

МЕТОДИ СТРУКТУРНОЇ БІОІНФОРМАТИКИ

Г. П. ВОЛИНЕЦЬ, В. Г. БДЖОЛА, С. М. ЯРМОЛЮК

Інститут молекулярної біології і генетики НАН України, Київ

E-mail: galina.volinetc@gmail.com

В огляді наведено найпоширеніші методи комп'ютерного моделювання в біології. Описано технології порівняння амінокислотної послідовності та структурного вирівнювання, що їх використовують для пошуку функціонально подібних або еволюційно споріднених протеїнів. Розглянуто програми докінгу, застосовувані для вивчення формування надмолекулярних комплексів, які є невід'ємним знаряддям раціонального дизайну ліків. Комбінування методів комп'ютерного моделювання з експериментальними підходами дозволяє значно знизити вартість дослідження.

Ключові слова: вирівнювання амінокислотних послідовностей, структурне вирівнювання протеїнів, функціональний сайт, молекулярний докінг.

Надзвичайно стрімке зростання кількості експериментальних даних, що стосуються структури протеїнів, разом зі значним прогресом у розвитку методів комп'ютерного моделювання дозволяють наблизити сучасну науку до розуміння сутності живих систем на молекулярному рівні.

Сьогодні важливою характеристикою біологічних досліджень є комбінування методів комп'ютерного моделювання із загальноприйнятими експериментальними підходами. Результати комп'ютерного моделювання обов'язково мають супроводжуватися підтвердженням дослідженнями *in vitro* та *in vivo*. З іншого боку, науковий проект з великим обсягом біологічних досліджень потребує значної кількості ресурсів. Вартість такого дослідження може бути істотно знижена завдяки використанню комп'ютерних методів аналізу.

Методи порівняння амінокислотної послідовності та структурного вирівнювання дозволяють знаходити функціонально подібні або еволюційно споріднені протеїни. Гомологічне моделювання проводять з метою передбачити невідому просторову структуру протеїнів на основі подібності їх амінокислотної послідовності до відомих структур. Аналіз поверхні протеїнів та інтерфейсу взаємодії здійснюють для відкриття зв'язку «структура — властивість» взаємодіючих молекул. Пошук сайтів зв'язування протеїнів-мішеней є важливим для дизайну лі-

ків. Міжмолекулярні взаємодії маленьких лігандів, нуклеїнових кислот і протеїнів прогнозують за процедурою докінгу. Докінг малих протеїнових фрагментів дає змогу також змодельовати фолдинг протеїну.

В огляді розглянуто сучасні досягнення в галузі комп'ютерного моделювання та можливості найбільш поширених методологій для дослідження структурних мотивів, вивчення молекулярних асоціацій, дизайну ліків та ін.

Вирівнювання амінокислотної послідовності протеїнів

Комп'ютерні методи вирівнювання послідовностей можна поділити на дві категорії: глобальне та локальне вирівнювання. Глобальне вирівнювання є формою глобальної оптимізації, що знаходить відповідності амінокислотних залишків по всій довжині досліджуваних послідовностей. Локальне вирівнювання ідентифікує схожі ділянки у межах довгих послідовностей, які істотно відрізняються на більшій частині своєї протяжності.

Для глобального вирівнювання застосовують алгоритм Нудельмана–Вунша [1], а для локального — Сміта–Вотермана [2]. Стосовно надзвичайно подібних послідовностей між локальним та глобальним вирівнюванням немає різниці.

Залежно від кількості послідовностей, що порівнюються, розрізняють попарне та

множинне вирівнювання. Існує три види методів для реалізації попарного вирівнювання — точково-матричний, динамічне програмування і так звані словесні методи.

Точково-матричний підхід використовують для вирівнювання дуже схожих послідовностей. Для побудови графіка одну з послідовностей записують уздовж верхнього рядка зліва направо, а іншу — вздовж крайньої лівої колонки зверху вниз. Коли амінокислотний залишок послідовності, що розміщений у верхньому рядку, відповідає тому, який міститься в колонці, то в місці перетину ставлять крапку. У результаті виходить двовимірний матриця. Точкові графіки близькоспоріднених послідовностей мають вигляд лінії вздовж діагоналі матриці. Із точково-матричного графіка дуже просто візуально визначити деякі характерні риси послідовності, зокрема інсерції, делеції, повтори чи інвертовані повтори [3].

Метод динамічного програмування дозволяє проводити як локальне, так і глобальне вирівнювання. Цей підхід включає три етапи: ініціалізації, заповнення матриці (скоринг) та вирівнювання. На першій стадії створюється матриця з кількістю колонок і рядків, що на 1 більше, ніж кількість мономерів у вирівнюваних послідовностях (якщо M — кількість мономерів у колонці, а N — кількість мономерів у рядку, то по вертикалі й горизонталі буде відповідно $M + 1$ та $N + 1$ клітинок). Першу колонку і перший рядок матриці заповнюють нулями. Далі заповнення розпочинають з лівого верхнього кутка матриці і для кожної позиції $M_{i,j}$ знаходять максимальне значення, що визначається таким чином: $M_{i,j} = \text{MAXIMUM}[M_{i-1,j-1} + 1$ (якщо є відповідність між мономерами рядка і колонки), або «0» (така відповідність відсутня); $M_{i,j-1}$ чи $M_{i-1,j}$]. Етап вирівнювання починається з позиції, що має найбільше числове значення. Визначають три сусідні клітинки — зліва, зверху і по діагоналі, серед них вибирають можливий попередник. Якщо всі три однакові, то попередником є значення, що розташоване по діагоналі. Якщо за такого вирівнювання по горизонталі поряд є декілька однакових чисел, то це означає, що в послідовності, яка розміщена вертикально, у цьому місці є пропуски. І навпаки, якщо по вертикалі поряд розміщені однакові значення, то це означає, що пробіли є в послідовності, яка розміщена горизонтально зверху. Вирівнювання закінчується в лівому верхньому кутку матриці. Метод динамічного програмування детально описано на сайті: <http://www.avatar.se/molbioinfo2001/dynprog/dynamic.html>.

Найуживанішими методами вирівнювання є так звані словесні методи. Вони визначають серію коротких амінокислотних послідовностей («слів»), що не перекриваються в структурі, яка порівнюватиметься з базою даних послідовностей. Ці методи найчастіше використовують у пошукових інструментах баз даних FASTA і BLAST [4]. У методі FASTA користувач встановлює значення k , що визначає довжину слова, за яким буде проводитися пошук. За малих значень k метод працює повільно, але результати більш точні. BLAST було розроблено як альтернативу для FASTA, що працює значно швидше і майже не поступається за точністю. Подібно до FASTA, BLAST здійснює пошук слова довжиною k , але оцінює лише найбільш важливі відповідності слів, а не кожне слово, як це робить FASTA. Вирівнювання за допомогою цих методів може бути виконано на веб-порталах: <http://www.ebi.ac.uk/Tools/fasta33/index.html> та <http://www.ncbi.nlm.nih.gov/BLAST/>.

Множинне вирівнювання послідовностей застосовують для одночасного порівняння більш ніж двох послідовностей. Цей вид вирівнювання є значно інформативнішим, ніж попарний, оскільки консервативність низки молекул є значно важливішою, ніж повторюваність амінокислотних залишків у двох протеїнів. Консервативні мотиви можна використовувати для визначення каталітично активних сайтів ензимів та еволюційної спорідненості протеїнів.

До методів множинного вирівнювання належать динамічне програмування, прогресивні методи, ітеративні методи та профільний аналіз для знаходження мотивів.

Динамічне програмування потребує конструювання n -вимірного еквівалента матриці, що була описана вище для двох послідовностей, де n — число послідовностей. Для реалізації цього методу потрібні значні комп'ютерні затрати. З метою зменшення комп'ютерних потреб було створено пакет програм MSA, який ґрунтується на «сумі пар» цільової функції [5].

Прогресивні методи — це так звані ієрархічні методи, які спочатку порівнюють найбільш схожі послідовності і поступово додають менш споріднені. Спочатку подібність визначають за допомогою попарного вирівнювання, наприклад за використання методу FASTA. Результати прогресивного вирівнювання залежать від вибору «найбільш споріднених» послідовностей на початковому етапі попарного вирівнювання. Прогресивні методи вирівнювання використовують

для конструювання філогенетичних дерев та передбачення структури протеїнів. Програми, що ґрунтуються на прогресивних методах вирівнювання, — Clustal [6] і T-Coffee [7]. Існують сайти, де можна провести вирівнювання за допомогою цих підходів: <http://www.ebi.ac.uk/Tools/clustalw/> та <http://www.ebi.ac.uk/Tools/t-coffee/index.html>.

Ітеративними методами спочатку здійснюють глобальне вирівнювання груп послідовностей, а потім проводять низку повторних вирівнювань для підмножин послідовностей. У процесі повторного вирівнювання ці алгоритми здатні виправляти помилки [8]. Ці методи можна застосовувати навіть до послідовностей із дуже низьким ступенем подібності.

Профільний аналіз здійснюється в процесі глобального множинного вирівнювання послідовностей, в результаті якого у досліджуваному ряді послідовностей знаходять короткі консервативні мотиви амінокислотних залишків. Такі консервативні регіони ізолюють і використовують для конструювання сету профільних матриць.

У профільній матриці значення частоти для кожної амінокислоти отримують з її розподілу безпосередньо в консервативних регіонах, а не із загального емпіричного розподілу [9].

Для множинного вирівнювання послідовностей також застосовують алгоритм загальної оптимізації — приховану модель Маркова. Його покладено в основу програми FSA (Fast Statistical Alignment). Ця програма спочатку здійснює попарне порівнювання послідовностей [10], а потім використовує технологію відбору послідовностей для реалізації множинного вирівнювання [11]. Цей підхід вирівнювання значно точніший порівняно з вищеописаними завдяки використанню алгоритму «крутого підйому». Програма FSA разом з інструментом візуалізації може бути використана через веб-інтерфейс на сайті: <http://orangutan.math.berkeley.edu/fsa/>, а ресурс кодів доступний через Інтернет-сторінку: <http://fsa.sourceforge.net/>.

Структурне вирівнювання протеїнів

Оскільки структура протеїнів є більш консервативною, ніж послідовність, структурне вирівнювання дозволяє точніше визначити функціональну та еволюційну спорідненість протеїнів. Окрім того, структурне вирівнювання є досить надійним способом для передбачення функції протеїнів. Для

структурного вирівнювання потрібна інформація щодо вторинної і третинної структури протеїнів. Отже, очевидним є те, що ці методи можуть бути застосовані лише до послідовностей з експериментально встановленою просторовою структурою. Результати структурного вирівнювання можуть також слугувати основою для гомологічного моделювання [12] та структурної класифікації [13].

Методи структурного вирівнювання можна застосовувати для порівняння двох або більшої кількості послідовностей і, як правило, для здійснення локальних вирівнювань.

Метод DALI використовують з метою попарного структурного вирівнювання протеїнових структур. За допомогою просторових координат кожного протеїну обчислюють матриці відстаней між C_{α} -атомами амінокислотних залишків. Матриці відстаней спочатку розкладають на прості контактні мотиви, наприклад на субматриці гексапептид-гексапептид. Потім подібні контактні мотиви у двох матрицях розташовують попарно і комбінують у більш послідовні ряди пар.

Оптимізацію показника подібності, виходячи з еквівалентних внутрішньомолекулярних відстаней, здійснюють методом Монте Карло. Деякі вирівнювання оптимізуються паралельно, що забезпечує одночасне визначення найкращого, гіршого та інших розв'язків. Метод повністю автоматичний і досить чітко ідентифікує структурну схожість і спільні структурні мотиви [14].

Метод DALI був використаний для конструювання бази даних структурного вирівнювання FSSP (Families of Structurally Similar Proteins). Існує веб-сервер DALI http://ekhidna.biocenter.helsinki.fi/dali_server/, де можна провести порівняння структури досліджуваного протеїну з наявними у PDB (Protein Data Bank).

За методом SSAP (Sequential Structure Alignment Program) з метою реалізації структурного вирівнювання двічі використовують динамічне програмування. Замість вирівнювання атомів C_{α} за допомогою методу SSAP будують вектори до цих атомів від β -карбонів. Це здійснюється для всіх амінокислотних залишків, окрім гліцину. Таким чином, SSAP враховує ротамерний стан кожного залишку разом із його положенням уздовж ланцюга. Потім SSAP конструює вектори відстаней від цих амінокислотних залишків до їхніх найближчих сусідніх залишків на кожному протеїні. Далі будуються матриці, що містять різниці векторів для кожної пари залишків. Динамічне

програмування застосовують до кожної отриманої матриці для визначення серій оптимальних локальних вирівнювань, що потім об'єднуються в одну матрицю, до якої знову застосовують динамічне програмування з метою здійснення загального структурного вирівнювання. SSAP було створено для парного структурного вирівнювання, однак згодом розширено для проведення множинного вирівнювання [15]. Цей метод використовували для конструювання ієрархічної бази даних протеїнових укладок CATH (Class, Architecture, Topology, Homology) [16].

Метод комбінаторного подовження дозволяє проводити попарне і множинне вирівнювання протеїнових структур. Ґрунтуючись на таких параметрах, як відстань між амінокислотними залишками, локальна вторинна структура і характеристики сусідніх амінокислотних залишків, зокрема гідрофобність, здійснюють локальне вирівнювання, в результаті якого отримують «пари вирівняних фрагментів». Комбінації «пар вирівняних фрагментів», які репрезентують можливі способи вирівнювання, додаються або видаляються, що зрештою зумовлює єдине оптимальне вирівнювання. Алгоритм, що є досить швидким і точним для знаходження оптимального структурного вирівнювання, використовують для сканування бази даних та детального аналізу великих родин протеїнів [17]. Веб-сервер, де застосовують цей метод, розміщено на сайті: [http:// web.archive.org/web/20080610070608/cl.sdsc.edu/](http://web.archive.org/web/20080610070608/cl.sdsc.edu/).

Деякі методи множинного вирівнювання дають змогу здійснювати глобальне попарне порівняння вхідних структур. Прикладом таких методів є PrISM [18]. Більш точним є одночасне розглядання всіх молекул. Таку схему використовують у методах MUSTA [19], MultiProt [20] і MASS [21]. За цими технологіями можна визначати структурно консервативні мотиви незалежно від послідовності.

Для розв'язання задачі множинного структурного вирівнювання можна також застосовувати оптимізацію Monte Carlo [22].

У більшості методів для структурного вирівнювання протеїни приймають за жорсткі тіла. Проте конформаційна рухливість протеїнів має надзвичайно важливе значення. Розглядають рухи двох типів — швидкі переміщення бічних ланцюгів та повільні великомасштабні згинання у шарнірній ділянці. Ефективна детекція шарнірних ділянок є важливою для ідентифікації сайтів зв'язування та дизайну інгібіторів чи ліків.

Використовуючи алгоритм гнучкого структурного вирівнювання, можна вирішити три основні завдання. По-перше, здійснити вирівнювання амінокислотних залишків, по-друге, — суперпозицію відповідних регіонів і, по-третє, — ідентифікацію шарнірної ділянки. Метод FlexMol дає змогу вирівнювати протеїни, розглядаючи їх як об'єкти, що складаються із жорстких частин, з'єднаних обертальними зв'язками [23]. Для ідентифікації шарнірних ділянок було розроблено алгоритм FlexProt, що одночасно вирівнює жорсткі частини молекул [24].

Більшість методів аналізують структурне вирівнювання лише на одному рівні: на рівні атомів, елементів вторинної структури чи більших структурних фрагментів. MolCom забезпечує багаторівневий аналіз множинно вирівняних структур протеїнів. Кожен рівень має відповідну скорингову функцію, що дозволяє кількісно визначати вирівнювання протеїнів на різних структурних рівнях [25]. Такий підхід є важливим для структурного аналізу в зв'язку з ієрархічною природою структури протеїнів.

Розпізнавання функціональних сайтів

Наявність подібних просторових структур протеїнів часто свідчить про схожі функції, але є випадки, коли протеїни з однаковою укладкою, наприклад ТІМ-бочка, виконують різну біохімічну роль [26]. З іншого боку, протеїни з різними укладками можуть мати однакові функції. Найбільш вірогідним є припущення, що такі зв'язувальні мотиви визначають схожі функції протеїнів.

Ідентифікація подібних функціональних сайтів може застосовуватися для дизайну ліків, зокрема для виявлення протеїнів, на які може впливати досліджувана сполука і таким чином призводити до побічних ефектів. Вважають, що найпростіший шлях вирішити цю проблему — проаналізувати структурно різні протеїни, закрystalізовані у присутності одного й того самого ліганду. Суперпозиція цих комплексів, що дозволяє вирівняти їхні ліганди, забезпечує також вирівнювання сайтів зв'язування [27]. Це є важливим для дослідження функціональних груп, які беруть участь у взаємодії. Проте головний недолік такого підходу полягає в тому, що один і той самий ліганд може альтернативно зв'язуватися з активним сайтом протеїну [28]. Це зумовлює необхідність дослідження фізико-хімічних властивостей амінокислот.

Для порівняння сайтів зв'язування протеїнів існують два основні підходи. Перший дозволяє розпізнавати специфічні просторові мотиви амінокислот. Більшість із цих методів було розроблено для ідентифікації амінокислотних залишків, що формують «каталітичні тріади». Прикладом такого підходу є алгоритм ASSAM [29]. Методи для розпізнавання подібних сайтів зв'язування без урахування послідовності амінокислотних залишків або укладки були отримані Brakoulias et al. [30] і Shulman-Peleg et al. [31]. Додаткові методи для пошуку функціональних сайтів у структурах протеїнів зазначено в огляді Jones et al. [32].

Молекулярний докінг

Докінг застосовують для дослідження структури надмолекулярних комплексів. За допомогою алгоритму докінгу можна передбачити взаємодію молекул, визначити просторову структуру їх комплексів та афінність взаємодії. Існує декілька взаємозалежних конформацій молекул, що є сприятливими для взаємодії. Це так звані способи зв'язування.

Алгоритми докінгу низькомолекулярних сполук до молекул рецептора є невід'ємним знаряддям для раціонального дизайну ліків. Докінг може застосовуватися на різних стадіях процесу розроблення ліків. Найчастіше його використовують для віртуального скринінгу бібліотек низькомолекулярних органічних сполук та фрагментів. Для прогнозування токсичності потенційні ліки можна тестувати *in silico* на метаболічно активних протеїнах. Таким чином, комп'ютерне фільтрування дозволяє зберегти не лише багато тестів *in vitro*, але й дорогі експерименти *in vivo*.

Методи докінгу також використовують для вивчення взаємодій між протеїнами. Незважаючи на постійне зростання кількості структур у базі даних PDB, число доступних протеїн-протеїнових комплексів відносно невелике. У разі застосування сучасних експериментальних методів, використуваних для визначення структури з високою роздільною здатністю, таких, зокрема, як РСА (рентгеноструктурний аналіз) та ЯМР (ядерно-магнітний резонанс), виникають певні труднощі з отриманням структур великих надмолекулярних асоціатів. Тому алгоритми докінгу можуть бути корисними для вивчення просторових структур таких комплексів. Структурне моделювання протеїн-протеїнових комплексів,

що утворюються при патологічних станах, є важливим для розроблення інгібіторів до інтерфейсів їх взаємодії з метою порушення агрегації.

Припускають, що для комп'ютерного моделювання надмолекулярного комплексу доцільнішим було б застосування підходів квантової механіки. У цьому разі взаємодія між лігандом і рецептором може бути з'ясована в результаті розв'язання комбінованого рівняння Шредінгера для обох систем. Але цей метод є надзвичайно складним, і навіть моделі квантового поля, які чисельно вирішуються за допомогою комп'ютера, не дають змоги отримати чітке вирішення. Тому значно частіше використовують більш примітивну, механічну модель. Ця модель ґрунтується на вивченні сил між взаємодіючими частинками [33].

У методі молекулярної механіки атоми розглядають як ньютонівські частинки, які взаємодіють через певні силові поля. Потенційна енергія взаємодії залежить від довжини зв'язків, кутів зв'язків, торсійних кутів і нековалентних взаємодій.

Набір параметрів, що складається з рівноважних значень довжин зв'язків, валентних кутів, величин парціальних зарядів, силових констант і ван-дер-ваальсових параметрів, називається силовим полем.

Силові поля створено на основі параметрів із відносно невеликого експериментально отриманого ряду молекул і, як правило, їх використовують, щоб спрогнозувати формування комплексів. Прикладами моделей силових полів є AMBER, CHARMM і CVFF.

Силове поле AMBER (Assisted Model Building with Energy Refinement) було розроблено для протеїнів і нуклеїнових кислот. Це поле дозволяє враховувати всі атоми окремо або ж вибрати опцію об'єднаного атома, під яким розуміють групу еквівалентних атомів з однаковими властивостями. В останньому випадку декілька атомів або груп атомів досліджують як один атом певного типу. Сет параметрів AMBER отриманий експериментально. Силові поля AMBER є найпоширенішими [34].

Силові поля CHARMM (Chemistry at HARvard Macromolecular Mechanics) розроблено для протеїнів, ДНК, РНК та ліпідів. Можливе також використання комбінації полів, наприклад для моделювання взаємодії ДНК із протеїнами. Існують поля, що розглядають об'єднаний атом, і такі, що враховують усі атоми [34]. CHARMM також включає поляризовані силові поля, що ґрунтуються на моделі флуктуації заряду або

дисперсній моделі осцилятора [35]. Параметри для силових полів CHARMM можна вільно завантажити із сайту http://mackerell.umaryland.edu/CHARMM_ff_params.html.

CVFF (Consistent Valence Force Field) містить параметри, одержані у процесі встановлення відповідності кристалічної і газової структур низькомолекулярних органічних молекул. CVFF створений переважно для органічних сполук і його використовують для вивчення структур та підрахунку енергій зв'язування [33].

Силове поле CFF (Consistent Force Field) має низку переваг над іншими полями, оскільки параметри для оптимізації в ньому передбачені більш точно. CFF може оптимізувати структури, враховуючи довжини зв'язків, значення валентних, торсійних кутів, відстаней між атомами, енергії решіток, молекулярної потенційної енергії, дипольних моментів та частоти вібрації. CFF дозволяє використовувати квантові розрахунки для визначення параметрів енергетичних функцій [36].

COMPASS (Condensed-phase Optimized Molecular Potentials for Atomistic Simulation Studies) — це *ab initio* модель силового поля. Подібно до CFF параметри цього поля встановлюються у підрахунках квантової механіки і підтверджуються емпіричними даними [37].

Існують узагальнені силові поля, які за точністю значно поступаються вищезазначеним. Їх застосовують до систем, що не охоплюються більш точними моделями силового поля. Прикладами узагальнених полів є ESFF і UFF.

ESFF (Extensible Systematic Force Field) створено для моделювання органічних, неорганічних та металоорганічних систем. У це силове поле було введено такі параметри, як псевдокут і функція скалярного добутку, яка характеризує енергію торсійних деформацій. Часткові атомні заряди визначаються із *ab initio* підрахованої електронегативності та жорсткості валентних орбіталей. ESFF є цінним інструментом у молекулярному моделюванні для дослідження і вивчення кристалічної структури та газової фази молекул [38].

UFF (Universal Force Field) включає всю таблицю Менделєєва. Параметри цього силового поля ґрунтуються на властивостях атомів, їх гібридизації та можливості формування зв'язку. Однак це поле не є достатньо точним і його застосовують до систем, що не обчислюються іншими силовими полями [39].

Алгоритми докінгу

Основним завданням молекулярного докінгу є пошук можливих способів зв'язування. Точний пошуковий алгоритм має визначати й оцінювати якомога більшу кількість способів зв'язування між двома молекулами. Проте цей процес може потребувати надто багато комп'ютерного часу. Тому потрібно досягти узгодження між комп'ютерною вартістю процесу та пошуковим простором.

У літературі відомі такі пошукові алгоритми:

- молекулярна динаміка;
- метод Монте Карло;
- генетичні алгоритми;
- методи, що ґрунтуються на фрагментах;
- методи комплементарності точок;
- пошук заборон;
- систематичний пошук.

Методи молекулярної динаміки дають змогу розв'язувати рівняння руху Ньютона. Мінімум енергії докованого комплексу знаходять з використанням алгоритмів стандартної оптимізації, включаючи:

- прямий пошук із застосуванням лише потенціальної функції; не придатний для великих молекул, підходить лише для грубої оптимізації маленьких молекул, визначає стан системи, що надзвичайно далекий від мінімального значення енергії, наприклад метод симплекс;

- градієнтні методи — базуються на першій похідній функції потенціалу, забезпечують встановлення стану системи з енергією, що є далекою від мінімуму, придатні для початкової оптимізації, наприклад алгоритм крутого спуску;

- методи спряженого градієнта — характеризуються кращою здатністю до мінімізації досліджуваної системи, наприклад алгоритм Fletcher-Reeves;

- методи другої похідної — дозволяють ще ефективніше здійснювати мінімізацію, наприклад алгоритм Newton-Raphson;

- методи найменших квадратів — забезпечують встановлення стану системи, що є дуже близьким до мінімуму, але комп'ютерно дуже дорогі, наприклад Marquardt.

Часто використовують комбінацію зазначених методів, зокрема поєднання градієнтного методу для початкової оптимізації та методу спряженого градієнта для визначення мінімуму [33].

Метод Монте Карло посідає особливе місце в історії молекулярного моделювання, оскільки він перший був використаний у комп'ютерному моделюванні молекулярних систем. Алгоритм Монте Карло — це за-

гальна назва групи числових методів, що ґрунтуються на одержанні великої кількості реалізацій стохастичного (випадкового) процесу. Метод Монте Карло генерує випадкові рухи системи і потім приймає або відкидає рух залежно від імовірності перебування частки в стані з певною енергією (розподіл Больцмана) [40]. Методи Монте Карло відіграють важливу роль у здійсненні молекулярного докінгу. Програми, що використовують метод Монте Карло, — AutoDock, ProDock, ICM, MCDOCK, DockVision, QXP та Affinity.

Генетичний алгоритм — це еволюційний алгоритм пошуку, що його використовують для вирішення завдань оптимізації і моделювання шляхом послідовного підбору, комбінування і варіації досліджуваних параметрів. Важлива ідея генетичних алгоритмів — еволюція популяції можливих розв'язків через генетичні оператори (мутації, кросовери та міграції) до кінцевої популяції.

Процес застосування генетичних алгоритмів починається з кодування змінних, наприклад ступенів свободи, у так званий «генетичний код». Потім створюється випадкова початкова популяція розв'язувань. Далі до цієї популяції застосовуються генетичні оператори, що призводить до формування нової популяції. Потім нова популяція оцінюється, і вибір для наступних раундів залежить від параметрів енергії. Слід зазначити, що генетичні алгоритми добре підходять для паралельного обчислення [41]. Програми, в яких використовують генетичні алгоритми, — GOLD, AutoDock, DIVALI і DARWIN.

Методи, що ґрунтуються на фрагментах, дозволяють розділяти ліганди на частини, докувати їх і потім з'єднувати разом. Для цих методів вкрай важливим є вдалий вибір основного фрагмента — функціональної групи ліганду. Невдалий вибір може значно знизити якість результатів. Основний фрагмент має забезпечити домінуючі взаємодії з рецептором [42].

Програми, в яких використовують методи, що ґрунтуються на фрагментах, — FlexX і DOCK.

Методи комплементарності точок ґрунтуються на оцінюванні стеричної і/або хімічної комплементарності між взаємодіючими молекулами. Атоми взаємодіючих молекул представлені у вигляді сфер або кубів. Ліганд переміщується і обертається для отримання максимальної кількості відповідностей із поверхнею протеїну, не враховуючи перекривання. Можуть зазначатися додаткові умови, наприклад, нормалі взаємодіючих поверхонь мають бути спрямовані у протилежних напрямках.

Деякі алгоритми застосовують просторові решітки, які включають в себе протеїн і ліганд. Потім створюється кореляційна функція, що оптимізується під час переміщення або обертання ліганду. При цьому часто використовують традиційні алгоритми розпізнавання поверхні, такі як швидке перетворення Фур'є із кореляційною теорією Фур'є. Високий кореляційний показник вказує на достовірну відповідність між поверхнями молекул [43]. Приклади програм із використанням методів точкової комплементарності, — FTDOCK, SANDOCK, FLOG та алгоритм Soft Docking.

Методи пошуку заборон ґрунтуються на стохастичних процесах, у яких нові стани випадково генеруються з вихідного стану. Ці нові розв'язування потім оцінюються і сортують. Найкраще нове розв'язування потім вибирають як вихідний стан і процес повторюється знову. Для гарантування різноманітності вихідного стану використовують лист заборон, що діє як пам'ять. Він містить інформацію про попередні вихідні стани, і якщо нове розв'язування дуже нагадує попереднє, його видаляють [44]. Приклад алгоритму докінгу, в якому використовують пошук заборон, — PRO_LEADS.

Систематичний пошук визначає усі можливі конформації при утворенні комплексу. Молекули в цьому алгоритмі розглядають як жорсткі і енергію взаємодії оцінюють з моделі силового поля [45].

Програмне забезпечення

AutoDock — автоматичний інструмент для реалізації докінгу. Програма визначає спосіб зв'язування низькомолекулярних субстратів із рецептором, у якого відома просторова структура. AutoDock складається з двох основних програм: власне AutoDock, що здійснює докінг ліганду до сету решіток, які описують протеїн-мішень, і AutoGrid, що розраховує ці решітки. Для створення сету можливих конформацій AutoDock використовує алгоритм модельного «загартування» Монте Карло та генетичний алгоритм Ламарка. Останній застосовують для глобальної оптимізації та для мінімізації енергії. Вірогідні орієнтації оцінюють за допомогою моделі силового поля AMBER разом зі скоринговими функціями вільної енергії та великим сетом комплексів протеїну з лігандами з відомими константами взаємодії. AutoDock дозволяє враховувати гнучкість бічних ланцюгів амінокислот активного сайту рецептора. Веб-сторінка програми AutoDock: <http://autodock.scripps.edu/>.

DOCK — одна з перших і найвідоміших програм для докінгу. Початкова версія давала змогу використовувати негнучкі ліганди; гнучкість було введено пізніше завдяки поступовому конструюванню ліганду у зв'язувальній кишені. Іншими словами, DOCK — це метод, який ґрунтується на фрагментах і використовує методи хімічної та стеричної комплементарності для генерації можливих орієнтацій ліганду. Процес докінгу за допомогою програми DOCK включає декілька етапів. На першому відбувається додавання протонів та часткових зарядів як до низкомолекулярної сполуки, так і до мішені. Потім створюється негативне зображення поверхні активного сайту за допомогою сфер. Далі програма розраховує енергетичну решітку мішені. Кожна точка решітки містить інформацію про заряд відповідної ділянки та ван-дер-ваальсові характеристики. На останній стадії DOCK наближує атоми ліганду до центрів сфер і оцінює відносно решітки. Програма DOCK є надзвичайно зручною для швидкого проведення докінгу, але за точністю дещо поступається іншим методам. Сайт програми DOCK: <http://dock.compbio.ucsf.edu/>.

FlexX — інший метод, який ґрунтується на фрагментах із застосуванням гнучких лігандів і негнучких рецепторів та бази даних торсійних кутів MIMUMBA для створення конформерів. Для оцінювання використовують функцію Бема (з незначними адаптаціями, необхідними для докінгу). Метод FlexX має дещо меншу швидкість, ніж DOCK, але забезпечує краще оцінювання середньоквадратичного відхилення для лігандів із коректно передбаченим сайтом зв'язування. Розширена версія FlexX враховує гнучкість рецепторів і називається FlexE. Веб-адреса цієї програми: <http://www.biosolveit.de/FlexX/index.html?ct=1>.

Gold має високий коефіцієнт ефективності пошуку, який дещо знижений у випадку гідрофобних сайтів. Для проведення докінгу гнучких лігандів і рецепторів із рухомими ОН-групами Gold використовує генетичний алгоритм. При цьому протеїн розглядається як негнучкий. Ця програма є адекватним вибором, якщо кишеня зв'язування містить амінокислоти, здатні формувати водневі зв'язки з лігандом. Gold дозволяє використовувати скорингову функцію, що ґрунтується на вигідних конформаціях, виявлених у Кембриджській структурній базі даних та на емпіричних результатах про хімічні взаємодії. Gold має найбільшу кількість сетів із тестами валідації. Сайт Gold:

http://www.ccdc.cam.ac.uk/products/life_sciences/gold/.

ZDOCK — програма для проведення протеїн-протеїнового докінгу. Для пошуку можливих способів зв'язування протеїнів у ZDOCK застосовують швидке перетворення Фур'є. Оцінювання здійснюється на основі комплементарності поверхонь, енергії десольватації та електростатики. Найкращі результати ZDOCK використовуються програмою RDOCK для мінімізації за допомогою силового поля CHARMM з метою поліпшення енергетичних характеристик. RDOCK також дозволяє точніше перерахувати енергію десольватації та електростатичну енергію. Веб-сторінка програми: <http://zdock.bu.edu/>.

Програма DOT (Daughter of Turnip) дає змогу знаходити протеїн-протеїнові комплекси з низькою енергією, здійснюючи систематичний пошук за шістьма ступенями свободи, підраховувати енергію взаємодії для багатьох орієнтацій рухомої молекули і зберігати окремі листи з оцінкою електростатичної енергії, енергії ван-дер-ваальсових взаємодій чи із сумою обох. Вільна енергія є сумою фактора Больцмана за всіма напрямками обертання в кожній точці решітки. Програма DOT розташована за електронною адресою: <http://www.sdsc.edu/CCMS/DOT/>.

Програма FTDock (Fourier Transform Dock) дозволяє здійснювати жорсткий докінг двох молекул для прогнозування правильної геометрії їх зв'язування, а також застосовувати кореляційний алгоритм Фур'є та електростатичний фільтр, який за бажанням можна відключити.

Програма RPScore (Residue level Pair potential Score) повністю інтегрована із FTDock. У RPScore використовують емпіричну матрицю потенціалів на рівні амінокислотних залишків для оцінювання кожного можливого комплексу. Кожен потенціал відповідає емпірично отриманій вірогідності формування пари двох типів залишків, обмеженій лише параметрами відстані.

Програму MultiDock (Multiple copy side-chain refinement Dock) було розроблено для деталізації інтерфейсу між двома взаємодіючими протеїнами на атомному рівні. Програма уможливує моделювання ефектів конформаційних змін бічних ланцюгів і рухи жорстких частин взаємодіючих протеїнів. Протеїн описується на атомному рівні електростатичними та ван-дер-ваальсовими взаємодіями, в якому бічні ланцюги амінокислотних залишків змодельовано багатокопійним представленням на фіксованому пептидному ланцюгу згідно з ротамерною

бібліотекою. Для кожного циклу оптимізації усі взаємодії між бічними ланцюгами амінокислотних залишків оцінює і зберігає програма. MultiDock також дає можливість виконувати стандартну мінімізацію енергії жорстких частин молекул для зменшення напруги інтерфейсу. Сайт цього пакета програм: <http://www.sbg.bio.ic.ac.uk/docking/>.

Програма MolFit дає змогу проводити протеїн-протеїновий докінг. MolFit репрезентує кожен молекулу через просторову решітку і визначає ступінь геометричної та хімічної комплементарності поверхонь. У програмі MolFit вперше було використано швидке перетворення Фур'є для знаходження кореляційної функції. Веб-сторінка MolFit: http://www.weizmann.ac.il/Chemical_Research_Support/molfit/.

ICM-Docking можна використовувати для проведення протеїн-протеїнового, пептидно-протеїнового та ліганд-протеїнового докінгу. Стандартна процедура докінгу за допомогою ICM передбачає наявність гнучкого ліганду і напівжорсткого рецептора. Гнучкість рецептора може досягатися або завдяки використанню множинних конформацій рецептора, або завдяки гнучкості бічних ланцюгів амінокислотних залишків. У ICM-Docking для пошуку можливих конформацій застосовують алгоритм Монте Карло. Сайт програми ICM-Docking: <http://www.molsoft.com/docking.html>.

Програму Glide використовують для високопродуктивного віртуального скринінгу мільйонів сполук та передбачення способів зв'язування ліганду з рецептором. Ця програма має дві різні скорингові функції — SP і XP GlideScore. Три способи вибору конформацій ліганду і ступенів свободи доступні для визначення оптимальної орієнтації ліганду відносно жорсткого протеїнового рецептора. Веб-сторінка програми Glide: <http://www.schrodinger.com/products/14/5/>.

Hex — інтерактивна програма для проведення докінгу. Алгоритм докінгу ґрунтується на сферичних полярних кореляціях Фур'є

та електростатичних характеристиках. Дозволяє розпізнавати структури протеїнів та ДНК у PDB-форматі, а також читати SDF-файли низькомолекулярних сполук. Сайт програми Hex: <http://www.loria.fr/~ritchied/hex/>.

У роботах іноземних авторів проводилося порівняння інструментів докінгу за низкою параметрів. Одним з найважливіших аспектів комп'ютерних технологій для раціонального дизайну ліків є точність докінгу (теоретичне передбачення структури, отриманої за допомогою PCA). Програми, що найбільшою мірою відповідають цій умові, — FlexX, Gold, Glide і ICM-Docking. Недоліком цих програм є низька швидкість [46, 47]. Програмі Dock притаманна досить висока швидкість роботи, що дозволяє проводити паралельні обчислення і робить її зручною для здійснення віртуального скринінгу великих бібліотек сполук. Програми ZDOCK, DOT, MultiDock і MolFit застосовують для протеїн-протеїнового докінгу.

Підсумовуючи, слід зазначити, що в сучасних біохімічних дослідженнях широко застосовують комбінування методів вирівнювання послідовностей/структур, молекулярного докінгу та експериментального підтвердження. В огляді описано основні методологічні підходи біоінформатики. Деякі алгоритми є специфічними для протеїнів, інші можуть застосовуватися до різних типів молекул, наприклад ДНК, РНК, ліпідів та низькомолекулярних сполук. Структурний аналіз має значні переваги над вирівнюванням послідовності, оскільки дає змогу отримувати більш точну інформацію про функціональну та еволюційну спорідненість протеїнів. Методи молекулярного докінгу широко використовують для раціонального дизайну ліків. Незважаючи на значний прогрес у галузі комп'ютерного моделювання, алгоритми докінгу потребують удосконалення багатьох параметрів, зокрема підвищення їхньої точності, швидкості, а також можливості врахування гнучкості взаємодіючих молекул.

ЛІТЕРАТУРА

1. Needleman S. B., Wunsch C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins // *J. Mol. Biol.* — 1970. — V. 48, N 3. — P. 443–453.
2. Smith T. F., Waterman M. S. Identification of Common Molecular Subsequences // *Ibid.* — 1981. — V. 147, N 1. — P. 195–197.
3. Mount D. M. Dot Matrix Pairwise Sequence Comparison // *Cold Spring Harb. Protoc.* — 2007; doi:10.1101/pdb.top31.
4. Mount D. M. *Bioinformatics: Sequence and Genome Analysis* (2nd ed.). — Cold Spring Harbor Labor. Press: Cold Spring Harbor, NY. — 2004.
5. Lipman D. J., Altschul S. F., Kececioglu J. D. A tool for multiple sequence alignment // *Proc. Natl. Acad. Sci. USA.* — 1989. — V. 86, N 12. — P. 4412–4415.
6. Chenna R., Sugawara H., Koike T. et al. Multiple sequence alignment with the Clustal series of programs // *Nucl. Acids Res.* — 2003. — V. 31, N 13. — P. 3497–3500.

7. *Notredame C., Higgins D. G., Heringa J. T. T.* Coffee: A novel method for fast and accurate multiple sequence alignment // *J. Mol. Biol.* — 2000. — V. 302, N 1. — P. 205–217.
8. *Hirosawa M., Totoki Y., Hoshida M., Ishikawa M.* Comprehensive study on iterative algorithms of multiple sequence alignment // *Comput. Appl. Biosci.* — 1995. — V. 11, N 1. — P. 13–18.
9. *Marshall T., Rahmann S.* Efficient exact motif discovery // *Bioinformatics.* — 2009. — V. 25, N 12. — P. i356–i364.
10. *Bradley R. K., Roberts A., Smoot M. et al.* Fast statistical alignment // *PloS Comput. Biol.* — 2009. — V. 5, N 5. — P. e1000392.
11. *Schwartz A. S., Pachter L.* Multiple alignment by sequence annealing // *Bioinformatics.* — 2007. — V. 23, N 2. — P. e24–e29.
12. *Goldsmith-Fischman, S., Honig, B.* Structural genomics: Computational methods for structure analysis // *Protein Sci.* — 2003. — V. 12, N 9. — P. 1813–1821.
13. *Dietmann S., Park J., Notredame C., Heger A. et al.* A fully automatic evolutionary classification of protein folds: Dali Domain Dictionary version 3 // *Nucl. Acids Res.* — 2001. — V. 29, N 1. — P. 55–57.
14. *Holm L., Sander C.* Protein structure comparison by alignment of distance matrices // *J. Mol. Biol.* — 1993. — V. 233, N 1. — P. 123–138.
15. *Taylor W. R., Flores T. P., Orengo C. A.* Multiple protein structure alignment // *Protein Sci.* — 1994. — V. 3, N 10. — P. 1858–1870.
16. *Orengo C. A., Michie A. D., Jones S. et al.* CATH — a hierarchic classification of protein domain structures // *Structure.* — 1997. — V. 5, N 8. — P. 1093–1108.
17. *Shindyalov I., Bourne P.* Protein structure alignment by incremental combinatorial extension (CE) of the optimal path // *Protein Eng.* — 1998. — V. 11, N 9. — P. 739–747.
18. *Yang A., Honig B.* An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance // *J. Mol. Biol.* — 2000. — V. 301, N 3. — P. 691–711.
19. *Leibowitz N., Nussinov R., Wolfson H. J.* MUSTA — a general, efficient, automated method for multiple structure alignment and detection of common motifs: application to proteins // *J. Comp. Biol.* — 2001. — V. 8, N 2. — P. 93–121.
20. *Shatsky M., Nussinov R., Wolfson H. J.* MultiProt — a multiple protein structural alignment algorithm // *Work. Algor. Bioinform.* — 2002. — V. 2452. — P. 235–250.
21. *Dror O., Benyamini H., Nussinov R., Wolfson H.* Multiple structural alignment by secondary structures: algorithm and applications // *Protein Sci.* — 2003. — V. 12, N 11. — P. 2492–2507.
22. *Bi C.* A Monte Carlo EM algorithm for de novo motif discovery in biomolecular sequences // *IEEE/ACM Trans. Comput. Biol. Bioinform.* — 2009. — V. 6, N 3. — P. 370–386.
23. *Verbitsky G., Nussinov R., Wolfson H.* Flexible structural comparison allowing hinge-bending, swiveling motions // *Proteins.* — 1999. — V. 34, N 2. — P. 232–254.
24. *Shatsky M., Nussinov R., Wolfson H.* Flexible protein alignment and hinge detection // *Proteins: Structure, Function, and Genetics.* — 2002. — V. 48, N 2. — P. 242–256.
25. *O’Hearn S., Kusalik A., Angel J.* MolCom: a method to compare protein molecules based on 3-D structural and chemical similarity // *Protein Eng.* — 2003. — V. 16, N 3. — P. 169–178.
26. *Nagano N., Orengo C. A., Thornton J. M.* One fold with many functions — the evolutionary relationships between TIM barrel families based on their sequences, structures and functions // *J. Mol. Biol.* — 2002. — V. 321, N 5. — P. 741–765.
27. *Kuttner Y. Y., Sobolev V., Raskind A., Edelman M.* A consensus-binding structure for adenine at the atomic level permits searching for the ligand site in a wide spectrum of adenine-containing complexes // *Proteins.* — 2003. — V. 52, N 3. — P. 400–411.
28. *Denessiouk K. A., Rantanen V., Johnson M. S.* Adenine recognition: a motif present in ATP-, CoA-, NAD-, NADP-, and FAD-dependent proteins // *Ibid.* — 2001. — V. 44, N 3. — P. 282–291.
29. *Spriggs R. V., Artymiuk P. J., Willett P.* Searching for patterns of amino acids in 3D protein structures // *J. Chem. Inf. Comput. Sci.* — 2003. — V. 43, N 2. — P. 412–421.
30. *Brakoulias A., Jackson R. M.* Towards a structural classification of phosphate binding sites in protein-nucleotide complexes: An automated all-against-all structural comparison using geometric matching // *Proteins.* — 2004. — V. 56, N 2. — P. 250–260.
31. *Shulman-Peleg A., Nussinov R., Wolfson H. J.* Recognition of functional sites in protein structures // *J. Mol. Biol.* — 2004. — V. 339, N 3. — P. 607–633.
32. *Jones S., Thornton J. M.* Searching for functional sites in protein structures // *Curr. Opin. Chem. Biol.* — 2004. — V. 8, N 1. — P. 3–8.
33. *Kaapro A., Ojanen J.* Protein docking // <http://www.lce.hut.fi/teaching/S-114.500/k2002/Protdock.pdf>; 2002.
34. *Cornell W. D., Cieplak P., Bayly C. I. et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules // *J. Am. Chem. Soc.* — 1996. — V. 117. — P. 5179–5197.
35. *Lamoureux G., Roux B.* Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm // *J. Chem. Phys.* — 2003. — V. 119, N 6. — P. 3025–3039.

36. *Jynsdyttir S. Y., Rasmussen K.* The consistent force field. Part 6: an optimized set of potential energy functions for primary amines // *New J. Chem.* — 2000. — V. 24. — P. 243–247.
37. *Sun H.* COMPASS: an ab initio force-field optimized for condensed-phase applications — overview with details on alkane and benzene compounds // *J. Phys. Chem. B.* — 1998. — V. 102, N 38. — P. 7338–7364.
38. *Shi S., Yan L., Yang Y. et al.* An extensible and systematic force field, ESFF, for molecular modeling of organic, inorganic, and organometallic systems // *J. Comput. Chem.* — 2003. — V. 24, N 9. — P. 1059–1076.
39. *Rappe A. K., Casewit C. J., Colwell K. S. et al.* UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations // *J. Am. Chem. Soc.* — 1992. — V. 114, N 25. — P. 10024–10035.
40. *Metropolis N., Ulam S.* The Monte Carlo Method // *J. Am. Stat. Assoc.* — 1949. — V. 44, N 247. — P. 335–341.
41. *Fogel D. B.* Evolutionary Computation: Towards a New Philosophy of Machine Intelligence. — New York: IEEE Press, 2000. — P. 140.
42. *Taylor R. D., Jewsbury P. J., Essex J. W.* A review of protein-small molecule docking methods // *J. Comput. Aid. Mol. Des.* — 2004. — V. 16, N 3. — P. 151–166.
43. *Gabb H. A., Jackson R. M., Sternberg M. J. E.* Modeling Protein Docking using shape complementarity, electrostatics and biochemical information // *J. Mol. Biol.* — 1997. — V. 272, N 1. — P. 106–120.
44. *Baxter C. A., Murray C. W., Clark D. E. et al.* Flexible docking using Tabu search and an empirical estimate of binding affinity // *Proteins.* — 1998. — V. 33, N 3. — P. 367–382.
45. *Wang J., Kollman P. A.* Automatic parameterization of force field by systematic search and genetic algorithms // *J. Comput. Chem.* — 2001. — V. 22, N 12. — P. 1219–1228.
46. *Kellenberger E., Rodrigo J., Muller P., Rognan D.* Comparative evaluation of eight docking tools for docking and virtual screening accuracy // *Proteins: Structure, Function, and Bioinformatics.* — 2004. — V. 57. — P. 225–242.
47. *Cross J. B., Thompson D. C., Rai B. K. et al.* Comparison of several molecular docking programs: pose prediction and virtual screening accuracy // *J. Chem. Inf. Model.* — 2009. — V. 59, N 6. — P. 1455–1474.

МЕТОДЫ СТРУКТУРНОЙ БИОИНФОРМАТИКИ

*Г. П. Вольнец
В. Г. Бджола
С. М. Ярмолюк*

Институт молекулярной биологии и генетики
НАН Украины, Киев

E-mail: galina.volinetc@gmail.com

В обзоре представлены наиболее распространенные методы компьютерного моделирования в биологии. Описаны технологии сравнения аминокислотной последовательности и структурного выравнивания, используемые для поиска функционально подобных и гомологичных протеинов. Рассмотрены программы докинга, которые применяются для изучения формирования надмолекулярных комплексов и являются неотъемлемым инструментом рационального дизайна лекарств. Комбинирование методов компьютерного моделирования с общепринятыми экспериментальными подходами позволяет значительно снизить стоимость исследований.

Ключевые слова: выравнивание аминокислотных последовательностей, структурное выравнивание протеинов, функциональный сайт, молекулярный докинг.

METHODS OF STRUCTURAL BIOINFORMATICS

*G. P. Volynets
V. G. Bdzholo
S. M. Yarmoluk*

Institute of Molecular Biology and Genetics
of the National Academy of Sciences
of Ukraine, Kyiv

E-mail: galina.volinetc@gmail.com

The most widespread methods of computer modeling in biology is given in the review. Protein sequence and structure alignments assist in detection of functional/evolutionary relationship. Molecular docking algorithms are powerful tools for rational drug design. Protein-protein docking may be used to predict the intermolecular complexes structure. The integration of computational tools with experimental approaches allows to reduce significantly the research cost.

Key words: amino acids sequence alignment, protein structure alignment, functional site, molecular docking.