



БЛЮМ Ярослав Борисович — академік НАН України, доктор біологічних наук, професор, директор ДУ «Інститут харчової біотехнології та геноміки НАН України»



КАРПОВ Павло Андрійович — кандидат біологічних наук, завідувач лабораторії біоінформатики та структурної біології відділу геноміки та молекулярної біотехнології ДУ «Інститут харчової біотехнології та геноміки НАН України»

ЕФЕКТИВНІСТЬ ВИКОРИСТАННЯ ПОТЕНЦІАЛУ ГРІД ДЛЯ ВИСОКОПРОПУСКНОГО ВІРТУАЛЬНОГО СКРИНІНГУ БІОЛОГІЧНО АКТИВНИХ РЕЧОВИН

У статті наведено короткий історичний огляд розвитку ґрід-обчислень, розглянуто їх актуальний стан в Україні. Особливу увагу приділено біологічному і медико-біологічному сегменту в національній ґрід-інфраструктурі України. Надано стислу інформацію про діяльність та спеціалізацію офіційно зареєстрованих віртуальних організацій медико-біологічного напрямку. На прикладі віртуальної організації CSLabGrid проаналізовано переваги від впровадження ґрід-обчислень у наукові дослідження та їх економічний потенціал, наведено результати розрахунків для проектів, спрямованих на поєднання ґрід-обчислень і завдань високопропускового віртуального скринінгу (high-throughput screening) біологічно активних речовин — нових агентів з протибактеріальною та антимітоотичною активністю.

Ключові слова: ґрід, УНГ, CSLabGrid, економічний ефект, високопропусковий віртуальний скринінг, біологічно активні речовини.

Ідея раціонального комп'ютингу з використанням потужностей комп'ютерів під час їх простою зародилася наприкінці 1990-х років і набула популярності завдяки волонтерським обчислювальним проектам (volunteer computing): GIMPS (1996), distributed.net (1997) і SETI@home (1999) [1]. Однак уже в 1994 р. Єн Фостер, Карл Кессельман і Стів Тікі сформулювали основні ідеї ґрід-комп'ютингу, зокрема базові принципи розподілених обчислень, об'єктно-орієнтованого програмування, раціонального використання обчислювальних ресурсів, веб-сервісів тощо [2, 3]. Саме тоді сформувалося поняття «ґриду», або «ґрід-інфраструктури» як розподіленого програмно-апаратного комп'ютерного середовища з розподіленою організацією високопродуктивних обчислень, здатного керувати робочим навантаженням і потоками даних [4]. Як показав час, розроблені базові принципи виявилися життєздатними, і ґрід-обчислення вже міцно ввійшли в такі галузі науки, як ядерна фізика, інженерне моделювання, моніторинг стану на-

вколишнього середовища, прогнозування погоди, матеріалознавство, фармацевтика, біомедичне моделювання, біоінформатика та ін. [5]. Сьогодні грід-технології в різних країнах світу активно застосовують не лише наукові спільноти, а й державні органи управління, обори, організації сфери комунальних послуг, приватні компанії, найчастіше фінансового та енергетичного спрямування.

В Україні офіційне впровадження грід-технологій розпочалося з 23 вересня 2009 р., коли було прийнято Постанову Кабінету Міністрів України «Про затвердження Державної цільової науково-технічної програми впровадження і застосування грід-технологій на 2009–2013 роки» [6]. Одночасно з цим було прийнято Положення про Український національний грід (УНГ) [7]. За цією програмою було створено перший грід-кластер, що об'єднав комп'ютерні мережі та обчислювальні ресурси Інституту теоретичної фізики НАН України, Головної астрономічної обсерваторії НАН України, кластер Інституту харчової біотехнології та геноміки (ІХБГ) НАН України (на той час кластер Інституту клітинної біології та генетичної інженерії НАН України) та кластер Інституту молекулярної біології і генетики (ІМБГ) НАН України [8–10]. На початок 2016 р. УНГ уже склався як єдина дослідницька е-інфраструктура національного рівня, яка об'єднує 39 ресурсних центрів наукових організацій України (з них 29 належать НАН України). Основою УНГ є 12 ресурсних центрів, що координуються національним операційним центром (NGI-UA) і зараз являють собою інтегровану в європейський простір (EGI – <http://www.egi.eu/>) грід-інфраструктуру. Останній факт є особливо знаменним для УНГ. Саме з моменту приєднання до EGI національний грід-проект перейшов у якісно новий формат і став частиною загальноєвропейського грід-простору, сформувавши національний домен NGI-UA [6, 8].

Від початку заснування Українського національного гріду значна частина грід-обчислень припадає на вирішення молекулярно-біологічних завдань. Загалом в УНГ зареєстровано

14 віртуальних організацій¹, шість з яких використовують грід для розв'язання обчислювальних задач у галузі біології та медицини.

1. MolDynGrid² – віртуальна лабораторія, створена на базі кластера ІМБГ НАН України (IMBG Cluster) для вирішення завдань структурної біології. Насамперед вона орієнтована на ресурсомісткі обчислення молекулярної динаміки білків і макромолекулярних комплексів.

2. SysBio³ – створена лабораторією системної біології ІМБГ НАН України і орієнтована на дослідження регуляції генів еукаріотів. Основна спеціалізація – реконструкція мереж генної регуляції з використанням обчислень у гріді.

3. CompuChemGridUA⁴ – створена відділом молекулярної біофізики Фізико-технічного інституту низьких температур ім. Б.І. Веркіна (Харків) і орієнтована на використання гріду для неемпіричних квантово-механічних розрахунків фізичних властивостей фрагментів металізованої ДНК.

4. Multiscale⁵ – створена в Інституті фізики конденсованих систем НАН України (Львів) на базі обчислювального кластера ICMP і спрямована на використання грід-обчислень для розв'язання фундаментальних і науково-прикладних задач фізики, хімії, матеріалознавства, а також біології. Зараз у її межах виконуються різноманітні розрахунки, починаючи з квантово-хімічних і закінчуючи мезоскопічним рівнем моделювання.

5. Medgrid⁶ – створена для розв'язання обчислювальних задач у галузі медичних досліджень. Ця віртуальна організація передусім спрямована на виконання популяційних досліджень у галузі кардіології, ґрунтуючись на постійно оновлюваній базі даних електрокардіограм, має прикладний характер і в остаточ-

¹ http://infrastructure.kiev.ua/monitoring/rubrik/5/sub_rubrik/0/?p=0.

² <https://moldyngrid.org/>.

³ <http://sysbio.org.ua>.

⁴ <http://infrastructure.kiev.ua/monitoring/54/>.

⁵ <http://www.icmp.lviv.ua/multiscale/>.

⁶ <http://medgrid.immsp.kiev.ua/>.

ному вигляді оперативно надає інформацію, необхідну лікарям, а також фахівцям, що працюють у сфері охорони здоров'я і страхової медицини.

6. CSLabGrid. У 2011 р. на базі кластера ДУ «Інститут харчової біотехнології та геноміки НАН України» – IFBG Cluster⁷ було створено віртуальну організацію CSLabGrid⁸ (CytoSkeleton Laboratory for Grid computations) [10, 11]. Вона була призначена для розв'язання комплексу обчислювальних задач, пов'язаних з дослідженням молекулярної організації цитоскелета, механізмів клітинного поділу, а також з використанням основних білків цитоскелета і білків, асоційованих з ним (БАМ, протеїнкінази, протеїнфосфатази та ін.), як молекулярних мішеней для біологічно активних сполук. Стратегічно обчислювальні задачі, що розв'язуються CSLabGrid, включають віртуальний високопропускний молекулярний скринінг (virtual high-throughput screening, vHTS), молекулярний докінг, обчислення молекулярної динаміки білків, 3D-QSAR, задачі класичної біоінформатики, завдання, пов'язані зі зберіганням та обробкою даних лазерної конфокальної та електронної мікроскопії [11, 12].

На перших етапах діяльності CSLabGrid найбільшу увагу приділяли розвитку інфраструктури, відпрацюванню методів та їх поєднанню з ґридом. Було створено відповідні віртуальні бібліотеки 3D-структур білків-мішеней (зокрема, тубулінів і FtsZ-білків з патогенних мікроорганізмів), а також бібліотеки лігандів. Так, для зберігання та повторного використання було створено базу даних структурних моделей цитоскелетних білків – CSMoDB⁹. Віртуальну бібліотеку структур низькомолекулярних агентів, яку застосовують для молекулярного докінгу і високопропускного скринінгу, було створено, спираючись на можливості бібліотеки AKos¹⁰, а також бібліотеки сполук, синтезованих співробітниками Інституту органічної хімії НАН України, доступної для

⁷ <http://ifbg.org.ua/uk/grid-vuzol-institutu>.

⁸ <http://ifbg.org.ua/uk/csllabgrid>.

⁹ <http://csmoedb.ifbg.org.ua/>.

¹⁰ www.akosgmbh.de/.

подальших експериментальних досліджень у рамках співпраці інститутів [11]. Нині загальний обсяг бібліотеки структур низькомолекулярних агентів становить 7 003 000 сполук.

В останні роки основна діяльність Українського національного ґриду безпосередньо пов'язана з цільовою комплексною програмою наукових досліджень НАН України «Грид-інфраструктура і ґрид-технології для наукових і науково-прикладних застосувань» на 2014–2018 рр., затвердженою постановою Президії НАН України від 11.12.2013 № 164. Ця академічна програма стала правонаступницею згаданої раніше Державної цільової науково-технічної програми, але орієнтована на практичне використання попередніх напрацювань УНГ. У зв'язку з головними завданнями програми наша увага також спрямована на практичне використання напрацювань ВО CSLabGrid, насамперед у напрямі високопропускного віртуального скринінгу. Як приклад, за підтримки Державної цільової науково-технічної програми впровадження і застосування ґрид-технологій ми виконали високопропускний молекулярний скринінг біологічно активних речовин з антимиіотичною активністю. Для цього було використано вищезгадану бібліотеку низькомолекулярних лігандів (7 003 000 сполук) і здійснено їх молекулярний докінг на поверхні молекул білка тубуліну (основний білок мііотичного веретена поділу в клітинах еукаріотів) та його гомолога – FtsZ-білка, який є головним білком клітинного поділу у мііоорганізмів, у тому числі й патогенних бактерій. Як мііені для докінгу ми використовували сайти зв'язування гетероциклічних сполук на поверхні молекул тубуліну і FtsZ-білка. Загальна кількість білків-мііеней з різних патогенних організмів становила 120. Отже, з урахуванням кількості мііеней необхідно було виконати 3 361 440 000 операцій лііганд-біілкового докінгу.

На реалізацію цього завдання із застосуванням ґрид-ресурсу УНГ та алгоритму UCSF DOCK загалом знадобилося приблизно 3 місяці, включно з докінгом, обчисленням молекулярної динаміки та відбором сполук-лідерів

з подальшою експериментальною перевіркою їх активності в лабораторних умовах (усього по 10–12 сполук для кожної білкової мішені). Якби подібні дослідження виконували з використанням локальної робочої станції на основі процесора Intel Core2 Quad Q6600, лише розрахунки молекулярного докінгу тривали б приблизно 389 днів для кожного білка. Загалом, за умови актуальної конфігурації кластера IFBG Cluster (грід-кластер ІХБГ НАН України), один цикл віртуального скринінгу з використанням 7 003 000 речовин для одного типу білка-мішені займає приблизно 3 місяці і коштує 1448 дол. США з урахуванням експериментальної перевірки активності сполук-лідерів.

Якби всі 7 003 000 сполук перевіряв би експериментально один дослідник, то вартість відбору сполук з антимітотичною активніс-

тю становила б близько 195 млн дол. США, а сполук з антибактеріальною активністю — приблизно 125 млн дол. США. Для реалізації проекту з пошуку як інгібіторів тубуліну, так і FtsZ-білків знадобилося б 87 500 робочих днів, або близько 240 років.

І це лише окремих приклад ефективності віртуального скринінгу з використанням грід-інфраструктури. Однак він дуже яскраво і наочно демонструє економічний потенціал застосованого віртуальною організацією CSLabGrid підходу. Використання гріду дозволило значно скоротити фінансові та часові затрати на пошук нових агентів з протибактеріальною та антимітотичною дією, що відкриває широкі перспективи для створення нових препаратів з антипротозойною, антигельмінтною, протипухлинною, фунгіцидною та гербіцидною активністю.

REFERENCES

[СПИСОК ЛІТЕРАТУРИ]

1. Cusack C., Martens C., Mutreja P. Volunteer Computing Using Casual Games. In: *Future of Game Design and Technology: Proc. Int. Conf. FuturePlay'2006*. (October 9–11, 2006, London, Ontario, Canada). P. 1–8.
2. Foster I., Kesselman C., Tuecke S. The Open Grid Services Architecture (Chapter 17). In: *The Grid 2*. (Elsevier Inc., 2004). P. 215–257.
3. Foster I., Kesselman C. *The Grid: blueprint for a new computing infrastructure*. (Elsevier, 1998).
4. Joseph J., Fellenstein C. *Grid computing*. (IBM Press, 2004).
5. Talbi E.-G., Zomaya A.Y. *Grid Computing for Bioinformatics and Computational Biology*. (John Wiley & Sons, 2008).
6. Zagorodny A.G., Svistunov S.Ya., Belous L.F., Golovinskiy A.L. UA-Grid: Ukrainian National Grid Program. In: *Parallel and Distributed Computing Systems: Proc. Int. Conf. PDCS'2013* (13–14 March, 2013, Kharkiv, Ukraine). [in Russian].
[Загородний А.Г., Свистунов С.Я., Белоус Л.Ф., Головинский А.Л. UA-Grid: Украинская национальная грид-программа. В кн.: *Parallel and Distributed Computing Systems: матер. наук. конф.* (13–14 березня 2013, Харків). С. 346–356].
7. Ukrainian National Grid Policy. <http://infrastructure.kiev.ua/en/about/statute/>.
[Положення про Український національний грід. http://infrastructure.kiev.ua/upload/ung_fin.pdf].
8. Zagorodny A.G., Svistunov S.Ya., Belous L.F., Golovinskiy A.L. UA-Grid: Ukrainian National Grid Program and its achievements. In: *Scientific service on the internet: all facets of parallelism: Proc. Int. Conf.* (23–28 September 2013, Novorossiysk, Russia). P. 453–457. [in Russian].
[Загородний А.Г., Свистунов С.Я., Белоус Л.Ф., Головинский А.Л. Украинская национальная программа UA-GRID и ее достижения. В кн.: *Научный сервис в сети Интернет: все грани параллелизма: труды Междунар. конф.* (23–28 сентября 2013, Новороссийск). С. 453–457].
9. Martynov E., Zinovjev G., Svistunov S. Academic segment of Ukrainian Grid infrastructure. *System research and information technologies*. 2009. (3): 31. [in Ukrainian].
[Мартинов Е., Зінов'єв Г. Академічний сегмент української грід-інфраструктури. *Системні дослідження та інформаційні технології*. 2009. № 3. С. 31–42].
10. Svistunov S.Ya., Shevchenko A.Yu. Status and prospects of development of the Ukrainian national grid. Analysis and logic of possible development. *System Research and Information Technologies*. 2014. (2): 40. [in Russian].

- [Свистунов С.Я., Шевченко А.Ю. Состояние и перспективы развития Украинского национального грид. Анализ и логика возможного развития. *Системні дослідження та інформаційні технології*. 2014. № 2. С. 40–52].
11. Karpov P.A., Brytsun V.M., Rayevsky A.V., Demchuk O.M., Pydiura N.O., Ozheredov S.P., Samofalova D.A., Spivak S.I., Yemets A.I., Kalchenko V.I., Blume Ya.B. High-Throughput Screening of New Antimitotic Compounds Based on CSLabGrid Virtual Organization. *Sci. Innov.* 2015. **11**(1): 85. [in Ukrainian].
[Карпов П.А., Брицун В.М., Раевський О.В. та ін. Високопропускний скринінг речовин з антимітотичною активністю на базі віртуальної організації CSLabGrid. *Наука та інновації*. 2015. Т. 11, № 1. С. 85–93].
12. Pydiura N., Karpov P., Blume Ya. Hardware environment for CSLabGrid: Reaching maximum efficacy of computations in structural biology and bioinformatics. In: *Cluster Computing: Proc. II Int. Conf. CC-2013* (June 3–5, 2013, Lviv).

Стаття надійшла 13.05.2016.

P.A. Karpov, Ya.B. Blume

Institute of Food Biotechnology and Genomics of National Academy of Sciences of Ukraine (Kyiv)

EFFICIENCY AND PERSPECTIVES OF HIGH-THROUGHPUT SCREENING IN GRID – CSLABGRID EXPERIENCE IN EFFECTIVE SEARCHING FOR NEW BIOLOGICALLY ACTIVE COMPOUNDS

The article provides a brief historical review of grid computing and its current state in Ukraine. Particular attention is paid to biological and biomedical segments of Ukrainian National Grid. A brief information on officially registered medical and biological virtual organizations (VO) and their specialization is presented. Based on experience of VO CSLabGrid, the main economic benefits from grid-computing implementation and their economical potential are analyzed. Economic impact is revealed for two CSLabGrid projects unifying grid-computing and protocols of virtual High-Throughput Screening for new antibacterial and antimitotic agents.

Keywords: grid, Ukrainian National Grid, CSLabGrid, economic benefit, high-throughput screening, biologically active compounds.