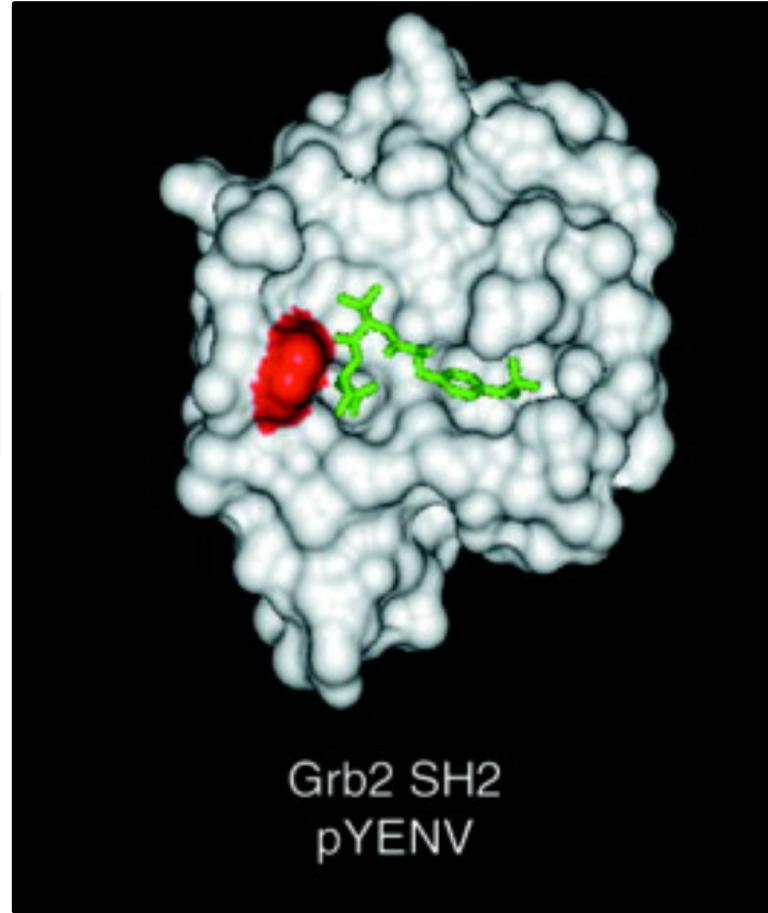
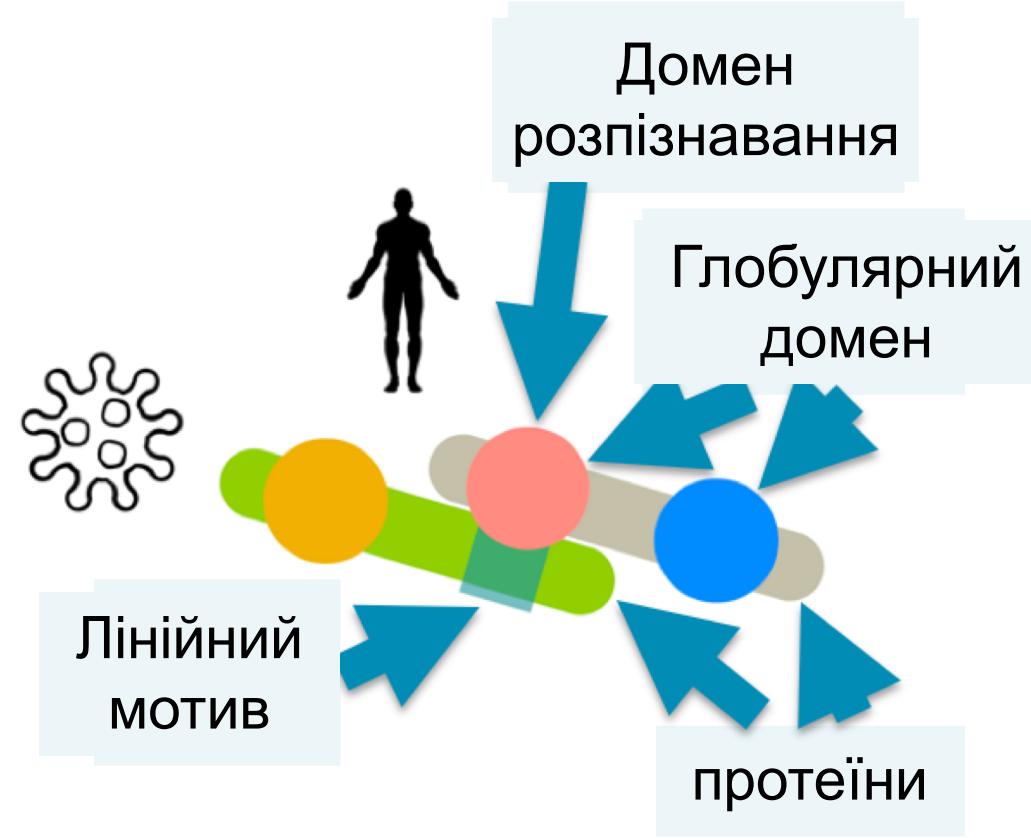


# Пошук функціональних лінійних мотивів протеїнів з використанням сукупності білкових взаємодій

Випускна кваліфікаційна робота  
студента ІІ курсу магістратури  
денної форми навчання  
**Клещевнікова Віталія Віталійовича**  
Науковий керівник від кафедри  
к.б.н., доцент Гребінік Д.М.

Робота виконана в Європейській молекулярно-біологічній лабораторії –  
Європейському інституту біоінформатики (EMBL-EBI), Хінкстон, Кембридж,  
Велика Британія, під керівництвом керівника групи Доктора Євангелії  
Петсалакі



- Швидко еволюціонують і тому широко використовуються вірусами
- 15-40% очікується проте відомо лише про 1%

**Рис. 1. Властивості коротких лінійних мотивів**

Van Roey K. [et al.]. // Chem. Rev. – 2014. – Vol. 114, № 13 – P. 6733–6778.

Pawson T. [et al.]. // Trends Cell Biol. – 2001. – Vol. 11, № 12 – P. 504–511.

# **Мета роботи**

Пошук нових функціональних лінійних мотивів у мережі  
білкових взаємодій людини

## **Задачі роботи**

1. Отримати та обробити дані експериментальної взаємодії з публічних баз даних та вивчити властивості мережі вірусно-людської білкової взаємодії.
2. Визначити домени білкової послідовності у всіх вірусних і людських білках. Оцінити домени людини, ймовірно, опосередковують взаємодію з кожним вірусним білком.
3. Застосувати інструменти імовірного пошуку мотивів до мережі взаємодій білків людини та вірусів, щоб визначити короткі лінійні мотиви *de novo*. Використати послідовності вірусних білків, щоб підвищити чутливість пошуку.
4. Оцінити наш метод пошуку мотивів за допомогою еталонного набору даних відомих вірусних мотивів.
5. Реалізувати цей метод пошуку мотиву в статистичній мові програмування R, за допомогою інструментів командного рядка і високопродуктивного обчислювального кластера

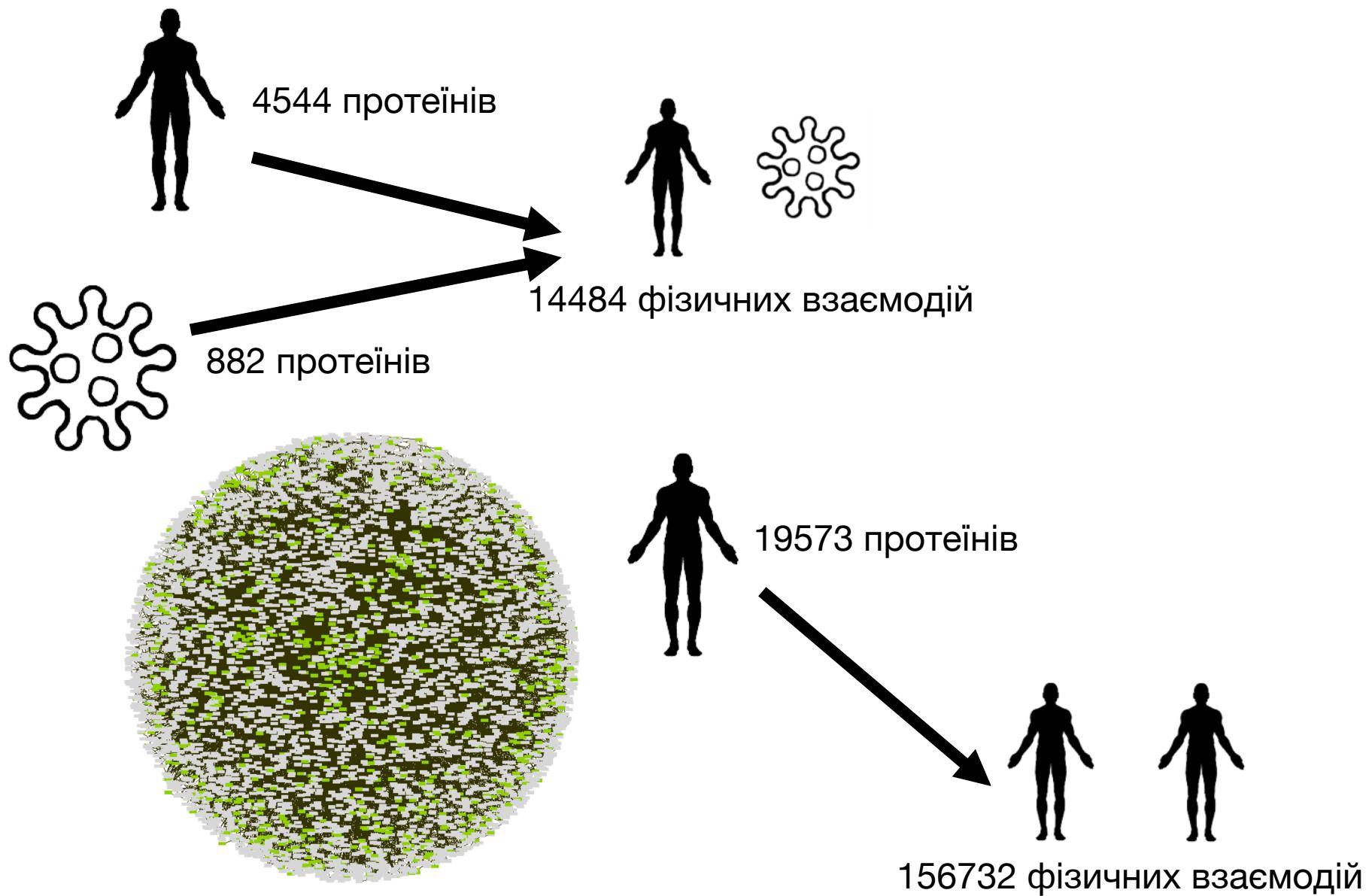


Рис. 2. Схема отриманих даних про взаємодію білків

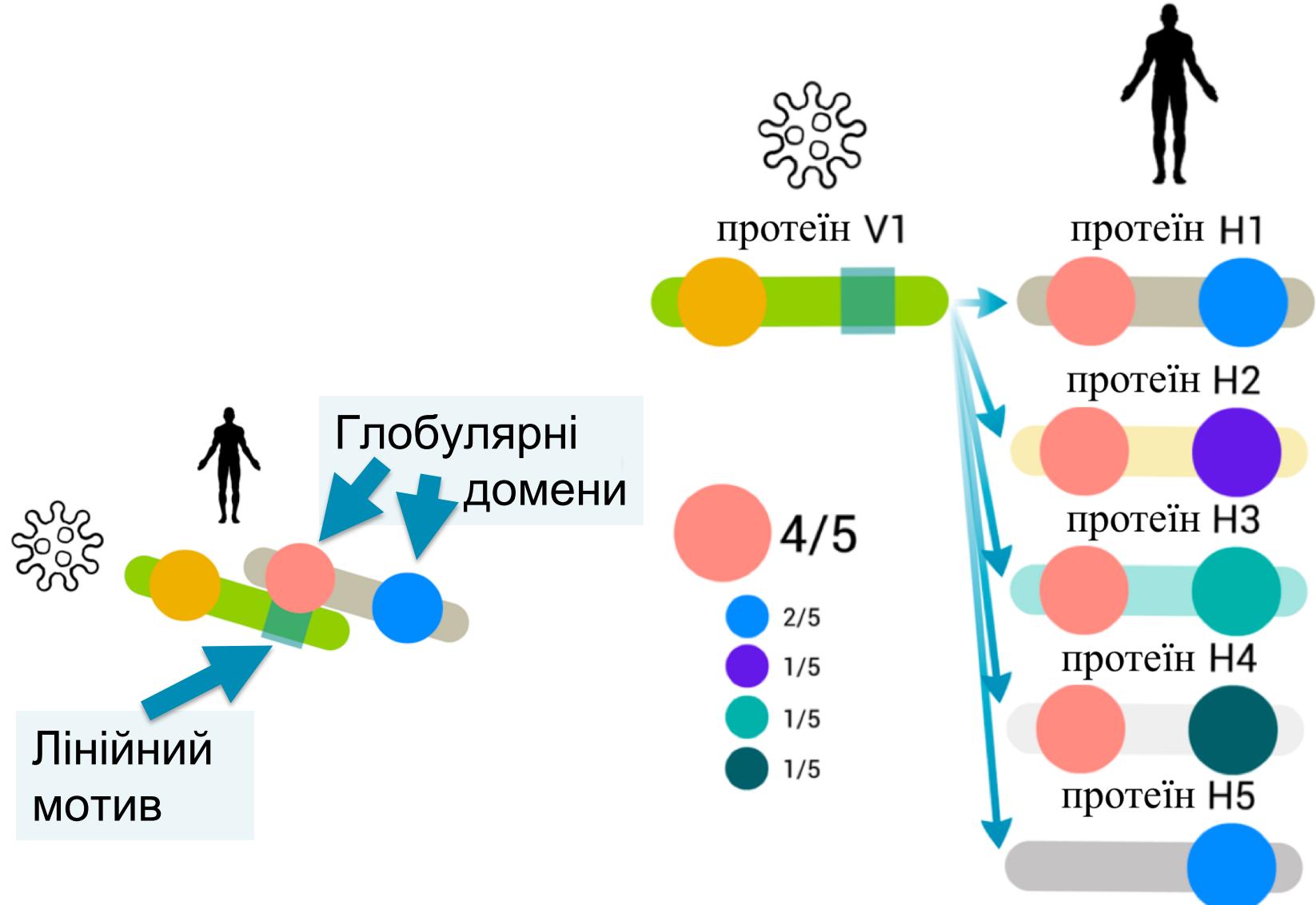
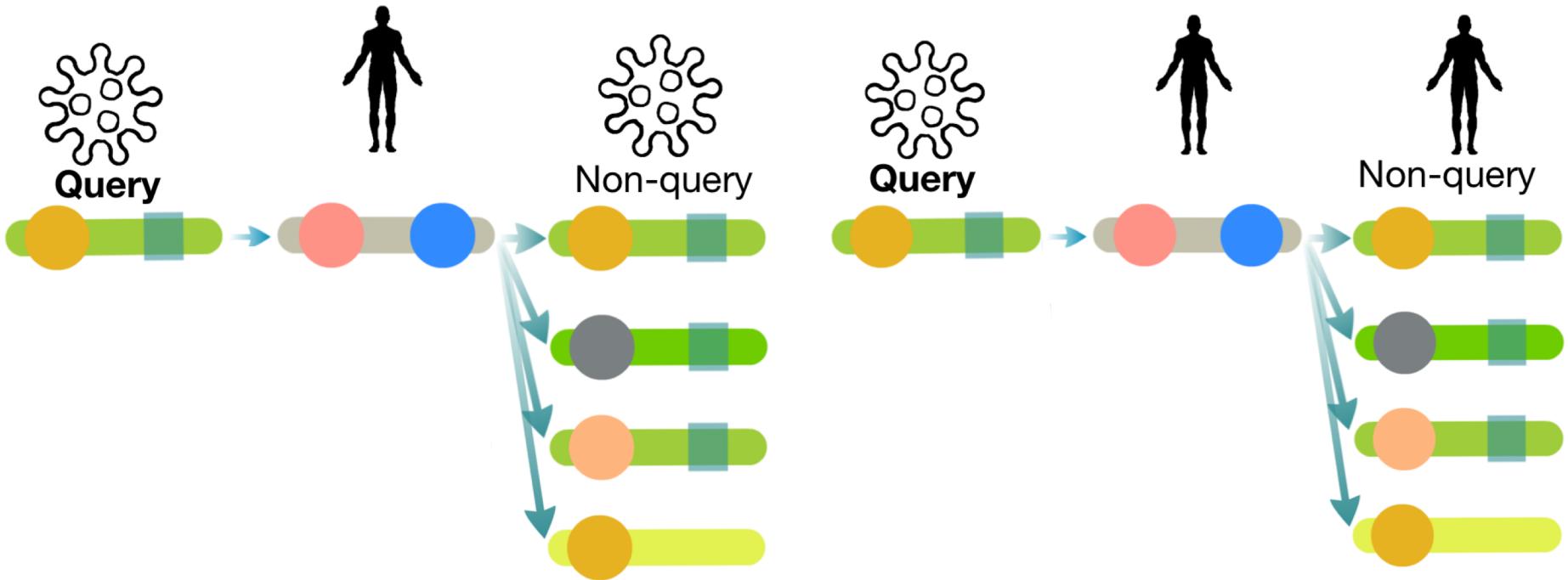


Рис. 3. Схема методу визначення доменів, що ймовірно опосередковують взаємодію



Мотиви мають бути присутніми у протеїнах запиту (query) :

- **Підвищена чутливість**
- **Підвищена специфічність** до прямих взаємодій

Ми тестували **8 варіантів** процедури

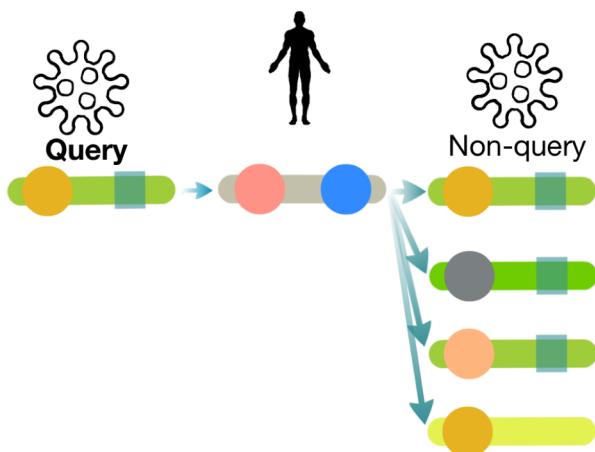
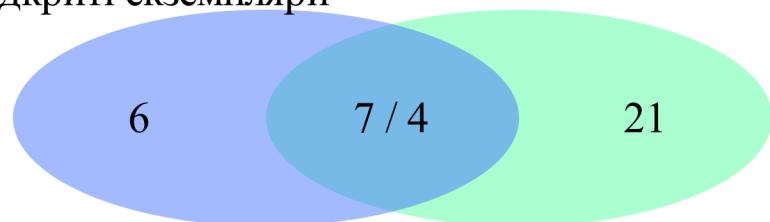
**Рис. 4. Схема методу пошуку мотивів**

## взаємодії між білками людини та вірусів

суворий поріг: значення  $p < 0.04$

16 білки з відомими мотивами (4 відкрито заново)  
159 мотиви передбачені в інших білках

відкриті екземпляри



## взаємодії білків людини (дані IntAct)

суворий поріг: значення  $p < 0.002$

16 білки з відомими мотивами (3 відкрито заново)  
149 мотиви передбачені в інших білках

відкриті екземпляри

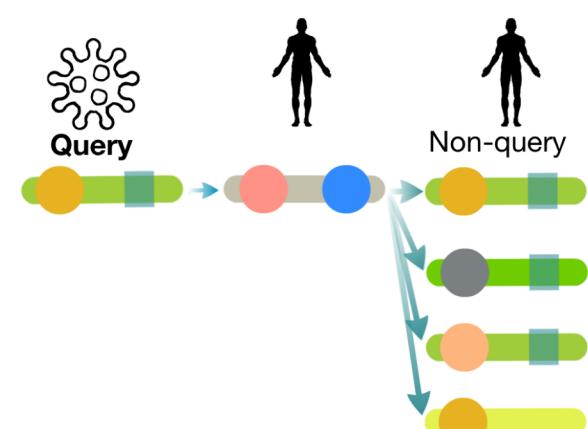
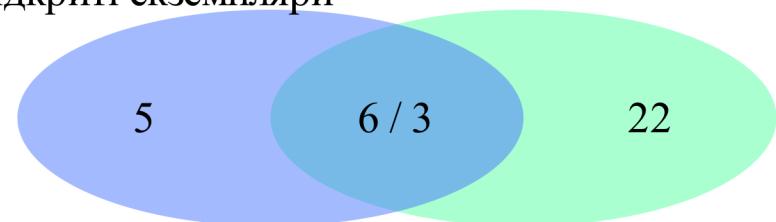
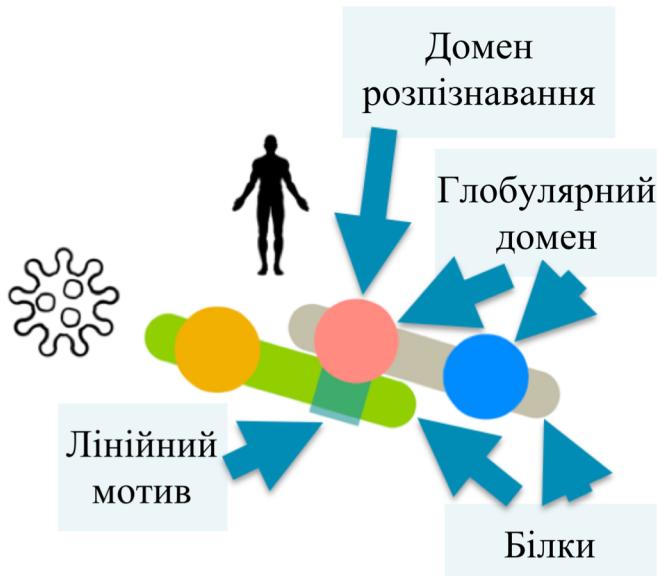


Рис. 5. Діаграми Вена, що показують кількість виявлених мотивів-кандидатів та заново визначених відомих мотивів.

# Підсумок передбачених мотивів та доменів, що їх розпізнають



Класичні ліганд-зв'язуючі мотиви:

- 4 мотиви **WD40**
- 1 мотив **SH3**
- 1 мотив **EF-hand**
- 1 мотив РН-подібного домену

Найбільш представлені:

- 26 варіантів **мотивів PDZ**
- 12 варіантів **сигналів ядерної локалізації**

Інші мотиви  
(за назвою домену):

- 1 циклін-подібний домен
- 1 кератинова головка типу 2
- 2 Gro-EL-подібний домен
- 7 BAG домен
- 1 домен, що зв'язує дволанцюгову РНК

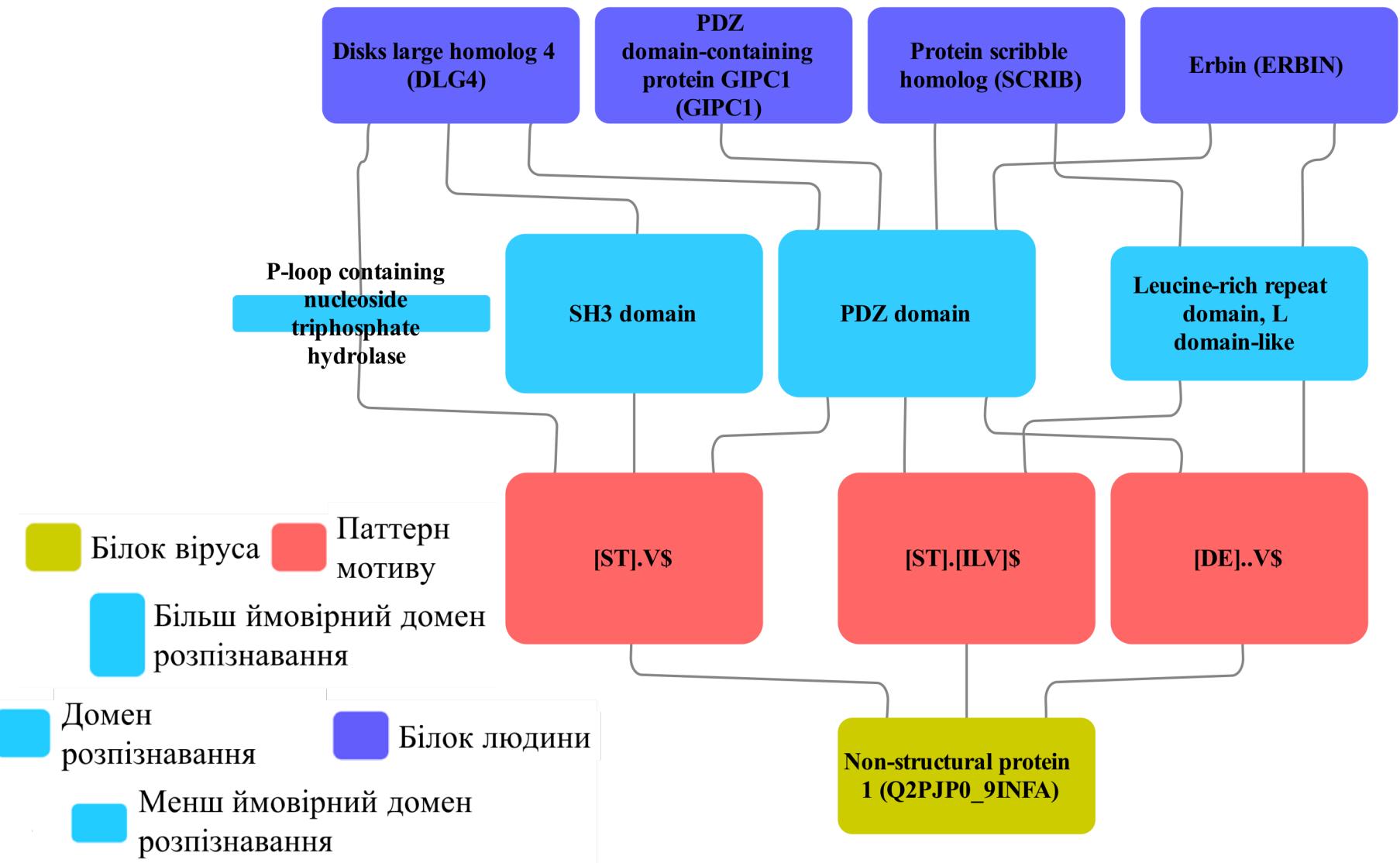


Рис. 6. Схема мережі мотиву-кандидату PDZ в неструктурному білку 1 (NS1) вірусу H5N1 грипу були підтвердженні в попередньому дослідженні, але не входив до еталонних даних

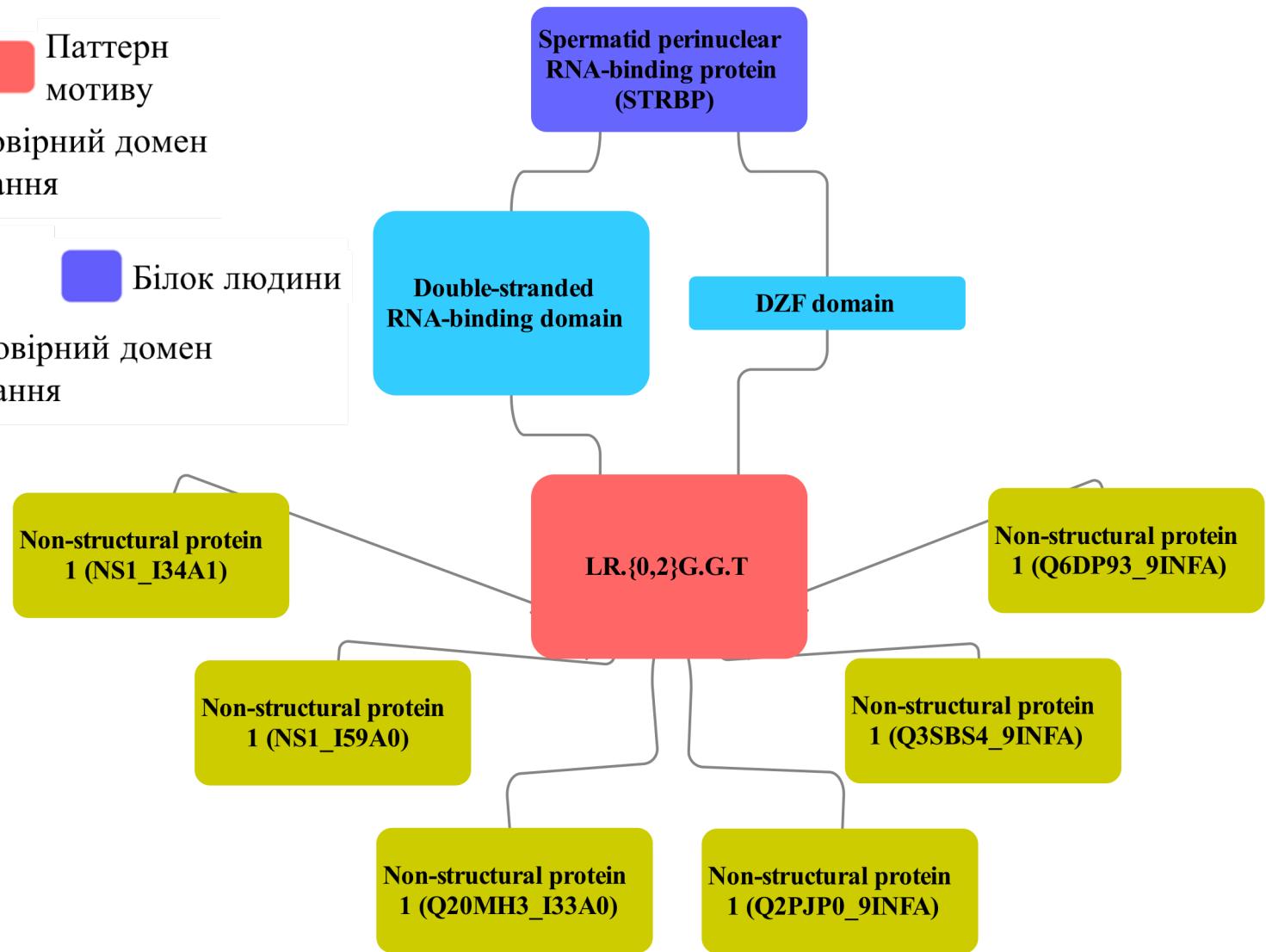
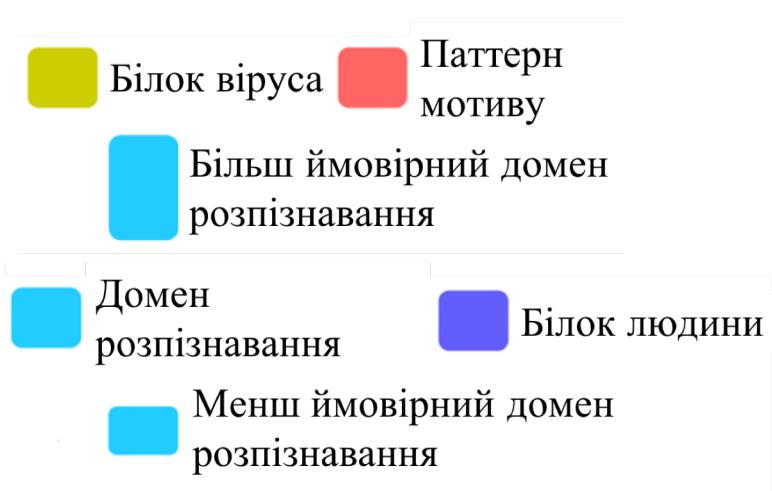


Рис. 7. Схема мережі мотиву-кандидату LR.{0,2}G.G.T передбачений в 6 неструктурних білках різних штамів грипу А.

# ВИСНОВКИ

1. Ми передбачили екземпляри мотивів та домени що їх розпізнають використовуючи дані білкових взаємодій та статичні моделі. Використання вірусних білків та передбачення доменів підвищило чутливість передбачення мотивів.
2. Ми отримали та обробили дані експериментальної взаємодії з публічних баз даних. Ми вивчили властивості мережі вірусно-людської взаємодії. Білки людини - мішенні вірусів виступають як центральні, але є результатом дослідницької упередженості у даних про білкові взаємодії.
3. Ми заново визначили відомі приклади коротких лінійних мотивів у вірусних білках і передбачили нові мотиви-кандидати за допомогою ймовірнісних інструментів пошуку мотивів, мережі взаємодій білків людини і вірусів та використовуючи послідовності вірусних білків, щоб підвищити чутливість пошуку.
4. Ми визначили домени у всіх людських білках. Ми оцінили, які домени людини, ймовірно, опосередковують взаємодію з кожним вірусним білком. Ці домени є збагаченими на відомі домени розпізнавання мотивів. Фільтрація можливих доменів покращує відкликання. Інтеграція передбачення домену та мотиву спрощує інтерпретацію результатів.

# ВИСНОВКИ

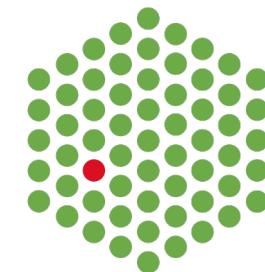
5. За жорсткого порогу з точністю 50% ми визначили de novo 3 відомі екземпляри мотивів з нашого тренувального набору, та визначили 6 екземпляри відомих мотивів, які не були в нашему тренувальному наборі. Ми передбачили 43 екземпляри нових мотивів кандидатів.
6. Ми розробили цю процедуру пошуку мотивів статистичною мовою програмування R, використовуючи інструменти командного рядка та обчислювальний кластер. Ця процедура може бути використана для передбачення мотивів, коли будуть отримані нові дані про взаємодію білків.
7. Ця робота сприяє нашему розумінню коду взаємодії лінійного мотиву - домену розпізнавання та спрямовує вибір доменів людини-вірусних мішеней для подальших експериментальних досліджень невпорядкованого вірусного протеому за допомогою фагового дисплею.

Подяка:

доктору Евангелії Петсалакі  
та її групі



Європейському інституту  
біоінформатики **EMBL-EBI**



Кафедрі біохімії  
Гребінику Д.М., к.б.н., доценту

**ДЯКУЮ ЗА УВАГУ !**

**ДЯКУЮ ЗА УВАГУ !**