показник потенційного блокування hERG – 0,5244. Низьке значення тесту Еймса свідчить про низьку ймовірність генотоксичної активності сполуки. Водночас для молекул з хімічною архітектурою, подібною до GFB-887, яка не містить типових електрофільних або реакційноздатних фрагментів, такий прогноз є очікуваним на підставі загальних структурних ознак. Тому в цьому випадку прогноз тесту Еймса доцільно розглядати не як ключовий індикатор безпеки, а як підтвердження узгодженості токсикологічного прогнозу з очікуваними властивостями молекули з огляду на її хімічну будову. Для параметра DILI отримано високе значення, однак у доступних джерелах відсутні дані, придатні для прямого або опосередкованого порівняння. Для hERG отримано значення 0,5244, що відповідає ймовірній блокаді каналу; водночас у клінічному описі зазначено, що безпековий моніторинг включав електрокардіографічний контроль і не виявив клінічно значущих відхилень [12]. Це не замінює *in vitro* тесту блокування hERG, однак дозволяє зіставити наявні клінічні факти з отриманим прогнозом.

Підсумовуючи, прогнозований ADMET-профіль сполуки GFB-887 характеризується сприятливими параметрами абсорбції (висока проникність через Caco-2, HIA), помірним об'ємом розподілу у поєднанні з високим зв'язуванням з білками плазми, ймовірною участю ізоформ CYP3A4 та CYP2C9 у біотрансформації сполуки без ознак значущого інгібування цих ферментів, а також низькою прогнозованою мутагенністю за тестом Еймса. Окремі параметри, передусім DILI та hERG, потребують експериментальної верифікації з огляду на відсутність відкритих експериментальних даних для прямого порівняння, а також на клінічно значущі наслідки, пов'язані з цими параметрами.

## РОЗДІЛ 6

### ЕКСПЕРИМЕНТАЛЬНА ВАЛІДАЦІЯ ВПЛИВУ МОДУЛЯТОРІВ НА СКОРОТЛИВІСТЬ ГЛАДЕНЬКИХ М'ЯЗІВ ТОНКОЇ КИШКИ МИШІ

Структурно-обчислювальні прогнози, отримані в попередньому розділі для Pico145 та GFB-887, вимагають функціональної перевірки в нативних тканинних системах. У цьому розділі наведено результати експериментальної валідації дії обох модуляторів на TRPC4-залежні процеси у гладеньких м'язах тонкої кишки миші, виконаної методами петч-клемп та ізометричної тензометрії. Функціональну оцінку здійснено у двох взаємодоповнюючих моделях: методом петч-клемп у ізольованих міоцитах клубової кишки миші та методом ізометричної тензометрії на ізольованих сегментах тонкої кишки миші. Отримані експериментальні дані зіставлено з результатами молекулярного докінгу, що дало змогу сформувати цілісну структурно-функціональну інтерпретацію механізму дії досліджуваних сполук.

#### 6.1. Електрофізіологічне дослідження впливу Pico145 на $mI_{\text{CAT}}$

З метою експериментальної перевірки ефективності блокування Pico145 рецептор-керованих TRPC4-каналів у нативних клітинах було досліджено дію цього блокатора на  $mI_{\text{CAT}}$  в ізольованих міоцитах клубової кишки миші з використанням методу петч-клемп у стандартній конфігурації «ціла клітина». У попередніх дослідженнях детально охарактеризовано біофізичні та фармакологічні властивості  $mI_{\text{CAT}}$  [2]. Реєстрації проводилися за кімнатної температури (20–22 °C) з використанням симетричних розчинів, що містили  $\text{Cs}^+$  (125 ммоль/л), з концентрацією  $[\text{Ca}^{2+}]_i$  100 нмоль/л та буферним розчином 10 ммоль/л ВАРТА/4,6 ммоль/л  $\text{CaCl}_2$ , що давало змогу надійно ізолювати  $mI_{\text{CAT}}$  [2].

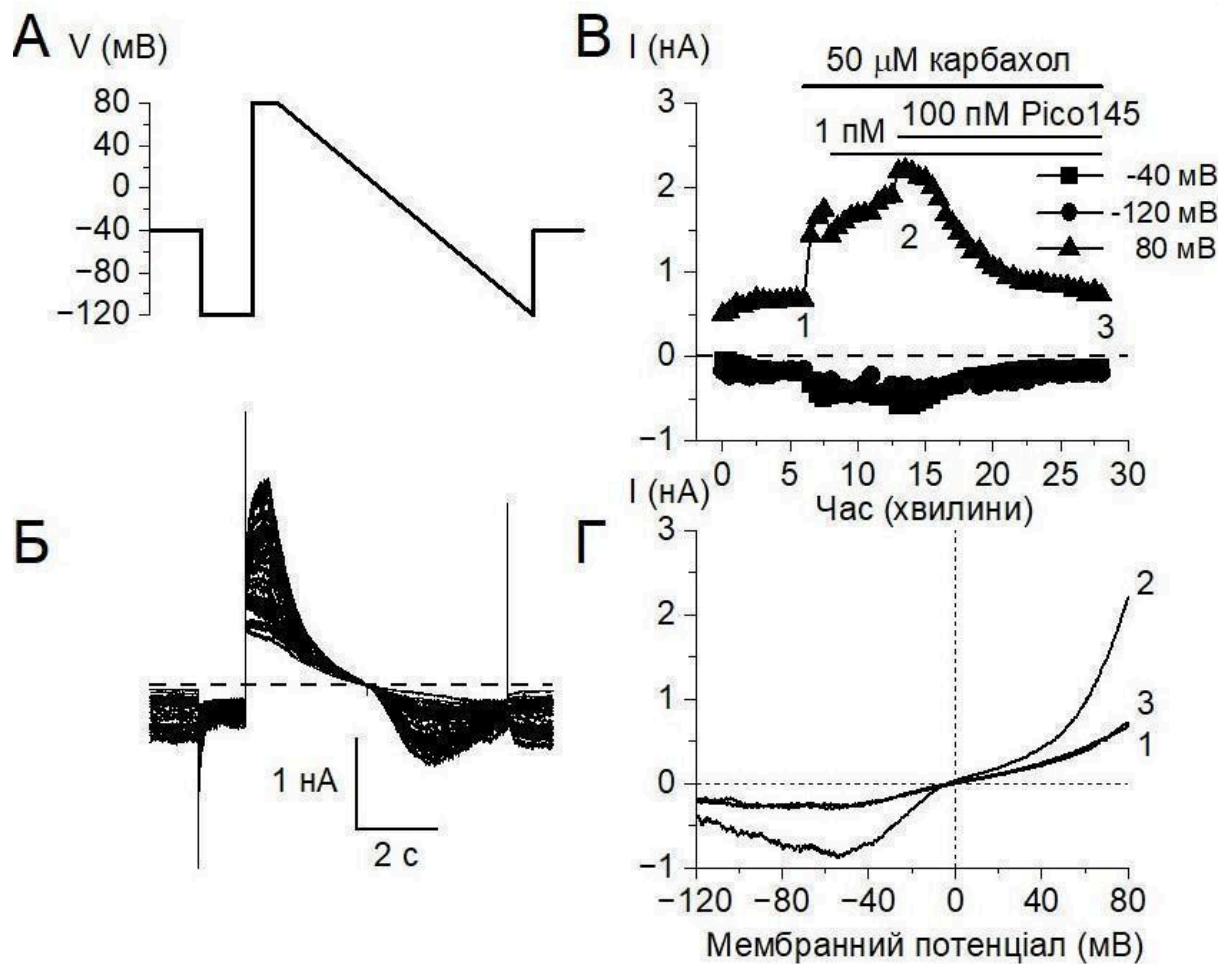


Рис. 6.1. Інгибування  $mI_{CAT}$ , зареєстрованого в міоцитах клубової кишки миші, селективним блокатором TRPC4-каналів Pico145 [120]. А: схема ramp-протоколу з поступовим зміщенням потенціалу (з інтервалом 30 с від 80 до -120 мВ протягом 6 с), що застосовувалася для дослідження біофізичних властивостей провідності каналів; Б: відповідно зареєстровані за допомогою цього протоколу оригінальні катіонні струми; В: інгибування струму з часом, яке вимірювали при -40 мВ (підтримуючий потенціал), -120 мВ та 80 мВ, як зазначено. На фоні карбахол-індукованого струму (КХ, 50 мкмоль/л) (1) спочатку аплікували 1 пМ Pico145, а потім на піку розвитку струму (2) збільшили концентрацію блокатора до 100 пМ, що викликало майже повне інгибування  $mI_{CAT}$  (3). Г: вольт-амперна характеристика  $mI_{CAT}$ , що вимірювалася за допомогою повільних ramp-імпульсів до (крива 1) і після аплікації карбахолу у контролі (крива 2) та блокування струму у присутності

Pico145 (крива 3); ці моменти часу позначені відповідними цифрами на панелі В.

Такий протокол експерименту дозволяв паралельно вимірювати зміни стаціонарної амплітуди струмів при потенціалах  $-40$ ,  $-120$  та  $+80$  мВ у часі, як проілюстровано на Рис. 6.1, В, і також стаціонарні вольт-амперні характеристики струму (Рис. 6.1, Г).  $mI_{CAT}$  був активований карбахолом (КХ), який додавали до зовнішнього розчину у субмаксимальній концентрації (50 мкмоль/л).  $mI_{CAT}$  починав пригнічуватися Pico145 вже при 1 пМ, а при додаванні Pico145 у концентрації 100 пМ спостерігалось повне блокування струму (Рис. 6.1, В, Г). При цьому інгібування розвивалось досить повільно, протягом десятків секунд (Рис. 6.1, Г), що, очевидно, пов'язано з тим, що сайт зв'язування блокатора знаходиться у площині мембрани клітини, найбільш імовірно у гідрофобній «кишені». Це може бути причиною повільної дифузії блокатора і таким чином пояснювати повільний розвиток його ефекту, що дуже контрастує з високим афінітетом у пікомолярному діапазоні. Інгібування розвивалось досить повільно, із середньою постійною часу  $140 \pm 6$  с ( $n = 9$ ). Більше того, спостерігалась потенціалзалежність блокування: при  $-120$  мВ інгібування було помітно менш ефективним порівняно з 80 мВ (Рис. 6.1, В). Це також є прямим доказом того, що інгібування TRPC4 відбувається всередині плазматичної мембрани клітини, адже сайт зв'язування молекули блокатора є чутливим до трансмембранної різниці потенціалів.

## **6.2. Тензометричне дослідження впливу GFB-887 на скоротливість тонкої кишки**

Оцінку впливу GFB-887 на карбахол-індуковану скоротливість гладеньких м'язів тонкої кишки миші проводили методом ізометричної

тензометрії, що дало змогу охарактеризувати функціональні ефекти сполуки у тканинній моделі *ileum* тонкої кишки миші *in vitro*. З наших власних та досліджень інших авторів добре відомо, що центральним компонентом холінергічного збудження і скорочення гладеньких м'язів кишечника є саме рецептор-керований TRPC4-канал, тому саме ця експериментальна модель була обрана для дослідження фармакологічних ефектів GFB-887 на TRPC4-канали у нативних клітинах [1, 19, 62].

Перед застосуванням агоніста мускаринових ацетилхолінових рецепторів карбахолу (КХ) підтверджували збереження функціональної активності препарату шляхом додавання гіперкалієвого розчину з концентрацією  $K^+$  60 ммоль/л, що викликало виражене скорочення та свідчило про фізіологічну інтактність і активність тканини (Рис. 6.2).

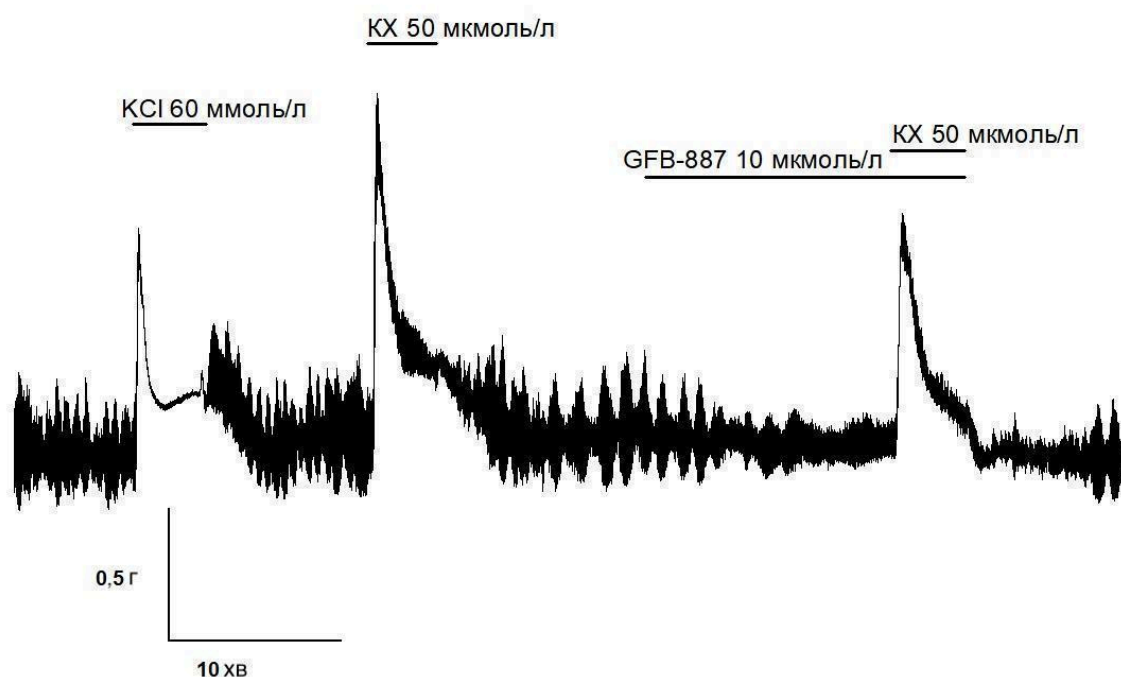


Рис. 6.2. Демонстративна тензометрична реєстрація пригнічення інгібітором TRPC4-каналів GFB-887 (10 мкмоль/л) карбахол-викликаних (КХ, 50 мкмоль/л) фазних та тонічних скорочень тонкої кишки миші [121].

Після відмивання гіперкалієвого розчину аплікація карбахолу (50 мкмоль/л) спричиняла пікову фазну скоротливу відповідь із наступним плато, тобто тонічним скороченням. Далі, після проведення контрольного тесту з карбахолом, препарат відмивали у розчині Кребса протягом 20 хв, після чого до перфузійного середовища додавали GFB-887 у концентрації 10 мкмоль/л. По завершенні 20-хвилинної інкубації у розчині з GFB-887 повторно додавали карбахол у тій самій концентрації, що й у контролі (50 мкмоль/л). За цих умов скоротлива реакція на карбахол після попередньої обробки GFB-887 була помітно зниженою: амплітуда карбахол-індукованого скорочення зменшувалася з  $1,44 \pm 0,05$  г у контролі до  $0,87 \pm 0,07$  г на фоні GFB-887 (mean  $\pm$  SEM,  $n = 3$ ), що відповідає зниженню скоротливої відповіді на  $39,80 \pm 2,94\%$  (до  $\approx 60\%$  контрольного рівня). Різниця статистично значуща ( $p < 0,05$ , парний  $t$ -критерій Стьюдента) (рис. 6.2). Виявлене тензометрією пригнічення холінергічно індукованої скоротливості узгоджується з клінічними проявами порушень з боку ШКТ та підкреслює значущість ацетилхолін-залежного шляху ініціації скорочень кишечника.

### 6.3. Співставлення результатів *in silico* та *in vitro* досліджень

Отримані експериментальні дані зіставлено з результатами молекулярного моделювання (розділ 4), що дозволяє поєднати електрофізіологічні та тензометричні спостереження та результати комп'ютерного моделювання у цілісну картину механізму дії досліджуваних модуляторів TRPC4.

Встановлені у підрозділі 6.1 особливості дії Pico145 на  $mI_{\text{CAT}}$ , пікомолярна ефективність блокади, повільна кінетика ( $\tau \approx 140$  с) та потенціалзалежність, вказують, що взаємодія ліганду з каналом відбувається у гідрофобному трансмембранному середовищі, де дифузія до сайту обмежена, а сама взаємодія чутлива до трансмембранної різниці потенціалів.

Ці функціональні ознаки відповідають глибоко локалізованому мембранному сайту зв'язування Pico145 у структурі TRPC4. Сайт такого типу вперше експериментально ідентифіковано для комплексів TRPC5 і TRPC1/TRPC4 людини у міжсубодиничній гідрофобній кишені трансмембранного домену, з гіпотезою про його консервативність у підродині TRPC1/4/5 [10, 11].

Молекулярне моделювання Pico145 з мишачим TRPC4 (PDB ID: 5Z96) відтворило позу ліганда у тій самій міжсубодиничній гідрофобній кишені з характерною орієнтацією ксантинового ядра, а єдина відмінність оточення (заміна Thr599 у миші на Val624 у людині) стосується полярної складової контактної мережі, а не структурного каркаса кишені, і має підсилюючий, а не визначальний характер, що загалом свідчить про консервативність сайту між ортологами. Незалежно від цього, функціональні характеристики блокади  $mI_{CAT}$  узгоджуються з локалізацією сайту у площині мембрани. Сукупно структурні та функціональні дані підтримують гіпотезу про консервативний механізм зв'язування Pico145 у межах підгрупи TRPC1/4/5.

Тензометричне тестування GFB-887 виявило пригнічення карбахол-індукованої скоротливості *ileum* тонкої кишки миші на  $\approx 40\%$  (до  $\approx 60\%$  контролю) після інкубації з GFB-887 (10 мкмоль/л). З огляду на центральний внесок TRPC4-каналів у холінергічне спряження збудження і скорочення вісцеральних гладеньких м'язів цей ефект відповідає фармакологічному блокуванню TRPC4 і підтримується структурними даними розділу 4: докінг GFB-887 у структуру 5Z96 дав енергетично сприятливу позу ( $-8,11$  ккал/моль) у трансмембранному регіоні S1–S4, узгоджену з кріо-ЕМ структурами комплексів TRPC4 *Danio rerio* з інгібіторами серії GFB (PDB ID: 7B05, 7B0S, 7B16) та TRPC5 людини з клемізолом, для яких показано локалізацію ліганду у VSL-домени [27, 32].

Зіставлення обох сполук виявляє спільні закономірності фармакологічного модулювання TRPC4. По-перше, отримані функціональні ефекти незалежно підтверджують роль TRPC4 як основної молекулярної ланки холінергічного збудження і скорочення гладеньких м'язів кишечника.

По-друге, в обох випадках експериментальні дані підтверджують прогноз *in silico*: сайти зв'язування локалізовані всередині трансмембранного домену (міжсубодинична гідрофобна кишеня для Pico145 та VSL-домен для GFB-887), а не у цитозольних ділянках, що узгоджується з повільною кінетикою ефекту Pico145, зумовленою обмеженою дифузією ліганду у ліпідному середовищі. При цьому у наших попередніх дослідженнях Pico145 дозозалежно послаблював спонтанні та карбахол-індуковані скорочення сегментів тонкої кишки й уповільнював кишковий транзит, що функціонально збігається з тензометричним ефектом GFB-887 [20]. По-третє, відмінності в активності (пікомолярна для Pico145 проти мікромолярної для GFB-887) і в оточенні сайтів відповідають різним хімічним класам (ксантинове ядро Pico145 проти піридазинон-похідного GFB-887) і відкривають можливість розробки різних за механізмом TRPC4-селективних модуляторів. У сукупності результати формують структурно-функціональну основу для раціонального дизайну селективних блокаторів TRPC-каналів і демонструють застосовність інтегрованого *in silico* / *in vitro* підходу до характеристики модуляторів TRPC4.



## РОЗДІЛ 7

### УЗАГАЛЬНЕННЯ РЕЗУЛЬТАТІВ РОБОТИ

Дисертаційна робота присвячена встановленню особливостей молекулярних взаємодій двох структурно та фармакологічно різних інгібіторів Pico145 і GFB-887 з рецептор-керованим катіонним каналом TRPC4 шляхом поєднання обчислювального моделювання з функціональним тестуванням у нативних гладеньком'язових препаратах тонкої кишки миші. Концептуальною особливістю роботи є послідовне об'єднання чотирьох методичних блоків: гібридного методу пріоритизації сайтів зв'язування на основі геометричного аналізу білкових порожнин і витягу літературних даних великими мовними моделями, обчислювальної оцінки ADMET-параметрів методом Mol2Vec Best, молекулярного докінгу і молекулярної динаміки модуляторів TRPC4 та функціонального тестування у препаратах *ileum* миші [14, 17, 120, 121]. Кожен блок самостійно вирішує окрему задачу обчислювального дизайну ліків, а їх послідовне застосування формує переносний методичний підхід до структурно-функціональної характеристики ліганд-білкової взаємодії, придатний до перенесення на інших представників родини TRP-каналів та інші мембранні білкові мішені.

Розроблений гібридний метод пріоритизації сайтів зв'язування поєднує геометричне виявлення кандидатних кишень за допомогою Focket з літературно-керованою валідацією амінокислотних залишків, витягнутих із публікацій моделлю GPT-4o-mini [14, 77]. Його принципова перевага полягає у заміні тривалого суб'єктивного ручного аналізу літератури автоматизованим відтворюваним протоколом, незалежним від наявності співкристалізованих лігандів у структурах PDB, на відміну від ресурсів структурної анотації BioLip2 та Q-BioLiP [109, 110]. На бенчмарк-датасеті з 10 різнотипних мішеней і 31 публікації повна конфігурація конвеєра демонструє Pocket Number Accuracy 0,710, Pocket Recall 1,0 і Pocket Specificity 0,657, а етап фільтрації забезпечує нульовий FNR при FPR близько

0,363 [14]. Ці метрики свідчать, що жодна релевантна публікація не втрачається на вході конвеєра, а вихідний набір кишень з високою імовірністю містить усі експериментально підтверджені сайти за раціонально низької частки хибнопозитивних. Бенчмарк, програмний код і протокол опубліковано у відкритому доступі (GitHub та Zenodo), що робить інструмент готовим до повторного застосування іншими групами без додаткової розробки чи перенавчання.

Застосування методу до мишачої кріо-ЕМ структури TRPC4 (PDB ID: 5Z96) дозволило з 267 кандидатних кишень Focket виокремити чотири біологічно унікальні сайти, кожен представлений чотирма еквівалентними копіями відповідно до гомотетрамерної симетрії каналу: катіон-зв'язувальний сайт у межах регіону S1–S4 кожної субодиниці (частина VSL-домену); сайт зв'язування фосфатидної кислоти на інтерфейсі порової спіралі і спіралі S6 сусідніх субодиниць;  $\text{Ca}^{2+}$ -зв'язувальний сайт у VSL-домени; та сайт у VSL-домени, у якому локалізовано зв'язування низькомолекулярних інгібіторів серії GFB. Принципово, що три з чотирьох сайтів є складовими VSL-домену TRPC4, який у цьому каналі не виконує класичної функції потенціал-сенсорного гейтингу, але формує ключову структурно-регуляторну і фармакологічну область каналу. Кільцева РА-кишеня розташована поза VSL-доменом в області порового домену і визначає ліпідне оточення каналу у мембрані. Отже, BMM-керована валідація скоротила геометричний надлишок Focket до набору сайтів з прямою експериментальною підтримкою у літературі без апріорного використання інформації про конкретні ліганди.

Молекулярне моделювання комплексу TRPC4:Pico145 виконано з попередньою валідацією докінг-протоколу на референтній структурі гетеротетрамеру TRPC1/TRPC4 людини з Pico145 (PDB ID: 8WPM), після чого ліганд вдоковано в апо-структуру мишачого TRPC4 (PDB ID: 5Z96). Отримана модель з енергією зв'язування  $-9,9$  ккал/моль локалізує Pico145 у міжсубодиниичній гідрофобній кишені, що відповідає консервативному ксантин-зв'язувальному сайту, вперше ідентифікованому для людського

TRPC5 [11]. Детальний перелік контактів ліганда у комплексі з мишачим TRPC4 наведено у підрозділі 4.1 [120]. Порівняння оточення кишені між мишачим TRPC4 і людським TRPC1/TRPC4 виявило високу консервативність контактних залишків за наявності лише однієї напівконсервативної заміни Thr599 (миша) на Val624 (людина). Ця заміна стосується не структурного каркаса кишені, а її полярної складової: гідроксильна група Thr599 безпосередньо задіяна в контактах із лігандом, тоді як у людській структурі відповідна позиція гідрофобна. Втім, вона має підсилюючий, а не визначальний характер, оскільки Pico145 зв'язується і з людським комплексом попри цю заміну. Отже, отримані результати поширюють гіпотезу Wright *et al.* про консервативність міжсубодиничного ксантин-зв'язувального сайту підгрупи TRPC1/4/5 на мишачий ортолог TRPC4, для якого структурні дані у комплексі з Pico145 наразі відсутні [11].

Оцінка стабільності комплексу TRPC4:Pico145 методом молекулярної динаміки впродовж  $\approx 215$  нс у POPC-бішарі (силове поле CHARMM36m, GROMACS) показала збереження ключових контактів у всіх чотирьох симетричних копіях [91, 93]. Дві копії (PICO 1, PICO 2) демонструють стаціонарний RMSD 1,5–2,5 Å, тоді як дві інші (PICO 3, PICO 4) – дискретні «сходінкові» переходи, що відповідають фліпам ароматичних кілець і обертанням навколо зв'язків C–C трифторметоксифенільного фрагмента; у жодному випадку ліганд не залишає кишеню. Профіль водневих зв'язків стабільний у часі ( $\approx 1,5$ – $2,0$  контактів на ліганд) із сприятливою геометрією (підрозділ 4.2). Систематична розбіжність профілів RMSD між парами PICO 1/PICO 2 та PICO 3/PICO 4 при ідентичному старті вказує, що повне семплування внутрішньолігандних ступенів свободи потребує продовження траєкторії; водночас отриманої тривалості достатньо для висновку про стабільність зв'язування у міжсубодиничній гідрофобній кишені.

Структурна основа взаємодії GFB-887 з TRPC4 принципово відмінна: структурні дані у комплексі з будь-яким представником підгрупи TRPC1/4/5 для цієї сполуки відсутні, що робить обчислювальне моделювання єдиним

практичним інструментом встановлення сайту зв'язування. Сайт GFB-887 локалізовано у трансмембранному регіоні S1–S4 (VSL-домен) TRPC4 з оточенням, ідентифікованим розробленим методом пріоритизації; докінг у структуру 5Z96 дав енергетично сприятливу позу (–8,11 ккал/моль) з мережею гідрофобно-ароматичних і полярних контактів (ключові залишки Phe413, Tyr373, Tyr646, His369, Glu417, Asp438, Asn442, Arg491, Ser494; детально – підрозділ 4.3) [121]. Отримана локалізація просторово сумісна з кріо-ЕМ структурами TRPC4 *Danio rerio* у комплексах з аналогами серії GFB, такими як, GFB-8438, GFB-8749 і GFB-9289 (PDB ID: 7B0S, 7B05, 7B16) та з TRPC5 людини у комплексі з клемізолом, для якого встановлено аналогічну локалізацію у VSL-домени зі стабілізацією закритої конформації каналу [27, 32]. Молекулярна динаміка комплексу TRPC4:GFB-887 у POPC-бішарі (~200 нс) підтвердила стабільність полярних «якорів» (His369, Asn442, Arg491, Ser494) і збереження гідрофобно-ароматичного оточення (Phe413, Tyr373, Tyr646, Leu492/Leu495) упродовж усієї траєкторії.

Зіставлення моделей TRPC4:Pico145 і TRPC4:GFB-887 показує, що обидва інгібітори локалізуються у трансмембранному домені, але займають структурно різні кишені: Pico145 – міжсубодиничну гідрофобну порожнину на інтерфейсах сусідніх субодиниць, GFB-887 – внутрішньосубодиничний сайт VSL-домени. Ця відмінність відповідає двом різним молекулярним механізмам інгібування TRPC4 і пояснює розбіжності у профілях контактів. Практично це відкриває можливість раціонального дизайну як селективних лігандів окремого сайту, так і подвійних лігандів з комбінованою дією на дві структурно відокремлені кишені одного каналу. З урахуванням функціональних даних, що підтверджують ефективність обох механізмів, отримана пара моделей є основою для побудови фармакофорних моделей і скринінгу віртуальних бібліотек у напрямі сполук другого покоління.

Розроблений метод прогнозування ADMET-параметрів Mol2Vec Best базується на перенавченні Mol2Vec на корпусі 883 897 271 сполук ZINC20 з підвищенням розмірності ембедингу до 512 і конкатенуванні з розширеним

набором класичних дескрипторів і фінгерпринтів (Mordred, MCE-18, Morgan, RDKit, Layered, Atom, Topological, Avalon, PharmBase, MACCS keys), що формує простір 94 729 кандидатних ознак на кожну ADMET-задачу [17]. Чотириступеневий конвеєр відбору ознак (видалення константних ознак, фільтрація за дисперсією, кореляційний відбір, оцінка важливості за Random Forest) скорочує цей простір до задач-специфічних підмножин (~600–8000 ознак), які подаються на 16 окремих MLP-моделей з оптимізацією гіперпараметрів засобами Optuna TPE [124]. На бенчмарках Therapeutics Data Commons Mol2Vec Best посідає перше місце у 10 з 16 задач – абсолютний максимум серед моделей лідерборду TDC на дату оцінки (для порівняння: MapLight + GNN – 8, ZairaChem – 3, MapLight – 2, DeepPurpose – жодного) [17, 97, 100]. Стабільна першість простого MLP на збагачених ознаках над графовими нейронними мережами і підходами на основі мовних моделей підтверджує, що добре сконструйований простір ознак з принциповим відбором конкурує з істотно складнішими архітектурами при значно нижчих обчислювальних вимогах.

Прикладну цінність методу продемонстровано на GFB-887, для якої у відкритих джерелах відсутні експериментальні значення стандартних ADMET-параметрів. Прогнозний профіль характеризується сприятливою абсорбцією ( $C_{aco-2} P_{app} 1,46 \times 10^{-5}$  см/с; HIA 0,9735), типовою для тривало циркулюючих сполук конфігурацією розподілу (зв'язування з білками плазми 95,8 %; VDss 0,95 л/кг), імовірною метаболізацією через CYP3A4 і частково CYP2C9 без ознак значущого інгібування цих ферментів та низькою мутагенністю за тестом Еймса (0,2113) [121]. Параметри DILI (0,8861) і блокування hERG (0,5244) визначено як пріоритетні для експериментальної верифікації з огляду на клінічну значущість цих параметрів. Прогнози узгоджуються з опосередкованими клінічними спостереженнями для GFB-887: фармакокінетичним профілем, сумісним із режимом одноразового добового дозування у дослідженні TRACTION-2, вимірюваною концентрацією препарату у плазмі та тканинах органів у щурячих моделях

і відсутністю клінічно значущих відхилень за електрокардіографічним моніторингом у дослідженні «перший у людині» [12, 141]. Це узгодження засвідчує практичну придатність методу для оцінки фармакокінетичного профілю сполук, для яких прямі експериментальні ADME-дані недоступні.

Експериментальна валідація виконана у двох взаємодоповнюючих моделях відповідно до ролі TRPC4 як ключової молекулярної ланки  $mI_{CAT}$  у вісцеральних гладеньких м'язах ШКТ [2, 3]. Методом петч-клемп у конфігурації «ціла клітина» в ізольованих міоцитах клубової кишки миші встановлено, що Pico145 блокує TRPC4-опосередкований  $mI_{CAT}$  у пікомолярному діапазоні (повне блокування при 100 пМ) з повільною кінетикою ( $\tau = 140 \pm 6$  с,  $n = 9$ ) та вираженою потенціалзалежністю (підрозділ 6.1) [120]. Сукупність ознак, а саме висока афінність, повільна кінетика і чутливість до трансмембранної різниці потенціалів є прямим функціональним свідченням локалізації сайту у площині мембрани з потребою дифузії ліганду через ліпідне середовище, що узгоджується саме з обчислювально передбаченим міжсубодиничним гідрофобним сайтом у трансмембранному домені каналу. Раніше показана здатність Pico145 дозозалежно послаблювати спонтанні і карбахол-індуковані скорочення сегментів тонкої кишки та уповільнювати кишковий транзит додатково підтверджує функціональну значущість блокування TRPC4 цією сполукою [20].

Методом ізометричної тензометрії на ізольованих сегментах *ileum* миші встановлено, що 20-хвилинна інкубація з GFB-887 (10 мкмоль/л) пригнічує карбахол-індуковану (50 мкмоль/л) скоротливу відповідь на  $\approx 40\%$  (до  $\approx 60\%$  контролю), що є першим функціональним свідченням фармакологічної активності GFB-887 у TRPC4-залежних нативних тканинах і водночас підтвердження обчислювально передбаченого VSL-сайту [121]. Зіставлення обох інгібіторів показує, що, попри відмінності у афінності (пікомолярна для Pico145 проти мікромолярної для GFB-887) і в оточенні сайтів, обидві сполуки спрямовані на один функціональний процес –

TRPC4-опосередковане холінергічне збудження гладеньких м'язів кишечника, що зберігає чутливість до інгібіторів двох різних хемотипів і двох структурно відокремлених сайтів зв'язування.

Інтегрування обчислювальних прогнозів і функціональних спостережень у єдину структурно-функціональну картину дозволяє сформулювати висновки поза межами власне TRPC4-фармакології. По-перше, функціональні характеристики інгібування  $mI_{CAT}$  Pico145 повністю узгоджуються з існуванням саме міжсубодиничного гідрофобного сайту, раніше ідентифікованого у TRPC5 і TRPC1/TRPC4; таким чином, електрофізіологічні спостереження підтверджують структурні гіпотези, сформульовані на основі кріо-ЕМ, а молекулярне моделювання надає їм просторове пояснення для мишачого TRPC4, що дозволяє розглядати запропонований механізм інгібування Pico145 як універсальний для каналів підгрупи TRPC1/4/5 [10, 11]. По-друге, для GFB-887 виявлений функціональний ефект на скоротливість тонкої кишки у поєднанні з обчислювально передбаченим VSL-сайтом формує доказову базу для розгляду TRPC4 як перспективної терапевтичної мішені при порушеннях моторики ШКТ, зокрема при синдромі подразненого кишечника, – тематично новому напрямі застосування сполуки порівняно з її основною клінічною орієнтацією на патології нирок [12].

Практичну значущість результатів доцільно розглядати на трьох рівнях. На рівні методології чотириблочна стратегія (пріоритизація сайтів зв'язування, моделювання міжмолекулярних взаємодій, ADMET-прогноз і функціональне тестування) формує методичний шаблон структурно-орієнтованого пошуку модуляторів іонних каналів. На рівні інструментів гібридний метод пріоритизації оприлюднено як рецензований програмний інструмент з відкритим кодом (GitHub) з присвоєнням DOI у архіві Zenodo, а ADMET-модель Mol2Vec Best представлено у форматі препринту bioRxiv з повним описом протоколу і програмною реалізацією за обґрунтованим запитом, що забезпечує придатність обох інструментів для

академічних та індустріальних груп без перенавчання чи адаптації під конкретну мішень. На рівні структурної фармакології TRPC4 побудовані моделі комплексів з Pico145 і GFB-887, ідентифіковані критичні залишки і підтверджена консервативність ключового сайту між мишачим і людським ортологами формують підґрунтя для раціонального дизайну селективних модуляторів каналів підгрупи TRPC1/4/5.

Водночас слід окреслити обмеження застосованих підходів. Метод пріоритизації сайтів валідовано для конкретної мовної моделі GPT-4o-mini, і його перенесення на альтернативні ВММ потребує повторної валідації; крім того, конвеєр успадковує обмеження геометричного предиктора Focket у випадках, коли дві фізично відокремлені кишені об'єднано у спільну геометричну порожнину. Для ADMET-моделі Mol2Vec Best характерна нижча ефективність на параметрах Caco-2 і VDss порівняно з графовими нейронними мережами, що вказує на доцільність ансамблевого поєднання архітектур. Молекулярний докінг у AutoDock 4 виконано з жорсткою моделлю білка, що обмежує моделювання індукованих лігандом конформаційних перебудов кишені; цей фактор частково компенсовано подальшою молекулярною динамікою повнорозмірних комплексів у РОРС-бішарі, проте остаточної валідація моделей потребує експериментальної структурної верифікації. Функціональні дані на нативних препаратах ілеум миші, зокрема петч-клемп реєстрації трансмембранних струмів, надійно констатують фармакологічну активність сполук у TRPC4-залежних процесах, але є менш прямим свідченням щодо молекулярних деталей зв'язування; для підтвердження ролі конкретних залишків необхідні спрямовані мутаційні дослідження. Ці обмеження відображають сучасний стан методологічного арсеналу обчислювально-структурного дизайну ліків.

Перспективи розвитку підходу окреслюються у трьох напрямках. Перший – це розширення бенчмарк-датасету пріоритизації сайтів за рахунок додаткових мішеней і типів сайтів, інтеграція з базою BioLip2 для перехресної валідації витягнутих залишків та випробування альтернативних



геометричних предикторів і архітектур BMM. Другий – ансамблеве комбінування Mol2Vec Best з графовими нейронними мережами для підвищення якості прогнозу на параметрах, де поточна реалізація поступається GNN, та розширення переліку охоплених параметрів за межі 16 бенчмарків TDC. Третім напрямом є повномасштабні розрахунки молекулярної динаміки комплексів TRPC4 з Pico145 і GFB-887 з істотно більшою тривалістю траєкторій для повного семплювання внутрішньолігандних ступенів свободи у POPC-бішарі із застосуванням силового поля CHARMM36m у GROMACS. Ці напрями не змінюють загальної архітектури підходу, а лише розширюють сферу застосування і підвищують точність окремих складових. Перенесення підходу на інших представників родини TRP-каналів та інші класи іонних каналів є логічним продовженням роботи і відкриває шлях до структурно-функціональної характеристики широкого спектру мішеней мембранної фармакології.

Узагальнюючи, у дисертаційній роботі вперше побудовано молекулярні моделі комплексів мишачого TRPC4 з інгібіторами Pico145 і GFB-887, встановлено, що Pico145 зв'язується у консервативній міжсубодиничній гідрофобній кишені підгрупи TRPC1/4/5, а GFB-887 у внутрішньосубодиничній VSL-кишені, охарактеризовано кінетичні та потенціалзалежні особливості блокади Pico145 рецептор-керованого катіонного струму  $mI_{CAT}$  і експериментально показано пригнічуючий вплив GFB-887 на карбахол-індуковану скоротливість тонкої кишки миші. Розроблені обчислювальні інструменти, такі як, гібридний метод пріоритизації сайтів зв'язування і метод ADMET-профілювання Mol2Vec Best мають універсальний характер, оприлюднені у відкритому доступі і покривають два важливі етапи пошуку лікарських засобів, що забезпечує їх самостійну наукову та практичну цінність і можливість інтеграції у дослідження інших наукових груп.

## ВИСНОВКИ

У дисертаційній роботі встановлено молекулярні механізми інгібування рецептор-керованого катіонного каналу TRPC4 селективними модуляторами Pico145 і GFB-887 шляхом поєднання обчислювального моделювання (ідентифікація сайтів зв'язування, молекулярний докінг, молекулярна динаміка, ADMET-профілювання) з функціональними дослідженнями у нативних гладеньком'язових препаратах тонкої кишки миші методами петч-клемп та ізометричної тензометрії. У роботі розроблено й застосовано низку нових обчислювальних інструментів, а саме гібридний метод пріоритизації сайтів зв'язування та набір моделей для прогнозування ADMET-параметрів, ці підходи поєднано з такими відомими *in silico* методами, як молекулярний докінг, молекулярна динаміка, а також функціональним тестуванням. Застосований комбінований підхід придатний до застосування на інших мембранних білкових мішенях.

1. Проаналізовано потенційні сайти зв'язування Pico145 і GFB-887 у кріо-електронно-мікроскопічній структурі TRPC4 миші із застосуванням розробленого гібридного методу пріоритизації сайтів зв'язування. З 267 геометричних кандидатів Focket без використання інформації про ліганд виокремлено чотири структурно та функціонально значущі сайти: катіон-зв'язувальний сайт у цитоплазматичній частині S1–S4, кільцеву кишеню зв'язування фосфатидної кислоти,  $\text{Ca}^{2+}$ -зв'язувальний сайт у VSL-домені та інгібіторну VSL-кишеню GFB-сполук. Ідентифіковані сайти проаналізовані, частина з них були пріоритизовані для подальшого молекулярного докінгу. Бенчмарк, програмний код і повний протокол опубліковано у відкритому доступі, що робить метод придатним до повторного використання іншими дослідницькими групами для пріоритизації сайтів зв'язування на інших мембранних і розчинних білкових мішенях, зокрема, у задачах раціонального дизайну лігандів, де

експериментальна структурна інформація про комплекс з лігандом відсутня або неповна.

2. Виконано молекулярний докінг Pico145 у структуру TRPC4 миші, який відтворив експериментально спостережувану позу ліганда у консервативній міжсубодиничній гідрофобній кишені підродиної TRPC1/4/5 з енергією зв'язування  $-9,9$  ккал/моль. Зіставлення амінокислотного оточення кишені між мишачим TRPC4 і людським TRPC1/TRPC4 виявило високий ступінь консервативності контактних залишків за наявності лише однієї напівконсервативної заміни Thr599 (миша) на Val624 (людина). Оцінка стабільності комплексу методом молекулярної динаміки підтвердила збереження ключових контактів ліганда з міжсубодиничною гідрофобною кишенею впродовж усієї траєкторії моделювання ( $\sim 215$  нс). Також виконано молекулярний докінг GFB-887 у структуру TRPC4 миші, внаслідок якого отримано комплекс з лігандом у внутрішньосубодиничній VSL-кишені (регіон S1–S4) з енергією зв'язування  $-8,11$  ккал/моль. В отриманій моделі ліганд формує мережу гідрофобно-ароматичних та полярних контактів із залишками, ключовими серед яких є Phe413, Tyr373, Tyr646, His369, Asn442, Arg491, Ser494; отримана поза узгоджується з наявними кріо-ЕМ даними для комплексів TRPC4 з іншими інгібіторами серії GFB (PDB ID: 7B0S, 7B05, 7B16). Молекулярно-динамічне моделювання комплексу TRPC4–GFB-887 продемонструвало стабільність зазначених контактів і утримання ліганда у VSL-кишені протягом усієї траєкторії ( $\sim 200$  нс).
3. Оцінено ADMET-властивості Pico145 та GFB-887 за допомогою власноруч розробленого методу прогнозування Mol2Vec Best на основі молекулярних ембедингів та автоматизованого відбору ознак з простору класичних дескрипторів. Метод реалізовано як універсальний обчислювальний інструмент.
4. Методом петч-клемп у конфігурації «ціла клітина» в ізольованих міоцитах клубової кишки миші встановлено, що Pico145 блокує

TRPC4-опосередкований  $mI_{CAT}$  у пікомолярному діапазоні з повільною кінетикою розвитку ефекту ( $\tau = 140 \pm 6$  с,  $n = 9$ ) та вираженою потенціалзалежністю, що узгоджується з інгібуванням у міжсубодиничному гідрофобному трансмембранному сайті, на який очевидно мають впливати зміни мембранного потенціалу.

5. Методом ізометричної тензометрії на ізольованих сегментах *ileum* миші показано, що попередня 20-хвилинна інкубація з GFB-887 (10 мкмоль/л) пригнічує карбахол-індуковану (50 мкмоль/л) скоротливу відповідь на  $39,80 \pm 2,94\%$  ( $n = 3$ ), що узгоджується з центральною роллю TRPC4 у холінергічному збудженні гладеньких м'язів тонкої кишки та гіпотезою про інгібувальну дію GFB-887 через зв'язування у VSL-кишені каналу.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Wang H, Cheng X, Tian J. et al. TRPC channels: structure, function, regulation and recent advances in small molecular probes. *Pharmacol. Ther.* 2020;209:107497.
2. Zholos AV. Regulation of TRP-like muscarinic cation current in gastrointestinal smooth muscle with special reference to PLC/InsP3/Ca<sup>2+</sup> system. *Acta Pharmacol. Sin.* 2006;27(7):833–842.
3. Zholos AV, Melnyk MI, Dryn DO. Molecular mechanisms of cholinergic neurotransmission in visceral smooth muscles with a focus on receptor-operated TRPC4 channel and impairment of gastrointestinal motility by general anaesthetics and anxiolytics. *Neuropharmacology.* 2024;242:109776.
4. Liu H, Fu M, Zhang Y. et al. Small molecules targeting canonical transient receptor potential channels: an update. *Drug Discov. Today.* 2024;29(5):103951.
5. Zholos AV, Bolton TB. Muscarinic receptor subtypes controlling the cationic current in guinea-pig ileal smooth muscle. *Br. J. Pharmacol.* 1997;122(5):885–893.
6. Zholos AV. Muscarinic effects on ion channels in smooth muscle cells. *Neurophysiology.* 1999;31:173–187.
7. Freichel M, Suh SH, Pfeifer A. et al. Lack of an endothelial store-operated Ca<sup>2+</sup> current impairs agonist-dependent vasorelaxation in TRP4<sup>-/-</sup> mice. *Nat. Cell Biol.* 2001;3:121–127.
8. Seo JY, Jo HR, Lee SH. et al. TRPC4 deletion elicits behavioral defects in sociability by dysregulating expression of microRNA-138-2. *iScience.* 2024;27:108617.
9. Yang LP, Jiang FJ, Wu GS. et al. Acute treatment with a novel TRPC4/C5 channel inhibitor produces antidepressant and anxiolytic-like effects in

- mice. *PLoS One*. 2015;10(8):e0136255. Available from: <https://doi.org/10.1371/journal.pone.0136255>.
10. Won J, Kim J, Kim J. et al. Cryo-EM structure of the heteromeric TRPC1/TRPC4 channel. *Nat. Struct. Mol. Biol.* 2025;32(2):326–338.
  11. Wright DJ, Simmons KJ, Johnson RM. et al. Human TRPC5 structures reveal interaction of a xanthine-based TRPC1/4/5 inhibitor with a conserved lipid binding site. *Commun. Biol.* 2020;3:704. Available from: <https://doi.org/10.1038/s42003-020-01437-8>.
  12. Walsh L, Reilly JF, Cornwall C. et al. Safety and efficacy of GFB-887, a TRPC5 channel inhibitor, in patients with focal segmental glomerulosclerosis, treatment-resistant minimal change disease, or diabetic nephropathy: TRACTION-2 trial design. *Kidney Int. Rep.* 2021;6(10):2575–2584.
  13. Kim SH, Park H, Kim J. et al. Molecular architecture of the human TRPC1/C5 heteromeric channel. *Nat. Commun.* 2026;17(1):317. <https://doi.org/10.1038/s41467-025-67024-9>
  14. Stratiichuk R, Melnychenko M, Koleyev I. et al. Leveraging large language models for literature-driven prioritization of protein binding pockets. *Bioinformatics.* 2025;41(8):btaf449. Available from: <https://doi.org/10.1093/bioinformatics/btaf449>.
  15. LLM workflow for pocket prediction [Internet]. GitHub. Available from: <https://github.com/receptor-ai/LLM-benchmark-dataset>.
  16. LLM workflow for pocket prediction [Internet]. Zenodo. Available from: <https://doi.org/10.5281/zenodo.15798647>.
  17. Stratiichuk R, Shevchuk N, Kyrylenko R. et al. Improving ADMET prediction with descriptor augmentation of Mol2Vec embeddings. *bioRxiv* [Preprint]. 2025. Available from: <https://doi.org/10.1101/2025.07.14.664363>.
  18. Clapham DE. TRP channels as cellular sensors. *Nature*. 2003;426:517–524.

19. Tsvilovskyy VV, Zholos AV, Aberle T. et al. Deletion of TRPC4 and TRPC6 in mice impairs smooth muscle contraction and intestinal motility in vivo. *Gastroenterology*. 2009;137(4):1415–1424.
20. Dryn DO, Melnyk MI, Bon RS. et al. Pico145 inhibits TRPC4-mediated mICAT and postprandial small intestinal motility. *Biomed. Pharmacother.* 2023;168:115672. Available from: <https://doi.org/10.1016/j.biopha.2023.115672>.
21. Bröker-Lai J, Kollewe A, Schindeldecker B. et al. Heteromeric channels formed by TRPC1, TRPC4 and TRPC5 define hippocampal synaptic transmission and working memory. *EMBO J.* 2017;36(18):2770–2789.
22. Tai C, Chang CW, Yu GQ. et al. TRPC4 and GIRK channels underlie neuronal coding of firing patterns that reflect Gq/11–Gi/o coincidence signals of variable strengths. *Proc. Natl. Acad. Sci. U. S. A.* 2022;119(20):e2120870119. Available from: <https://doi.org/10.1073/pnas.2120870119>.
23. Tiruppathi C, Freichel M, Vogel SM. et al. Impairment of store-operated  $\text{Ca}^{2+}$  entry in TRPC4<sup>-/-</sup> mice interferes with increase in lung microvascular permeability. *Circ. Res.* 2002;91(1):70–76.
24. Hall G, Wang L, Spurney RF. TRPC channels in proteinuric kidney diseases. *Cells.* 2019;9(1):44. Available from: <https://doi.org/10.3390/cells9010044>.
25. Duan J, Li J, Zeng B. et al. Structure of the mouse TRPC4 ion channel. *Nat. Commun.* 2018;9:3102. Available from: <https://doi.org/10.1038/s41467-018-05247-9>.
26. Vinayagam D, Mager T, Apelbaum A. et al. Electron cryo-microscopy structure of the canonical TRPC4 ion channel. *eLife.* 2018;7:e36615. Available from: <https://doi.org/10.7554/eLife.36615>.
27. Vinayagam D, Quentin D, Yu-Strzelczyk J. et al. Structural basis of TRPC4 regulation by calmodulin and pharmacological agents. *eLife.* 2020;9:e60603. Available from: <https://doi.org/10.7554/eLife.60603>.

28. Miller M, Shi J, Zhu Y. et al. Identification of ML204, a novel potent antagonist that selectively modulates native TRPC4/C5 ion channels. *J. Biol. Chem.* 2011;286(38):33436–33446.
29. Rubaiy HN, Ludlow MJ, Bon RS. et al. Pico145 – powerful new tool for TRPC1/4/5 channels. *Channels.* 2017;11(5):362–364.
30. Lopez JJ, Jardin I, Sanchez-Collado J. et al. TRPC channels in the SOCE scenario. *Cells.* 2020;9(1):126. Available from: <https://doi.org/10.3390/cells9010126>.
31. Yu M, Ledeboer MW, Daniels M. et al. Discovery of a potent and selective TRPC5 inhibitor, efficacious in a focal segmental glomerulosclerosis model. *ACS Med. Chem. Lett.* 2019;10(11):1579–1585.
32. Song K, Wei M, Guo W. et al. Structural basis for human TRPC5 channel inhibition by two distinct inhibitors. *eLife.* 2021;10:e63429. Available from: <https://doi.org/10.7554/eLife.63429>.
33. Arabi AA. Artificial intelligence in drug design: algorithms, applications, challenges and ethics. *Future Med. Chem.* 2017;9(18):2167–2180.
34. Wu F, Zhou Y, Li L. et al. Computational approaches in preclinical studies on drug discovery and development. *Front. Chem.* 2020;8:726. Available from: <https://doi.org/10.3389/fchem.2020.00726>.
35. Chubanov V, Grimm C, Hill K. et al. Physiological functions and pharmacological targeting of transient receptor potential channels. *Pharmacol. Rev.* 2025;77(6):100089. Available from: <https://doi.org/10.1016/j.pharmr.2025.100089>.
36. Chen J, Luan Y, Yu R. et al. Transient receptor potential (TRP) channels, promising potential diagnostic and therapeutic tools for cancer. *Biosci. Trends.* 2014;8(1):1–10.
37. Dong XP, Shen D, Wang X. et al. PI(3,5)P2 controls membrane trafficking by direct activation of mucolipin  $\text{Ca}^{2+}$  release channels in the endolysosome. *Nat. Commun.* 2010;1:38. Available from: <https://doi.org/10.1038/ncomms1037>.



38. Kim H, Kim J, Jeon JP. et al. The roles of G proteins in the activation of TRPC4 and TRPC5 transient receptor potential channels. *Channels*. 2012;6(5):333–343.
39. Kang H, Kim J, Park CH. et al. Direct modulation of TRPC ion channels by G $\alpha$  proteins. *Front. Physiol.* 2024;15:1362987. Available from: <https://doi.org/10.3389/fphys.2024.1362987>.
40. Phelan KD, Mock MM, Kretz O. et al. Heteromeric canonical transient receptor potential 1 and 4 channels play a critical role in epileptiform burst firing and seizure-induced neurodegeneration. *Mol. Pharmacol.* 2012;81(3):384–392.
41. Dresviannikov AV, Bolton TB, Zholos AV. Muscarinic receptor-activated cationic channels in murine ileal myocytes. *Br. J. Pharmacol.* 2006;149(2):179–187.
42. Zholos AV, Bolton TB. Properties of the muscarinic cationic channel in longitudinal smooth muscle of guinea pig small intestine. In: Bolton TB, Tomita T, editors. *Smooth muscle excitation*. Academic Press; 1996. p. 155–173.
43. Zholos AV, Bolton TB. Muscarinic control of cationic channel opening in smooth muscle. *J. Physiol.* 1998;511:3S–4S.
44. Zholos AV, Tsvilovskyy VV, Bolton TB. Muscarinic cholinergic excitation of smooth muscle: signal transduction and single cationic channel properties. *Neurophysiology*. 2003;35:283–301.
45. Takahashi N, Kozai D, Mori Y. TRP channels: sensors and transducers of gasotransmitter signals. *Front. Physiol.* 2012;3:324. Available from: <https://doi.org/10.3389/fphys.2012.00324>.
46. Tang J, Lin Y, Zhang Z. et al. Identification of common binding sites for calmodulin and inositol 1,4,5-trisphosphate receptors on the carboxyl termini of TRP channels. *J. Biol. Chem.* 2001;276(24):21303–21310.

47. Otsuguro KI, Tang J, Tang Y. et al. Isoform-specific inhibition of TRPC4 channel by phosphatidylinositol 4,5-bisphosphate. *J. Biol. Chem.* 2008;283(15):10026–10036.
48. Duan J, Li J, Chen GL. et al. Cryo-EM structure of TRPC5 at 2.8-Å resolution reveals unique and conserved structural elements essential for channel function. *Sci. Adv.* 2019;5(7):eaaw7935. Available from: <https://doi.org/10.1126/sciadv.aaw7935>.
49. Schaefer M, Plant TD, Obukhov AG. et al. Receptor-mediated regulation of the nonselective cation channels TRPC4 and TRPC5. *J. Biol. Chem.* 2000;275(23):17517–17526.
50. Plant TD, Schaefer M. TRPC4 and TRPC5: receptor-operated  $\text{Ca}^{2+}$ -permeable nonselective cation channels. *Cell Calcium.* 2003;33(5–6):441–450.
51. Jeon J, Tian JB, Zhu MX. TRPC4 as a coincident detector of Gi/o and Gq/11 signaling: mechanisms and pathophysiological implications. *Curr. Opin. Physiol.* 2020;17:34–41.
52. Thakur DP, Tian JB, Jeon J. et al. Critical roles of Gi/o proteins and phospholipase C- $\delta$ 1 in the activation of receptor-operated TRPC4 channels. *Proc. Natl. Acad. Sci. U. S. A.* 2016;113(4):1092–1097.
53. Zholos AV, Tsytsyura YD, Gordienko DV. et al. Phospholipase C, but not InsP3 or DAG, -dependent activation of the muscarinic receptor-operated cation current in guinea-pig ileal smooth muscle cells. *Br. J. Pharmacol.* 2004;141(1):23–36.
54. Jeon JP, Hong C, Park EJ. et al. Selective  $\text{G}\alpha_i$  subunits as novel direct activators of TRPC4 and TRPC5 channels. *J. Biol. Chem.* 2012;287(21):17029–17039.
55. Mederos y Schnitzler M, Gudermann T, Storch U. Emerging roles of diacylglycerol-sensitive TRPC4/5 channels. *Cells.* 2018;7(11):218. Available from: <https://doi.org/10.3390/cells7110218>.

56. Schwarz Y, Oleinikov K, Schindeldecker B. et al. TRPC channels regulate  $\text{Ca}^{2+}$ -signaling and short-term plasticity of fast glutamatergic synapses. *PLoS Biol.* 2019;17(9):e3000445. Available from: <https://doi.org/10.1371/journal.pbio.3000445>.
57. Tanahashi Y, Komori S, Matsuyama H. et al. Functions of muscarinic receptor subtypes in gastrointestinal smooth muscle: a review using receptor-knockout mice. *Int. J. Mol. Sci.* 2021;22(2):926. Available from: <https://doi.org/10.3390/ijms22020926>.
58. Yan HD, Okamoto H, Unno T. et al. Effects of G-protein-specific antibodies and  $\text{G}\beta\gamma$  subunits on the muscarinic receptor-operated cation current in guinea-pig ileal smooth muscle cells. *Br. J. Pharmacol.* 2003;139(3):605–615.
59. Zholos AV, Bolton TB. G-protein control of voltage dependence as well as gating of muscarinic metabotropic channels in guinea-pig ileum. *J. Physiol.* 1994;478(2):195–202.
60. Zholos AV, Bolton TB, Dresvyannikov AV. et al. Cholinergic excitation of smooth muscles: multiple signaling pathways linking M2 and M3 muscarinic receptors to cationic channels. *Neurophysiology.* 2004;36:398–406.
61. Zholos AV, Zholos AA, Bolton TB. G-protein-gated TRP-like cationic channel activated by muscarinic receptors: effect of potential on single-channel gating. *J. Gen. Physiol.* 2004;123:581–598.
62. Bolton TB, Prestwich SA, Zholos AV. et al. Excitation-contraction coupling in gastrointestinal and other smooth muscles. *Annu. Rev. Physiol.* 1999;61:85–115.
63. Gordienko DV, Zholos AV. Modulation of muscarinic cation current by intracellular  $\text{Ca}^{2+}$  release: a central role of the inositol 1,4,5-trisphosphate receptors. *Neurophysiology.* 2003;35:320.
64. Gordienko DV, Zholos AV. Regulation of muscarinic cationic current in myocytes from guinea-pig ileum by intracellular  $\text{Ca}^{2+}$  release: a central role

- of inositol 1,4,5-trisphosphate receptors. *Cell Calcium*. 2004;36(5):367–386.
65. Tsvilovsky VV, Zholos AV, Bolton TB. Modulation of the muscarinic cation current by intracellular calcium ions in guinea-pig ileal myocytes. *Neurophysiology*. 2003;35:347.
  66. Gordienko DV, Zholos AV, Bolton TB. Membrane ion channels as physiological targets for local  $\text{Ca}^{2+}$  signalling. *J. Microsc.* 1999;196(3):305–316.
  67. Grimm S, Just S, Fuertig R. et al. TRPC4/5 inhibitors: phase I results and proof of concept studies. *Eur. Arch. Psychiatry Clin. Neurosci.* 2025;275:1599–1610.
  68. Englisch CN, Paulsen F, Tschernig T. TRPC channels in the physiology and pathophysiology of the renal tubular system: what do we know? *Int. J. Mol. Sci.* 2023;24(1):181. Available from: <https://doi.org/10.3390/ijms24010181>.
  69. Dryn DO, Melnyk MI, Al Kury LT. et al.  $\text{C}_{60}$  fullerenes disrupt cellular signalling leading to TRPC4 and TRPC6 channels opening by the activation of muscarinic receptors and G-proteins in small intestinal smooth muscles. *Cell. Signal.* 2018;43:40–46.
  70. Dryn DO, Melnyk MI, Zholos AV. The muscarinic cation current in murine ileal myocytes (mICAT) inhibited by ketamine can be recovered by TRPC4 agonist (-)-englerin A. In: *Proceedings of The Physiological Society Scientific Meeting "Physiology 2019"*; 2019; Aberdeen, UK. p. 243P–244P.
  71. Melnyk MI, Dryn DO, Al Kury LT. et al. Suppression of mICAT in mouse small intestinal myocytes by general anaesthetic ketamine and its recovery by TRPC4 agonist (-)-englerin A. *Front. Pharmacol.* 2020;11:594882. Available from: <https://doi.org/10.3389/fphar.2020.594882>.
  72. Ratnayake R, Covell D, Ransom TT. et al. Englerin A, a selective inhibitor of renal cancer cell growth, from *Phyllanthus engleri*. *Org. Lett.* 2009;11(1):57–60.

73. Alexander SPH, Mathie AA, Peters JA. et al. The Concise Guide to PHARMACOLOGY 2023/24: ion channels. *Br. J. Pharmacol.* 2023;180(Suppl. 2):S145–S222.
74. Minard A, Bauer C, Wright D. et al. Remarkable progress with small-molecule modulation of TRPC1/4/5 channels: implications for understanding the channels in health and disease. *Cells.* 2018;7(6):52. Available from: <https://doi.org/10.3390/cells7060052>.
75. Deore AB, Dhumane JR, Wagh HV. et al. The stages of drug discovery and development process. *Asian J. Pharm. Res. Dev.* 2019;7(6):62–67.
76. Shoichet BK. Virtual screening of chemical libraries. *Nature.* 2004;432(7019):862–865.
77. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. *BMC Bioinformatics.* 2009;10:168. Available from: <https://doi.org/10.1186/1471-2105-10-168>.
78. Liang J, Edelsbrunner H, Woodward C. Anatomy of protein pockets and cavities: measurement of binding site geometry and implications for ligand design. *Protein Sci.* 1998;7(9):1884–1897.
79. Tian W, Chen C, Lei X. et al. CASTp 3.0: computed atlas of surface topography of proteins. *Nucleic Acids Res.* 2018;46:W363–W367.
80. Liu Y, Yang X, Gan J. et al. CB-Dock2: improved protein–ligand blind docking by integrating cavity detection, docking and homologous template fitting. *Nucleic Acids Res.* 2022;50:W159–W164.
81. Ghersi D, Sanchez R. Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites. *Proteins.* 2009;74:417–424.
82. Jiménez J, Doerr S, Martínez-Rosell G. et al. DeepSite: protein-binding site predictor using 3D-convolutional neural networks. *Bioinformatics.* 2017;33:3036–3042.

83. Aggarwal R, Gupta A, Chelur V. et al. DeepPocket: ligand binding site detection and segmentation using 3D convolutional neural networks. *J. Chem. Inf. Model.* 2022;62:5069–5079.
84. Kandel J, Tayara H, Chong KT. PUPResNet: prediction of protein–ligand binding sites using deep residual neural network. *J. Cheminform.* 2021;13:65. Available from: <https://doi.org/10.1186/s13321-021-00547-7>.
85. Jeevan K, Palistha S, Tayara H. et al. PUPResNetV2.0: a deep learning model leveraging sparse representation for improved ligand binding site prediction. *J. Cheminform.* 2024;16:66. <https://doi.org/10.1186/s13321-024-00865-6>
86. An Y, Lim J, Glavatskikh M. et al. In silico fragment-based discovery of CIB1-directed anti-tumor agents by FRASE-Bot. *Nat. Commun.* 2024;15:5564. <https://doi.org/10.1038/s41467-024-49892-9>
87. Xia CQ, Pan X, Shen HB. Protein-ligand binding residue prediction enhancement through hybrid deep heterogeneous learning of sequence and structure data. *Bioinformatics.* 2020;36:3018–3027.
88. Zhao Y, He S, Xing Y. et al. A point cloud graph neural network for protein-ligand binding site prediction. *Int. J. Mol. Sci.* 2024;25:9280. Available from: <https://doi.org/10.3390/ijms25179280>.
89. Krivák R, Hoksza D. P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure. *J. Cheminform.* 2018;10:39. Available from: <https://doi.org/10.1186/s13321-018-0285-8>.
90. Guvench O, MacKerell AD. Comparison of protein force fields for molecular dynamics simulations. *Methods Mol. Biol.* 2008;443:63–88.
91. Huang J, Rauscher S, Nawrocki G. et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods.* 2017;14(1):71–73.

92. Wu EL, Cheng X, Jo S. et al. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J. Comput. Chem.* 2014;35(27):1997–2004.
93. Abraham MJ, Murtola T, Schulz R. et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 2015;1–2:19–25.
94. Yesylevskyy S. MolAR: a memory-safe library for analysis of molecular dynamics simulations written in Rust. *J. Comput. Chem.* 2025;46:e27536. Available from: <https://doi.org/10.1002/jcc.27536>.
95. Moriwaki H, Tian YS, Kawashita N. et al. Mordred: a molecular descriptor calculator. *J. Cheminform.* 2018;10:4. Available from: <https://doi.org/10.1186/s13321-018-0258-y>.
96. Rogers D, Hahn M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* 2010;50:742–754.
97. Turon G, Hlozek J, Woodland JG. et al. First fully-automated AI/ML virtual screening cascade implemented at a drug discovery centre in Africa. *Nat. Commun.* 2023;14:5736. Available from: <https://doi.org/10.1038/s41467-023-41512-2>.
98. Pudjihartono N, Fadason T, Kempa-Liehr AW. et al. A review of feature selection methods for machine learning-based disease risk prediction. *Front. Bioinform.* 2022;2:927312. Available from: <https://doi.org/10.3389/fbinf.2022.927312>.
99. Jaeger S, Fulle S, Turk S. Mol2vec: unsupervised machine learning approach with chemical intuition. *J. Chem. Inf. Model.* 2018;58(1):27–35.
100. Huang K, Fu T, Gao W. et al. Therapeutics Data Commons: machine learning datasets and tasks for drug discovery and development. *arXiv [Preprint]*. 2021. Available from: <https://doi.org/10.48550/arXiv.2102.09548>.
101. Huang K, Fu T, Gao W. et al. Artificial intelligence foundation for therapeutic science. *Nat. Chem. Biol.* 2022;18(10):1033–1036.

102. Morris GM, Huey R, Lindstrom W. et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 2009;30(16):2785–2791.
103. Morris GM, Goodsell DS, Halliday RS. et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* 1998;19(14):1639–1662.
104. Jo S, Kim T, Iyer VG. et al. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* 2008;29(11):1859–1865.
105. Jumper J, Evans R, Pritzel A. et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583–589.
106. Schrödinger, LLC. The PyMOL Molecular Graphics System. Version 2.x [software]. Available from: <https://pymol.org/>.
107. Європейська конвенція про захист хребетних тварин, що використовуються для дослідних та інших наукових цілей: ETS № 123. Страсбург; 1986.
108. Про захист тварин від жорстокого поводження: Закон України від 21.02.2006 № 3447-IV. Відомості Верховної Ради України. 2006;(27):Ст. 230.
109. Zhang C, Zhang X, Freddolino L. et al. BioLiP2: an updated structure database for biologically relevant ligand–protein interactions. *Nucleic Acids Res.* 2024;52:D404–D412.
110. Wei H, Wang W, Peng Z. et al. Q-BioLiP: a comprehensive resource for quaternary structure-based protein–ligand interactions. *Genomics Proteomics Bioinformatics.* 2024;22:qzae001. Available from: <https://doi.org/10.1093/gpbjnl/qzae001>.
111. GROBID – GeneRation Of Bibliographic Data [Internet]. GitHub. Available from: <https://github.com/kermitt2/grobid>.
112. Wei J, Wang X, Schuurmans D. et al. Chain-of-thought prompting elicits reasoning in large language models. *arXiv [Preprint]*. 2023. Available from: <https://doi.org/10.48550/arXiv.2201.11903>.



113. Yao S, Yu D, Zhao J. et al. Tree of thoughts: deliberate problem solving with large language models. arXiv [Preprint]. 2023. Available from: <https://arxiv.org/abs/2305.10601>.
114. Kim JJ, Gharpure A, Teng J. et al. Shared structural mechanisms of general anaesthetics and benzodiazepines. *Nature*. 2020;585:303–308.
115. Murphy JM, Lucet IS, Hildebrand JM. et al. Insights into the evolution of divergent nucleotide-binding mechanisms among pseudokinases revealed by crystal structures of human and mouse MLKL. *Biochem. J.* 2014;457:369–377.
116. Kruse AC, Kobilka BK, Gautam D. et al. Muscarinic acetylcholine receptors: novel opportunities for drug development. *Nat. Rev. Drug Discov.* 2014;13:549–560.
117. Ahuja S, Mukund S, Deng L. et al. Structural basis of Nav1.7 inhibition by an isoform-selective small-molecule antagonist. *Science*. 2015;350:aac5464. Available from: <https://doi.org/10.1126/science.aac5464>.
118. Hofmann T, Obukhov AG, Schaefer M. et al. Direct activation of human TRPC6 and TRPC3 channels by diacylglycerol. *Nature*. 1999;397(6716):259–263.
119. Bai Y, Yu X, Chen H. et al. Structural basis for pharmacological modulation of the TRPC6 channel. *eLife*. 2020;9:e53311. Available from: <https://doi.org/10.7554/eLife.53311>.
120. Стратійчук РО, Дринь ДО, Жолос ОВ. Блокування TRPC4-каналів інгібітором Pico145: експериментальне та молекулярне дослідження. Вісник Київського національного університету імені Тараса Шевченка. Біологія. 2026;1(104):25–31.
121. Стратійчук РО, Дринь ДО, Жолос ОВ. Структурне моделювання взаємодії GFB-887 із TRPC4 та оцінка його впливу на карбахол-індуковану скоротливість тонкої кишки миші. Фармакологія та лікарська токсикологія. 2026;20(1):45–58.

122. Irwin JJ, Tang KG, Young J. et al. ZINC20 – a free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Model.* 2020;60(12):6065–6073.
123. Landrum G, Tosco P, Kelley B. et al. rdkit/rdkit: 2022\_09\_1 (Q3 2022) release [Internet]. 2022. Available from: <https://doi.org/10.5281/zenodo.7235579>.
124. Akiba T, Sano S, Yanase T. et al. Optuna: a next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery; 2019. p. 2623–2631. Available from: <https://doi.org/10.1145/3292500.3330701>.
125. Wang NN, Dong J, Deng YH. et al. ADME properties evaluation in drug discovery: prediction of Caco-2 cell permeability using a combination of NSGA-II and boosting. *J. Chem. Inf. Model.* 2016;56(4):763–773.
126. Hou T, Wang J, Zhang W. et al. ADME evaluation in drug discovery. 7. Prediction of oral absorption by correlation and classification. *J. Chem. Inf. Model.* 2007;47:208–218.
127. Broccatelli F, Carosati E, Neri A. et al. A novel approach for predicting P-glycoprotein (ABCB1) inhibition using molecular interaction fields. *J. Med. Chem.* 2011;54:1740–1751.
128. Ma CY, Yang SY, Zhang H. et al. Prediction models of human plasma protein binding rate and oral bioavailability derived by using GA–CG–SVM method. *J. Pharm. Biomed. Anal.* 2008;47(4–5):677–682.
129. Martins IF, Teixeira AL, Pinheiro L. et al. A Bayesian approach to in silico blood-brain barrier penetration modeling. *J. Chem. Inf. Model.* 2012;52:1686–1697.
130. Experimental in vitro DMPK and physicochemical data on a set of publicly disclosed compounds (ChEMBL3301361) [Internet]. Available from: <https://www.ebi.ac.uk/explore/document/ChEMBL3301361>.

131. Lombardo F, Jing Y. In silico prediction of volume of distribution in humans. Extensive data set and the exploration of linear and nonlinear methods coupled with molecular interaction fields descriptors. *J. Chem. Inf. Model.* 2016;56(10):2042–2052.
132. Veith H, Southall N, Huang R. et al. Comprehensive characterization of cytochrome P450 isozyme selectivity across chemical libraries. *Nat. Biotechnol.* 2009;27:1050–1055.
133. Xu C, Cheng F, Chen L. et al. In silico prediction of chemical Ames mutagenicity. *J. Chem. Inf. Model.* 2012;52:2840–2847.
134. Xu Y, Dai Z, Chen F. et al. Deep learning for drug-induced liver injury. *J. Chem. Inf. Model.* 2015;55(10):2085–2093.
135. Karim A, Lee M, Balle T. et al. CardioTox net: a robust predictor for hERG channel blockade based on deep learning meta-feature ensembles. *J. Cheminform.* 2021;13:60. Available from: <https://doi.org/10.1186/s13321-021-00541-z>.
136. Huang K, Fu T, Glass LM. et al. DeepPurpose: a deep learning library for drug-target interaction prediction. *Bioinformatics.* 2021;36:5545–5547.
137. KatanaGraph. SimGCN-TDC [Internet]. GitHub. Available from: <https://github.com/KatanaGraph/SimGCN-TDC>.
138. Just S, Chenard BL, Ceci A. et al. Treatment with HC-070, a potent inhibitor of TRPC4 and TRPC5, leads to anxiolytic and antidepressant effects in mice. *PLoS One.* 2018;13(1):e0191225. Available from: <https://doi.org/10.1371/journal.pone.0191225>.
139. Bon RS, Wright DJ, Beech DJ. et al. Pharmacology of TRPC channels and its potential in cardiovascular and metabolic medicine. *Annu. Rev. Pharmacol. Toxicol.* 2022;62:427–446.
140. ClinicalTrials.gov. A first-in-human, phase 1, placebo-controlled study to evaluate the safety, tolerability, and pharmacokinetics of GFB-887. Identifier: NCT03970122. Available from: <https://clinicaltrials.gov/study/NCT03970122>.

141. Westerling-Bui AD, Fast EM, Soare TW. et al. Transplanted organoids empower human preclinical assessment of drug candidate for the clinic. *Sci. Adv.* 2022;8(27):eabj5633. Available from: <https://doi.org/10.1126/sciadv.abj5633>.

**Перелік опублікованих праць та відомості про апробацію досліджень****Праці, в яких опубліковані основні наукові результати дисертації:**

1. **Stratiichuk R.**, Melnychenko M., Koleiev I., Voitsitskyi T., Husak V., Shevchuk N., Ostrovsky Z., Bdzhola V., Yesylevsky S., Starosyla S., Nafiiiev A. Leveraging large language models for literature-driven prioritization of protein binding pockets. *Bioinformatics*. 2025. V. 41 (8). P. btaf449. *(особистий внесок здобувача – обробка даних (формування набору білків і пов'язаних публікацій для бенчмарка), формальний аналіз результатів, валідація).*
2. **Стратійчук Р. О.**, Дринь Д. О., Жолос О. В. Блокування TRPC4-каналів інгібітором Pico145: експериментальне та молекулярне дослідження. *Вісник Київського національного університету імені Тараса Шевченка. Біологія*. 2026. № 1 (104). С. 25–31. *(особистий внесок здобувача – концептуалізація дослідження, методологія експериментів із молекулярного моделювання, проведення молекулярного моделювання, аналіз даних та написання чернетки рукопису).*
3. **Стратійчук Р. О.**, Дринь Д. О., Жолос О. В. Структурне моделювання взаємодії GFB-887 із TRPC4 та оцінка його впливу на карбахол-індуковану скоротливість тонкої кишки миші. *Фармакологія та лікарська токсикологія*. 2026. Т. 20, № 1. С. 45–58. *(особистий внесок здобувача – інформаційний пошук, концептуалізація дослідження, методологія експериментів із молекулярного моделювання, проведення молекулярного моделювання, аналіз даних, графічна підготовка результатів до опублікування та написання чернетки рукопису).*

### Опубліковані праці апробаційного характеру

1. **Stratiichuk R.**, Zholos A. ADMET prediction using molecular embeddings based on Mol2Vec: a comparative study within the TDC. XXII Міжнародна наукова конференція студентів та молодих вчених. Київ, 23–25 квітня 2025.
2. **Stratiichuk R.**, Zholos A. Large language model-driven identification of biologically relevant protein binding pockets to accelerate drug discovery. IX Міжнародна наукова конференція «Психофізіологічні та вісцеральні функції в нормі і патології». Київ, 21–24 жовтня 2025.
3. **Stratiichuk R. O.**, Zholos O. V., Dryn D. O. Analysis of mouse TRPC4 inhibition by Pico145 and GFB-887 using molecular modeling, electrophysiology, and functional physiology methods. Український біофізичний тиждень 2026. Україна, 25–28 березня 2026.