

МЕТОДИ АНАЛІЗУ ТА МОДЕЛЮВАННЯ ПРОСТОРОВОЇ СТРУКТУРИ БІЛКІВ

У статті розглянуто методи комп'ютерного аналізу і моделювання просторової структури білків. Описано використання серверів Internet та "локальних" програм при комп'ютерному моделюванні. Запропоновані підходи до аналізу гомологічних послідовностей, візуалізації просторових структур, моделювання структури за гомологією, мінімізації вільної енергії, молекулярної динаміки, взаємодії макромолекул. Вказані Web-адреси понад 110 серверів і програм.

1. ВСТУП

Визначення просторової (тривимірної, 3D-) структури білків є необхідним етапом для встановлення взаємозв'язку між структурою та функцією білків. Значні успіхи у секвенуванні геномів організмів сприяли появі нових даних про амінокислотні (АК) послідовності (первинні структури) тисяч різноманітних білків [1–3]. З них лише приблизно для 10% відомих АК-послідовностей доступні структурні дані, що становить понад 15 тис. експериментально визначених структур [4]. Тому для аналізу більшості білків дедалі важливішими стають комп'ютерні методи передбачення їх просторової структури, які належать до методів "біології *in silico*" [5–10].

Переваги і обмеження експериментальних методів. Прогрес останніх десятиліть у визначенні просторових структур білків пов'язаний із розвитком та удосконаленням методу рентгеноструктурного аналізу (РСА), який дозволяє визначити координати атомів білків з високою точністю. Проте проведення РСА білків — це складна і трудомістка процедура, критична стадія якої полягає в отриманні стабільних кристалів високоочищеного білка з високою дифрагуючою здатністю. Різні білки мають індивідуальні (часто досить чутливі до зовнішнього впливу) умови кристалізації. Деякі білки можуть кристалізуватися тільки у вигляді протеолітичних фрагментів або не кристалізуються взагалі. Крім того, для мембранних білків необхідно розробляти спеціальні методи для кристалізації. Незважаючи на вражаючі успіхи рентгенівської кристалографії (наприклад, кристалізація та розшифровка просторової структури 70S рибосоми [11]), в багатьох випадках метод РСА не є ефективним при вивченні білок-білкових та білок-нуклеїнових комплексів.

Спектроскопія ядерного магнітного резонансу (ЯМР) є альтернативним методом визначення просторової структури білка. Перевагою методу ЯМР є

можливість визначити структуру білків у розчині в нативних умовах, проте значним недоліком є обмеженість встановлення структури відносно великих білків (понад 20 кДа).

На основі експериментально визначених структур білків були розроблені певні емпіричні правила та алгоритми передбачення їх 3D-структури, які згодом були покладені в основу молекулярного моделювання просторової організації білків. Ці методи розвивалися спочатку як допоміжні при уточненні РСА- та ЯМР-структур, що дало змогу встановити загальні правила упаковки поліпептидних ланцюгів, а згодом розвинути новий підхід на основі конформаційного аналізу та мінімізації вільної енергії молекул білків.

2. РОБОТА з Internet ТА ЛОКАЛЬНИМИ ПРОГРАМАМИ

У статті ми надавали перевагу серверам Internet та програмам, що керуються операційною системою (ОС) Windows. У табл. 1 наведені типові етапи роботи в процесі аналізу і моделювання структури білків, а також ресурси Internet, для цього призначені.

ОСНОВНІ СЕРВЕРИ, ПЕРЕЛІКИ СЕРВЕРІВ ТА КОМП'ЮТЕРНІ ПРОГРАМИ. Існують потужні сервери (Web-сайти), які є світовими центрами з біоінформатики. Прикладами таких серверів є *NCBI*, *ExpASy*, *BioPlanet*, *Frontier*. Також існують "вторинні" сайти, що містять дуже корисні переліки посилань на інші Web-сайти (links), впорядковані за темами, зокрема, за різними аспектами аналізу і моделювання структури білків. Характерною ознакою *on-line* серверів і програм є діалогові вікна, в які користувач має ввести потрібну інформацію, а також ряд перемикачів, за допомогою яких задаються параметри для роботи програми чи виводу результатів. Існує тенденція щодо інтеграції окремих програм до програмних комплексів, які виконують роботу за принципом конвеєра, що значно

полегшує проведення типових робіт. У майбутньому це має призвести до появи світових автоматизованих Internet серверів (services), які паралельно виконуватимуть різні замовлення на великих комп'ютерних системах, як це вже має місце на прикладі серверів *BLAST*, *Swiss-Model*, *META* та ін.

Локальні програми під ОС Windows та Linux. На жаль, поки що не всі етапи роботи, пов'язані з моделюванням білків, можна виконувати через Internet. Виникає потреба встановлення на локальному комп'ютері окремих програм і пакетів програм (насамперед "вьюверів" (viewers) для графічного зображення і аналізу, програм оптимізації структур тощо). Більшість локальних програм для мінімізації енергії макромолекул, докінгу та ін. керуються операційною системою Linux (версії Black Cat, Red Hat, Free BSD та ін.). Програми під ОС Linux переписують з Web-сайтів у вигляді дистрибутивів, переважно заархівованих стандартними Linux-архіваторами (файли типів *.tar, *.Z, *.gz). Linux-програми компілюються під конкретний комп'ютер (hardware), і тому вони не містять вже готових до виконання файлів, як Windows-програми. Бажано заздалегідь встановити повну колекцію компіляторів з різних комп'ютерних мов, оскільки деякі програмні модулі пишуть різними мовами. Компіляцію проводять програмою *make*. Більшість з наведених Linux-програм не мають графічного інтерфейсу (виняток — програма для докінгу *HEX*), тому при роботі з ними потрібно попередньо написати керуючий script-файл, який містить вихідні дані для їх роботи.

3. Методи аналізу білків на рівні первинної структури

На першому етапі роботи потрібно виходячи з амінокислотної послідовності білка передбачити його загальні властивості, такі як доменна організація, вторинна структура, консервативні та варіабельні ділянки тощо.

Банки даних послідовностей. Інформація про визначені послідовності (первинні структури) біополімерів зберігається у спеціалізованих банках даних, які безперервно і автоматизовано поповнюються. Доступ до цих послідовностей можливий через пошук за ключовими словами або методом пошуку за гомологією. В записках банків даних наводяться додаткові відомості та різні перехресні посилання (наукові статті, інформація про вторинну структуру, різні аналізи, їх похибки тощо). Для генів і геномів це банки даних *GenBank*, *EMBL*, *DDBJ* та ін., а для білків — *GenPept*, *TrEMBL*, *Swiss-Prot*, *OWL*, *NBRF-PIR* та ін. Рекомендується використовувати автоматизовані банки даних послідовностей *GenPept* та *TrEMBL*, у яких зберігаються "трансляції" всіх нуклеотидних послідовностей з *GenBank* (*EMBL*). Інший банк даних — *Swiss-*

Prot є "класичним" і містить розвинену систему анотацій та посилань на інші чисельні банки даних.

Передбачення фізико-хімічних властивостей білків. Чисельні властивості білків (молекулярна маса, ізоелектрична точка (pI), оптичні властивості та ін.) можна розрахувати, виходячи з їх АК-послідовностей за допомогою досить простих програм. Зручні переліки посилань на такі програми є у розділах *Proteomic Tools* на серверах *ExpASY*, *Frontier*, *ISREC*.

Пошук за гомологією. Важливим етапом комп'ютерного аналізу білка є пошук гомологічних АК-послідовностей серед усіх відомих послідовностей [12]. Існують численні програми пошуку за гомологією, які можуть використовувати різні матриці та алгоритми вирівнювання, і враховують не тільки ідентичні, а й схожі за властивостями АК-залишки. При використанні удосконалених методів вирівнювання багатьох послідовностей враховується не тільки формальна схожість двох послідовностей, а й консервативність окремих залишків. Найбільш вживаною є програма *BLAST* та її чисельні спеціалізовані клони [12]. Офіційний сервер — *BLAST*, існують варіанти цієї програми (*blastp*, *blastn*, *tblastn*, *blastx*, *tblastx*) для різних типів пошуку залежно від типу надісланої послідовності і банку даних, в якому іде пошук. Інтегровані бази даних NR (non redundant) — це ті, в яких вилучені всі дублюючі послідовності. Подібними за призначенням є також програми *FastA*, *Blitz*. Варіант пошуку за гомологією, коли процес вирівнювання повторюється ітеративно з уточненням ролі найбільш консервативних залишків, дозволяє знаходити значно віддалені гомологи (програми *PSI-BLAST* *PHI-BLAST*).

Множинне вирівнювання. Гомологічні АК-послідовності мають бути порівняні між собою методом множинного вирівнювання (multiple alignment) [13]. При цьому можна виявити окремі консервативні АК-залишки та їх скупчення (мотиви, сигнатури), які зберігаються в ході дивергентної еволюції генів цих білків. Як правило, це свідчить про важливість таких залишків для структури і функцій білка. Аналіз групи гомологічних білків дозволяє виявити схожі ділянки поліпептидного ланцюга, які відповідають структурно консервативним елементам просторової структури (SCR), та варіабельні ділянки, що відповідають структурним елементам з різною структурою (SVR). Множинне вирівнювання можна зробити за допомогою серверів *Clustal W*[14], *MultAlin* або безпосередньо сервером *PredictProtein*. Можна використовувати також локальні редактори множинного вирівнювання, наприклад, *GeneDoc*.

Пошук мотивів та передбачення функцій білків. Білкові мотиви — відносно короткі схожі АК-послідовності, які зустрічаються у різних біл-

ках. Як правило, вони відповідають найбільш консервативним ділянкам активних центрів ферментів, особливо важливим залишкам для утворення потрібної просторової структури, сайтам модифікації (наприклад, фосфорилування або глікозилування) та ін. Існують чисельні програми і банки даних, які виявляють ці мотиви та їх комбінації, і завдяки цьому передбачають функцію і властивості білків, виходячи з їх АК-последовностей. Для пошуку структурних та функціональних мотивів використовують *ProSite*, *PRINTS*, *Motif*, *NetPhos*, *NetOglyc* та ін.

До цієї ж групи програм можна приєднати також пошук сигнальних пептидів у секреторних білках (*SignalP*), трансмембранних ділянок (*TopPred*, *DAS*, *TMHMM*), передбачення сайтів протеолізу (*PESTFind*), локалізації в клітині (*PSORT*) та ін.

Філогенетичний аналіз АК-последовностей призначений для аналізу ступеня їх схожості і можливого еволюційного походження, а також для їх наочної візуалізації у вигляді різноманітних дерев, діаграм тощо. Філогенетичний аналіз тісно пов'язаний з алгоритмами множинного вирівнювання і поділяється на різні за призначенням і алгоритмами напрямки (*distance matrix methods*, *maximum likelihood methods*, *parsimony programs* та ін.) [15]. Знання поширеності білкових доменів серед різних білків та організмів різних філогенетичних груп і наприклад, еубактерій, архебактерій і еукаріотів) дозволяє оцінити час їх виникнення, функціональну роль та важливість для існування організмів. Слід підкреслити, що результати філогенетичного аналізу відчутно залежать від вибору методів і вихідних параметрів програм. Великий пакет програм *Phylip* є одним з найуживаніших і цитованих.

Ідентифікація границь доменів. Для більшості білків характерна модульна (мультидоменна) будова з окремих доменів, яким притаманні незалежні походження і структура [16–19]. Аналіз доменної будови білків (кількості і границь доменів) проводиться з банками даних доменів *ProDom*, *Pfam*, *HSSP*, *SBASE*, *ProtFam*, *BLOCKS*.

Передбачення вторинної структури і експонованості залишків. Важливу інформацію щодо структури білка дає передбачення його вторинної структури та рівня експонованості АК-залишків [20]. Таке передбачення допомагає ідентифікувати окремі елементи структури (α -спіралі, β -тяжі, повороти поліпептидного ланцюга, ділянки петель тощо), а потім і тип згортки окремих доменів, а також границі доменів. Якість передбачення значно залежить від кількості і рівня гомології різних последовностей у множинному вирівнюванні, з яким працюють програми передбачення вторинної структури. Прикладом таких програм є програма *PHD* [21], точність якої становить понад 72 % для розчинних глобулярних білків. Подібними за при-

значенням є програми *PSSP*, *PROF*, *SSpro*, *Predator*, *JPRED* та ін.

4. Аналіз на рівні просторової структури

Банки даних просторових структур. Дані про експериментально встановлені 3D-структури білків методами PCA та ЯМР депонуються і зберігаються у банках даних просторових структур. Найважливішим з них є *PDB* (*Protein Data Bank*) [4], — головний всесвітній депозитарій просторових структур. На сьогодні *PDB* містить понад 15 тис. встановлених структур різних білків [4]. Кожній структурі відповідає окремий файл з просторовими (декартівськими) координатами окремих атомів. В анотації до *PDB*-файлів описуються важливі структурні дані. Різні структури в цих банках даних можуть бути визначені з різною роздільною здатністю, яка традиційно вимірюється в ангстремах (Å). Діапазон для більшості білкових структур становить 1,5–2,5 Å, при якому можна розрізнити всі неводневі атоми (за винятком особливо рухливих АК-залишків і часто N- та C-кінців поліпептидного ланцюга).

До споріднених банків даних належать *MMDB*, *3Dee*, *FSSP* та ін., які є загальнодоступними через Internet. Існує багато форматів таких координатних файлів, які розпізнаються різними програмами. Стандартом *de facto* для білкових структур є формат банку даних *PDB* (файли типів *.pdb та *.ent). Існують зручні програми з конвертації таких файлів у потрібний формат, наприклад, *BabelWin*.

Визначення типу згортки доменів. Дані про структуру окремих білкових доменів систематизуються у спеціалізованих банках даних у вигляді певних ієрархічних систем (наприклад, Class — Fold — Superfamily — Family — Domain — Organism Variant). Визначають близько п'яти основних структурних класів доменів (переважно α -спіральні, β -складчасті, $\alpha + \beta$, α / β та "не структуровані"). Налічується кілька сотень різних типів просторових згорток (folds) [22]. До найбільш цитованих банків даних належать *SCOP* (*Structural Classification of Proteins*) [23], *CATH* [24], *CE*, які мають власні розвинені пошукові системи та програми.

Візуалізація та рендерінг просторових структур білків. Для візуалізації та зміни типу зображення (рендерінгу) просторових структур використовують програми-вьювери. Такі програми дозволяють проводити виділення за вибраними параметрами (електростатичний потенціал, тип атомів, елементи вторинної структури, окремі АК-залишки, доступна для розчинника поверхня, кути та довжини зв'язків, водневі зв'язки, Ван-дер-Ваальсові радіуси та ін.) Також можна вибрати різні типи графічного зображення молекули (wireframe, back-

bone, spacefill, ribbons, strands, ball-and-stick тощо). Програма *WebLab* — це простий, але потужний вьювер для перегляду та редагування координатних файлів різних форматів. Програма *Swiss-PDB viewer* [25] містить також функції різних необхідних підпрограм-аналізаторів, а також є інтегрованою з сервером *Swiss-Model*, що полегшує роботу з цим сервером. *Swiss-PDB viewer* дозволяє аналізувати одночасно декілька білків завдяки їх вирівнюванню та суперпозиції, а також виконувати "комп'ютерний мутагенез", розраховувати кути, довжини зв'язків та водневі зв'язки, зчитувати та будувати карти електронної густини та ін. Програма *MolMol* [26] дозволяє робити суперпозицію молекул із розрахуванням їх середньоквадратичного відхилення (RMSD), виводити карти Рамачандрана тощо. До поширених вживаних вьюверів можна також віднести *Protein Explorer* і *RasMol* [27]. До вьюверів належать допоміжні програми (plug-ins) *Chime* і *Cn3D*, які доповнюють можливості браузерів *Internet Explorer* і *Netscape Communicator* в on-line режимі.

Зручно використовувати Windows-програму загального призначення *ACDSee* для перегляду "фіксованих" зображень (слайдів), записаних у різних форматах, які легко генеруються програмами-вьюверами. Результатами роботи Linux-програм є, як правило, окремі координатні файли чи їх сукупності. Такі файли можна аналізувати у звичайних вьюверах, які керуються ОС Windows, що було розглянуто вище.

5. Моделювання структури білків

Теоретичні основи моделювання структури білків. Встановлення АК-последовностей та просторових структур великої кількості білків викликало підвищений інтерес до вивчення природи сил, що стабілізують їх нативну конформацію. Було сформульовано спрощені вихідні принципи організації структури білків [28]: 1) АК-последовність однозначно визначає нативну конформацію білка; 2) згортання поліпептидного ланцюга в нативну структуру здійснюється спонтанно, прямо і неопосередковано; 3) стабільна конформація білка відповідає мінімальній вільній енергії Гіббса. Нині вже зрозуміло, що жоден з цих принципів не є абсолютно правильним: для багатьох встановлених структур нативна конформація відповідає тільки одному з локальних мінімумів вільної енергії. Але доповнення скоріше уточнюють, а не скасовують ці принципи.

Природна (еволюційно відібрана) АК-последовність набуває за фізіологічних умов досить визначеної, унікальної конформації, яка детермінує біологічну функцію даного білка. Це має місце незважаючи на значну конформаційну свободу багатьох АК-залишків. Білкову глобулу формують як

ковалентні, так і слабкі нековалентні взаємодії — водневі зв'язки, гідрофобні, електростатичні та Ван-дер-Ваальсові взаємодії. Бічні радикали АК у молекулах білків визначають унікальні властивості нативної конформації білка.

Класифікація методів передбачення структури білків. Ці методи спрощено можна розділити на три класи за повнотою фізичного підходу до їх опису [28]: 1) квантово-механічні методи розрахунку *ab initio*, які базуються на вихідних фізичних принципах та враховують усі електрони атомів (дедуктивні методи); 2) напівемпіричні, які враховують взаємодію валентних електронів атомів; 3) методи класичної молекулярної механіки (емпіричні індуктивні методи), які не враховують електрони, але використовують статистично встановлені правила у вигляді певних силових полів. Теоретично можливо, виходячи з базових фізичних принципів, розраховувати структуру макромолекул без додаткових емпіричних даних, що становить напрямок "строного" моделювання конформації білків на потужних суперкомп'ютерах. Напівемпіричні методи характеризуються різними підходами (MNDO, CNDO тощо), що базуються на різних типах врахування взаємодії валентних електронів.

Принцип методу моделювання за гомологією. Для передбачення просторової конформації молекул білків доцільно використовувати інформацію про білки, які мають тривимірну структуру [29—38]. За такими емпіричними методами спочатку базується жорсткий каркас пептидних зв'язків, в який потім вписуються бічні радикали АК-залишків. Для побудови структурної моделі білка, як правило, використовують найефективніший серед емпіричних методів — метод моделювання за гомологією (порівняльне моделювання), який ґрунтується на фундаментальному принципі залежності між рівнем гомології АК-последовностей білків та схожістю їх просторової структури [39—45].

Встановлено, що білки, які мають гомологічні АК-последовності (споріднені за первинною структурою), також мають подібну просторову структуру [28]. Чим вищий рівень гомології, тим більш схожі їх тривимірні конформації. Існують приклади, коли білки виявляють значну схожість у просторових структурах без очевидної гомології АК-последовностей. Це можна пояснити тим, що може існувати обмежена кількість різних типів згорток білків (приблизно 1 тисяча), тому навіть негомологічні білки можуть мати однаковий тип згортки (конвергентна еволюція). З другого боку, невідомі приклади гомології АК-последовностей без подібності їх просторових структур (окрім дуже коротких пептидів) [40]. Отже, просторова структура білків постійніша за їх АК-последовність упродовж еволюційного процесу. І це зрозуміло, бо для природного добору, який діє на рівні цілого організму,

насамперед важлива просторова структура молекули білка, що визначає його функцію, і завдяки цьому є більш консервативною, ніж АК-последовність білка. Якщо для білка, що вивчається, існує гомологічний білок, просторова структура якого вже експериментально встановлена, то атомні координати такого білка можна використати в якості просторової матриці для моделювання за гомологією. Последовність білка, який моделюється, "накладається" на матрицю. При цьому інформація про координати атомів переноситься на білок, що моделюється. За допомогою "комп'ютерного мутагенезу" бічні радикали АК-залишків матриці замінюються потрібними, і таким чином отримують початкову структуру з неоптимізованою геометрією. Експоненційне зростання експериментально встановлених структур сприяє поширенню і розвитку методу моделювання за гомологією.

Моделювання за гомологією включає наступні етапи: 1) пошук у банках даних білків-матриць з експериментально визначеною просторовою структурою, які мають високий ступінь гомології з АК-последовністю білка-мішені, для якого будується структурна модель; 2) вирівнювання последовності, яку необхідно змоделювати, з однією чи декількома последовностями-матрицями. Якщо ідентичність досліджуваної последовності порівняно з матрицями понад 40 %, то якість вирівнювання визначає якість моделі; 3) корекція вирівнювання; 4) генерація ковалентно-неперервного ланцюга (каркаса) моделі на основі вирівнювання. Це може бути каркас найбільш гомологічної матриці або гібридний з кількох матриць; 5) генерація "канонічних" поверхневих петель, отриманих з банків даних. Петлі визначають переважно конформаційні відмінності гомологічних білків; 6) "вбудова" бічних радикалів у каркас та їх оптимізація. Консервативні залишки мають однакові конформації; 7) побудова петель *ab initio*, якщо в них є інсерції або делеції; 8) мінімізація вільної енергії всієї моделі, іноді з використанням молекулярної динаміки; 9) перевірка моделі вибірково повторенням попередніх етапів.

Чим коротшою є досліджувана последовність, тим більша схожість її з матрицями необхідна для отримання структурно значимої моделі. Структури, визначені методом PCA, мають перевагу над ЯМР-структурами в якості структурних матриць, оскільки вони відповідають, як правило, більш стабільній конформації.

Зрозуміло, що чим вище рівень гомології, тим краще якість моделі. Модель структури білка, побудована на основі матриці з гомологією понад 90 %, має такі ж незначні похибки, як і кристалографічно визначені структури. Коли рівень гомології з матрицею 75 % або більше, вставки (інсерції) петель у більшості випадків необхідні нечасто, а

якщо вони й потрібні, то є дуже короткими. Якщо матриця має 50 % ідентичності з досліджуваною последовністю, середньоквадратичне відхилення (RMSD) становить приблизно 1,5 Å зі значно більшими локальними похибками.

Ділянки петель. Моделювання петель, яким зазвичай відповідають ділянки SVR, часто проводиться з використанням структурної інформації про структуру петель інших білків. Кінцеві залишки таких петлевих фрагментів мають збігатися з краями делецій у моделі (між ділянками SCR). На наступному етапі розраховується *ab initio* конформація петлі для моделі. Інформація про ділянки петель зберігається у спеціалізованих банках даних *Loop*, *LoopDB*.

Сервери для моделювання за гомологією. Моделювання за гомологією можна безпосередньо проводити в Internet за допомогою моделюючих систем *Swiss-Model*, *Modeller*, *CPHmodels*, *SDSC1*, *FAMS*, *3D-JIGSAW* та ін. За допомогою цих серверів можна автоматично провести пошук білків-матриць, саме моделювання, оптимізацію геометрії і перевірку якості отриманої моделі. Деякі сервери надають повний контроль за процесом побудови моделі (крок за кроком). Результати (координатні файли моделей та протокол їхньої побудови) надсилаються електронною поштою. На думку авторів, сервер *Swiss-Model* [46, 47] є найбільш розвиненим, коректним і зручним у роботі. Якщо при пошуку будуть знайдені структури з рівнем гомології не нижче 25–30 % для повної последовності або окремого домену, то використання *Swiss-Model* сервера дозволить провести весь цикл моделювання.

Моделювання за схожістю (threading). Методи структурного передбачення — це більш складні і менш надійні методи, які використовують за відсутності гомологічних последовностей з достатнім рівнем гомології [40, 48–52]. Ідентифікується тип згортки та аранжування елементів вторинної структури, і на цій основі використовують схожі (з неявною гомологією чи негомологічні) структури як матриці для окремих фрагментів поліпептидного ланцюга. Для цього призначені такі сервери: *3D-PSSM* [53], *UCLA-DOE*, *FRSVR*, *Loopp*, *Sausage*, *Superfamily*.

Оптимізація структури (мінімізація вільної енергії). Отримана попередня модель просторової структури білка (промодель) має бути оптимізована, тобто має бути знайдений локальний чи глобальний мінімум енергії даної структури. Мінімізація енергії є оціночною процедурою, призначеною для пошуку найбільш імовірної конформації молекули, яка має найменшу вільну енергію. Ця процедура лежить в основі комп'ютерного уточнення структур білків, визначених методом PCA, а також моделювання фолдингу макромолекули білка.

На практиці робиться припущення, що мінімум на поверхні потенціальної енергії відповідає мінімуму на поверхні вільної енергії. Часто обчислюють і порівнюють значення вільної енергії в усіх локальних мінімумах. Різні методи розрахунку вільної енергії свідчать на користь висновку, що "розмитий" мінімум потенціальної енергії може відповідати більш низькому мінімуму вільної енергії. Різниця значень потенціальних енергій обох мінімумів не повинна бути надто великою.

Для оцінки фазового простору структурної моделі (зменшення вільної енергії молекули білка) використовують три типи методів: не емпіричні (*ab initio*), напівемпіричні та емпіричні. Сучасні потужні програми можуть враховувати квантово-механічні ефекти у великих молекулярних системах там, де це необхідно (розрив та утворення ковалентних зв'язків тощо), але для моделювання структури білків у більшості програм обмежуються емпіричними методами молекулярної механіки і не враховують квантово-механічні ефекти.

Силкові поля. До емпіричних методів оцінки атом-атомних і торсійних потенціалів належать різні комбінації силових взаємодій (електростатичних або кулонівських, Ван-дер-Ваальсових і торсійних). Силковим полем називається фізична модель опису молекули аналітичними функціями для енергії взаємодії між всіма парами атомів різних типів, які взаємодіють між собою, або для сукупності всіх торсійних кутів у макромолекулі. Ядра атомів розглядаються як ньютонівські частки, що перебувають під впливом функції потенціальної енергії або силового поля. Як правило, довжини ковалентних зв'язків та планарних кутів вважаються незмінними. В емпіричних методах (методах молекулярної механіки) рівняння класичної механіки використовуються для опису потенціальної енергії та фізичних властивостей молекули. Різниця між цими методами полягає у врахуванні Ван-дер-Ваальсових сил та у використанні емпіричних коефіцієнтів (наприклад, для розрахунків близько- та далекодіючих взаємодій), які наближують модельні значення до реальних. Використовують такі силкові поля, що дозволяють потенціальну енергію молекулярних систем розраховувати як функцію координат їхніх атомних ядер, а саме: *MM+*, *AMBER*, *OPLS*, *GROMOS*, *CHARMM*, *ECEPP*, *BIO+*, *GROMACS* та ін.

Метод *MM+* є найбільш загальним методом молекулярно-механічних розрахунків, спеціально розробленим для органічних молекул як розширений варіант методу *MM2(tm)*. Метод *AMBER* дозволяє за банками даних розраховувати суперпозицію кожного атома, враховуючи Ван-дер-Ваальсові взаємодії. Цей метод завдяки підвищеній чутливості до Ван-дер-Ваальсових взаємодій, найчастіше застосовується для оптимізації пептидних лан-

цюгів. Метод *OPLS* містить інформацію лише про певні групи атомів, враховує Ван-дер-Ваальсові сили, а також є більш чутливим щодо електростатичних взаємодій. Тому метод *OPLS* частіше використовують для аналізу макромолекулярних комплексів і передбачення їх утворення.

Математичні методи (алгоритми) для мінімізації функції потенціальної енергії (оптимізації геометрії структури) можна розділити на три групи: градієнтні, не градієнтні та генетичні алгоритми. У більшості програм, з якими доводилось працювати авторам, використовуються градієнтні методи. За допомогою цих методів шукають екстремуми на поверхні функції конформаційного простору (аналізують кривизну енергетичної поверхні) шляхом обчислювання значень перших і/або других похідних за конформаційними змінними. Градієнтні методи можна порівняти з "поведінкою м'яча, що котиться вниз по складній поверхні з пагорбами і ущелинами". Ознаками градієнтних методів є якість і швидкість збіжності процесу обчислювання, але вони мають недолік: пошук припиняється в кожному з локальних мінімумів, коли функція поводить себе недостатньо просто (а це типова ситуація). До градієнтних методів належать методи найшвидшого спуску (*Steepest Descent*), *Pollak—Ribiere*, *Newton—Raphson*, *Fletcher—Reeves*, *Davidone*, *Fletcher—Pauell*, *Eigenvector—Follow* та ін.

До не градієнтних методів мінімізації належать симплекс-методи, які в цілому є дуже надійними, але вимагають високих витрат машинного часу для отримання результатів. Симплекс-метод можна порівняти "з поведінкою сліпого восьминога, який намащує свою нору в кораловому рифі". Цей метод працює за наявності точок розриву в похідних чи у самій функції, а також дозволяє подолати неглибокі локальні мінімуми, що збільшує імовірність виявлення глобального мінімуму. Імовірність "стрибків" до іншого мінімуму залежить від глибини і профілю мінімуму, який зустрівся, ширини енергетичного бар'єру і від значень керуючих параметрів симплекс-процедури. Доцільність об'єднання різних методів реалізована при використанні подвійної мінімізації, коли градієнтний метод переключається на симплекс, якщо відбувається зупинка в локальному мінімумі чи зустрічається складна точка. І навпаки, відбувається повернення до градієнтного методу при виявленні нового мінімуму. Для "виходу" з локальних мінімумів енергії часто використовують метод *Monte Carlo*, який дозволяє знаходити глобальний мінімум.

Комп'ютерні програми для мінімізації енергії. У доступних комп'ютерних програмах моделювання білків (*What If* [54], *Quanta*, *Hydra*, *Insight*, *O*, *BRAGI* та ін.) для проведення процедур оптимізації структури білків здебільшого використовують

ються емпіричні методи. Більшість таких програм дозволяють користувачу самому вибрати вид силового поля і математичний алгоритм.

Програма *HyperChem* [55] — це система для молекулярного аналізу, моделювання та оптимізації, яка має в наявності кілька згаданих вище методів молекулярної механіки, і на основі різних математичних алгоритмів намагається знизити енергію молекулярної системи за рахунок упорядкування її геометрії. *HyperChem* може рахувати та мінімізувати вільну енергію моделей багатьма методами, включаючи молекулярну механіку, метод Monte Carlo, напівемпіричне та *ab initio*-моделювання для відносно простих молекул. Програма *HyperChem* має зручний інтуїтивно-зрозумілий інтерфейс і широко застосовується у комп'ютерній хімії в Україні.

Для мінімізації вільної енергії використовують також програму *Gromos*. Спочатку видаляються найбільші помилки (надто зближені атоми, напружена конформація водневих зв'язків, неправильні кути в каркасі). Однак на кожному кроці в модель додається багато дуже малих помилок, і тому на початку роботи програми виправлення великих помилок переважає кількість введених малих, але згодом велика кількість кроків збільшує кількість помилок і це знову псує модель.

Оцінка якості отриманих моделей. Перевірка відповідності моделі полягає у виявленні відсутності якихось неприпустимих значень структурних параметрів (довжин ковалентних зв'язків, двограних та торсійних кутів та ін.) [56—61]. Повний аналіз зручно виконувати на серверах, призначених для цього: *PROCHECK*, *Verify3D*, *WHAT_CHECK*, *EVA*, *Prove* та ін.

Загальний аналіз властивостей просторової структури. Методи аналізу можна використовувати як для експериментально визначених, так і змодельованих структур білків. До типових задач аналізу належать: обчислення геометричних параметрів (розмірів, об'єму, площини поверхні, відкритої для розчинника, поверхневого заряду та ін.), встановлення оптимальної сітки водневих зв'язків [62], електростатичних взаємодій (сольові містки), дисульфідних зв'язків, встановлення гідрофобних та гідрофільних властивостей поверхні, візуалізація форми поверхні і виявлення щілин, занурень (западин) або експонованих елементів структури, які часто дуже важливі для функцій білків. Проводиться локалізація консервативних залишків, визначених на попередніх етапах, відносно просторової структури моделі. Особлива увага приділяється поверхневим (експонованим) консервативним залишкам, які можуть відповідати за взаємодію з іншими молекулами

Порівняння (суперпозиція) просторових структур білків. Проводиться накладання (супер-

позиція) двох чи кількох структур та розраховується середньоквадратичне відхилення (RMSD) у положенні відповідних атомів білка-матриці і моделі гомологічного білка або різних білків. Це можуть бути лише Ca атоми АК-залишків, усі атоми каркасу білка чи всі неводневі атоми. Тому слід звертати увагу на те, якого типу суперпозицію проведено і на якій довжині поліпептидного ланцюга. Для суперпозиції білків можна використовувати вьювер *Swiss-PDB viewer* або сервери *VAST ma Lock*.

Аналіз поверхневих властивостей молекул білків. Поверхневі АК-залишки молекул білків є найбільш цікавими з точки зору функціонування білків, в той час як занурені у білкову глобулу залишки визначають переважно його структуру. Існує багато різноманітних програм аналізу експонованості залишків та поверхні (*KeyResidues*, *Access*, *server* *What If*), виявлення западин, кишень та порожнин на поверхні білків (*castP*, *PASS* та ін.). Дуже важливо також аналізувати зони контактів (інтерфейси) між окремими доменами або олігомерами на рівні четвертинної структури білка. Такі інтерфейси захищені від розчинника поверхнями взаємодії макромолекул. Це можна робити серверами *Protein-Protein Interactions (PPI)*, *Protein-Nucleic acid Interactions (PNI)*, *STING*, програмами *ProteinExplorer* та ін.

Врахування водного оточення. У реальних умовах білки перебувають у водному оточенні (сольватовані білки) або занурені у мембранні структури чи у великі агрегати, тому потрібно розглядати вплив молекул води на макромолекули. Молекули води часто фіксуються у "кишенях" на поверхні білка і відіграють важливу роль у взаємодії з лігандами, утворюючи водяні містки. Дуже важливим є також виявлення зв'язаних молекул води, які оточують молекулу білка (ім відповідають молекули води, визначені методом РСА). Для того, щоб отримати найбільш реальну конформацію поліпептидного ланцюга, необхідно симулювати присутність молекул води, тобто змодельувати середовище, наближене до природного середовища існування білкової молекули.

6. Моделювання просторової структури СООН-кінцевого цитокінподібного модуля цитоплазматичної тирозил-тРНК синтетази ссавців

Одним з прикладів моделювання просторової структури білка за гомологією може бути побудова моделі некаталітичного СООН-кінцевого цитокінподібного модуля цитоплазматичної тирозил-тРНК синтетази ссавців [63]. С-модуль цитоплазматичних тирозил-тРНК синтетаз гомологічний до цитокіну ЕМАР II ссавців (52,7 % ідентичності) і у вільному стані має схожі біологічні властивості в експериментах *in vitro*. З використанням методів по-

рівняльного моделювання за гомологією створена модель просторової структури С-модуля тирозил-тРНК синтетази бика (*Bos taurus*). Головною частиною С-модуля є домен з типом просторової укладки "OB-fold" (В-домен, залишки Val363-Lys470), який вміщує п'ять р-складок, характерних для цього типу укладки. Подібно до ряду "OB-fold"-доменів С-модуль має спорідненість до нуклеїнових кислот. Порівняння моделі С-модуля з антикодон-зв'язувальними доменами аспартил- і лізил-тРНК синтетаз дозволило передбачити поверхню взаємодії з РНК і функціонально важливі амінокислотні залишки. Структурна модель С-модуля тирозил-тРНК синтетази є основою для аналізу подвійної функції цього домену: структурно-специфічної взаємодії з тРНК, а також цитокінової функції.

7. Деякі наступні напрями досліджень

Побудову структурної моделі білка можна розглядати лише як початок подальших експериментів *in silico*, які можуть бути виконані. Моделювання структури білків, аналіз їх властивостей і порівняння з відомими структурами становить лише певний етап на шляху до розуміння біологічних властивостей білка. До наступних етапів можна віднести аналіз динамічних властивостей молекул білків, зміни конформаційних станів (конформерів), а також передбачення потенційних взаємодій з іншими молекулами. Це має бути темою іншого огляду, але завдяки важливості цієї інформації для повноти картини ми зупиняємось на деяких окремих моментах.

Молекулярна динаміка. Кристалографічна структура білка — це фактично статична фотографія макромолекули, її морфологія. На основі статичної структури може створитися хибне уявлення про жорсткість і незмінність структури молекул білків. Динамічні моделі функціонування нині обчислені для багатьох білків. Такі дані зберігаються на спеціалізованих серверах, наприклад, *Morpher*, *MoIMovDB*, *ChemVis* тощо. Аналіз молекулярної динаміки полягає у перебиранні дозволених конформацій в рамках певного енергетичного мініму-

му. Це допомагає оцінити діапазон можливих змін структури білка. Програми з молекулярної динаміки при роботі на сучасних персональних комп'ютерах все ще обмежені піко- та наносекундним часовим діапазоном. Визначення великих конформаційних змін білків у мілісекундному діапазоні, які мають місце при русі окремих доменів або у великих макромолекулярних ансамблів, потребує тривалих розрахунків. Молекулярну динаміку можна розраховувати за допомогою програм *Gromacs*, *Moldy* та ін. Отримані таким чином результати (дуже великі так звані файли траєкторій) можна візуалізувати у спеціалізованих вьюверах, таких як *VMD*.

Проблема згортання (фолдінгу) білків. Проблема згортання поліпептидного ланцюга білків є фундаментальною проблемою молекулярної біології [64, 65]. Оскільки згортання поліпептидного ланцюга є динамічним процесом, то найбільш фізично адекватним підходом до цієї проблеми є симуляція рухів усіх атомів білкової молекули (молекулярна динаміка). Молекулярна динаміка дозволяє також відобразити глобальні конформаційні зміни молекули за певний проміжок часу, локальні структурні флуктуації та ін.

Передбачення взаємодій з мікро- і макролігандами. Якщо доступні експериментальні дані про субодичну будову білка (четвертинна структура), то можна передбачити механізм олігомеризації (найчастіше димеризації) білка. Аналізуються як стерична відповідність поверхонь молекул, так і різні види взаємодії (електростатичні і гідрофобні взаємодії, водневі зв'язки та ін.) Багато сигнальних і регуляторних шляхів опосередковуються взаємодією між окремими доменами, які відповідають за взаємодії між білками [66] та між білками і нуклеїновими кислотами [67, 68]. Можна також аналізувати взаємодії молекули білка і малої молекули (лігандом, субстратом, інгібітором), що дає змогу здійснювати дизайн нових лікарських препаратів (drug design) [69, 70]. Для цього використовують програми докінгу: *HEX*, *FTDock*, *AutoDock*, *ICM-Dock*, *DOCK* та ін.

Таблиця 1. Сервери і програми Internet з аналізу і передбачення структури білків. (Сервери і програми поєднані у розділи, які наведені у послідовності, що відбиває типовий порядок їх використання)

Сервер/програма	Web-адреса	Сервер/програма	Web-адреса
Аналіз і передбачення структури білків			
Перелік курсів	www.rcsb.org/pdb/education.html	<i>META server</i>	dodo.cpmc.columbia.edu/predictprotein/submit_meta.html
Люстрований курс	molmod.angis.org.au/~shoba/asbmb/	<i>PNaRM</i>	www.temple.edu/medbiochem/pnarm.html
Курс <i>BioVM631</i>	www.mbg.cornell.edu/nicholson/biohm631/	<i>Pedro's Tools</i>	www.public.iastate.edu/~pedro/rt_1.html
Переліки посилань на сервери і програми			
<i>Bio Links</i>	indigol.biop.ox.ac.uk/biolinks.html	Сервери загального призначення	
		<i>NCBI</i>	www.ncbi.nlm.nih.gov/
		<i>EB/</i>	www.ebi.ac.uk

ExpASY www.expasy.ch/

BioPlanet www.bioplplanet.com/1inks.htm

BioNavigator www.bionavigator.com

Банки даних послідовностей

EMBL www.embl-heidelberg.de/Services/index.html

GenPept www.ncbi.nlm.nih.gov/protein/

TrEMBL www.hgmp.mrc.ac.uk/Bioinformatics/Databases/tremblhelp.ivtmi

Swiss-Prot www.expasy.ch/sprot/sprot_details.html

Пошук і вирівнювання гомологічних послідовностей

BLAST www.ncbi.nlm.nih.gov/BLAST

Fasta bioweb.pasteur.fr/seqanal/interfaces/fasta.html

Blitz (e-mail) blitz@embl-heidelberg.de

Clustal pfwww2.ebi.ac.uk/clustalw/

MultAlin protein.toulouse.inra.fr/multalin.html

PredictProtein www.embl-heidelberg.de/predictprotein/predictprotein.html

MSA alfredo.wustl.edu/msa.html

GeneDoc www.psc.edu/biomed/genedoc/

Пошук мотивів та сигналів, передбачення локалізації

ProSite www.expasy.ch/prosite/

PRINTS bioinf.man.ac.uk/dbbrowser/PRINTS/

TMHMM www.cbs.dtu.dk/services/TMHMM-1.0/

PESTFind www.at.embnct.org/embnct/tools/bio/PESTfind/

PSORT psort.nibb.ac.jp/

Філогенетичний аналіз послідовностей білків

Phylog evolution.genetics.washington.edu/

BWP www.lms.si.edu/PAUP/about.html

Аналіз і передбачення доменної організації білків

ProDom www.toulouse.inra.fr/prodom.html

CD-Search www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml

Sbase base.icgeb.trieste.it/sbase/

BLOCKS www.blocks.fhcrc.org/blocks_search.html

Pfam www.sanger.ac.uk/Pfam/

ProtFam www.mips.biochem.mpg.de/desc/protfam/

HSSP www.sander.ebi.ac.uk/hssp/

Передбачення вторинної структури білків

PHD dodo.cpmc.columbia.edu/predictprotein

GOR molbiol.soton.ac.uk/compute/GOR.html

PSSP dot.imgen.bcm.tmc.edu:9331/pssp/prediction/pssp.html

Predator www.embl-heidelberg.de/argos/predator/predator_form.html

DSSP www.cmbi.kun.nl/swinVdssp/

SOPM www.ibcp.fr/predict.html

Sspred www.embl-heidelberg.de/sspred/sspred_info.html

Банки даних просторових структур

PDB www.rcsb.org/pdb

RCSB www.rcsb.org/pdb/index.html

MMDB www.ncbi.nlm.nih.gov:80/Structure/MMDB/mddb.shtml

3Dee jura.ebi.ac.uk:8080/3Dee/help/help-intro.html

OWL www.leeds.ac.uk/bmb/owl/owl.html

FSSP www.ebi.ac.uk/dali/fssp/fssp.html

3DCrunch www.expasy.ch/swissmod/SM_3DCrunch.html

Зміна формату координатних файлів

BabelWin mercury.ichem.arizona.edu/babel.html

Ідентифікація і аналіз типу згортки доменів

SCOP scop.mrc-lmb.cam.ac.uk/scop/data/scop

CATH www.biochem.ucl.ac.uk/bsm/cath

CE ci.sdsc.edu/ce.html

Dali server www2.ebi.ac.uk/dali/

Побудова діаграм вторинної структури

TOPS www3.ebi.ac.uk/tops/AtlasHTML.html

Програми для візуалізації макромолекул

Swiss-PDB viewer www.expasy.ch/spdbv/mainpage.html

WebLab www.msi.com

RasMol www.bernstein-plus-sons.com/software/rasmol

www.umass.edu/microbio/rasmol

Chime www.mdli.com/support/chime/default.html

Cn3D www.ncbi.nlm.nih.gov/Structure/cn3d.html

POV-Ray www.povray.org/

Системи моделювання просторової структури білків

Swiss-Model www.expasy.ch/swissmod/SWISS-MODEL.html

META server dodo.cpmc.columbia.edu/predictprotein/submit_meta.html

CPH Model www.cbs.dtu.dk/services/CPHmodels/

What [//www.cmbi.kun.nl:1100/WIWWWI](http://www.cmbi.kun.nl:1100/WIWWWI)

Modeller guitar.rockefeller.edu/modeller/modeller.html

SCULPT www.productreview.net.com/abstracts/1/1322.htm

Пошук подібних структур за методом "threading"

UCLA-DOE www.fold.doe-mbi.ucla.edu/

3D-PSSM www.bmm.icnet.uk/~3dpssm/

Loopp ser-joopp.tc.cornell.edu/loopp.html

Мінімізація вільної енергії

HyperChem www.hyper.com

InsightII www.chassels.net/organic/InsightII/index.html

Порівняння просторових структур (суперпозиція) білків

VAST www.ncbi.vastrv.com

Lock gene.stanford.edu/lock/

3DPAI www.elcomsoft.com/3dplan.html

Аналіз поверхневих властивостей і ділянок взаємодії

PPI www.biochem.ucl.ac.uk/bsm/PP/server/

PNI www.biochem.ucl.ac.uk/bsm/DNA/server/

MolSurfer www.embl-heidelberg.de/lgabdoell/ads/imap/

castP cast.engr.uic.edu/cgi-bin/cast/index.pl

KeyResidues trantor.bioc.columbia.edu/~bissan/keyres/opening.html

CSE/bioinfo weizmann.ac.il:8500/oca-bin/

ACCESS www.csb.yale.edu/userguides/datamanip/access/access_descrip.html

Динаміка білкових макромолекул

Morphing server bioinfo.mbb.yale.edu/MolMovDB/morph/index.cgi

MolMovDB bioinfo.mbb.yale.edu/MolMovDB/

Chem Vis www2.ccc.uni-erlangen.de/projects/ChemVis/

Gromacs rugmdO.chem.rug.nl/~gmx/gmx.html

Moldy www.earth.ox.ac.uk/~keith/molody.html

VMD www.umass.edu/microbio/rasmol/otherof.htm#vmd

Взаємодія білків з мікро- та макролігандами (docking)

HEX www.biochem.abdn.ac.uk/hex/hex_manual.html

FTDock www.bmm.icnet.uk/ftdock/ck/ftdock.html

AutoDock www.scripps.edu/pub/olson-web/doc/autodock/

ICM-Dock www.molsoft.com/icmpages/icmdock.htm

TINKER dasher.wustl.edu/tinker/

HotDock www.uni-paderborn.de/~lst/HotDock/index.html

1. Schwede T., Diemand A., Guex N., Peitsch M. C. Protein structure computing in the genomic era // Res. Microbiol.— 2000.— V. 151.— № 2.— P. 107–112.
2. Marti-Renom M. A., Stuart A. C., Fiser A., Sanchez R., Melo F, Sali A. Comparative protein structure modeling of genes and genomes // Annu. Rev. Biophys. Biomol. Struct.— 2000.— V. 29.— P. 291–325.
3. Sanchez R., Pieper U., Melo F, Eswar N., Marti-Renom M. A., Madhusudhan M. S., Mirkovic N., Sali A. Protein structure modeling for structural genomics // Nat. Struct. Biol.— 2000.— V. 7.— Suppl.— P. 986–990.
4. Berman H. M., Westbrook J., Feng Z., Gilliland G, Bhat T. N., Weissig H., Shindyalov N., Bourne P. E. The Protein Data Bank // Nucl. Acids Res.— 2000.— V. 28.— № 1.— P. 235–242.
5. Peitsch M. C., Schwede T., Guex N. Automated protein modeling the proteome in 3D // Pharmacogenomics.— 2000.— V. 1.— № 3.— P. 257–266.
6. Peitsch M. C., Guex N. Large-scale comparative protein modeling // Proteome Research: new frontiers in functional genomics.— 1997.— P. 177.
7. Ranganathan S. Molecular modeling on the Web // BioTechniques.— 2001.— V. 30.— № 1.— P. 50–52.
8. Ranganathan S. Internet on-ramp // BioTechniques.— 2001.— V. 30.— № 1.— P. 51–52.
9. May A. C., Johnson M. S., Rufino S. D., Wako #., Zhu Z. Y, Sowdhamini R., Srinivasan N., Rodionov M. A., Blundell T. L. The recognition of protein structure and function from sequence: adding value to genome data // Philos. Trans. R. Soc. Lond. B Biol. Sci.— 1994.— V. 344.— № 1310.— P. 373–381.
10. Orengo C. A., Todd A. E., Thornton J. M. From protein structure to function // Curr Opin Struct Biol.— 1999 — V. 9.— № 3 — P. 374–382.
11. Yusupov M. M., Yusupova G. Z., Baucom A., Lieberman K., Earnest T. N., Cate J. H., Noller H. F. Crystal structure of the ribosome at 5.5 Å resolution // Science.— 2001.— V. 292.— № 5518.— P. 883–896.
12. Zhang J., Madden T. L. PowerBLAST: a new network BLAST application for interactive or automated sequence analysis and annotation // Genome Res.— 1997 — V. 7.— № 6.— P. 649–656.
13. Gaskell G. J. Multiple sequence alignment tools on the Web // BioTechniques.— 2000 — V. 29 — № 1.— P. 60–62.
14. Thompson J. D., Higgins D. G., Gibson T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice // Nucl. Acids Res.— 1994.— V. 22.— № 22.— P. 467–1680.
15. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap // Evolution.— 1985.— V. 39.— P. 783–791.
16. Шульц Г., Шурмер Р. Принципы структурной организации белков.— М.: Мир.— 1982.
17. Проблема белка: В 5 т. / В. Т. Иванов (ред.); Т. 3. Структурная организация белка / Е. М. Попов, Т. И. Соркина (ред.).— М.: Наука.— 1997.— 604 с. ил.
18. Соркина Т. А., Залевская И. Н. Структурно-функциональные свойства белков: Учеб. пособие.— К.: Выща шк.— 1989.— 216 с: ил.
19. Овчинников Ю. А. Биоорганическая химия.— М.: Просвещение.— 1987.— 815 с: ил.
20. Морозов В. М. Структурно-функциональный анализ аминокислотных последовательностей через Internet // Молекул. биол.— 1996.— Т. 30.— № 2.— С. 387–393.
21. Rost B., Sander C, Schneider R. PHD — an automatic mail server for protein secondary structure prediction // Comput. Appl. Biosci.— 1994.— V. 10.— № 1.— P. 53–60.
22. Jones D., Thornton J. Protein fold recognition // J. Comput. Aided. Moī. Des.— 1993.— V. 7.— № 4.— P. 439–456.
23. Murzin A. G., Brenner S. E., Hubbard T, Chotia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures // J. Mol. Biol.— 1995.— V. 247.— № 4.— P. 536–540.
24. Orengo C. A., Martin A. M., Hutchinson G, Jones S., Jones D. T, Michie A. D., Swindells M. B., Thornton J. M. Classifying a protein in the CATH database of domain structures // Acta Crystallogr. D Biol. Crystallogr.— 1998 — V. 54.— № 1 (Pt 6).— P. 1155–1167.
25. Guex N., Peitsch M. C. SWISS-MODEL and the Swiss-Pdb-Viewer: An environment for comparative protein modeling // Electrophoresis.— 1997.— V. 18.— № 15.— P. 2714–2723.
26. Koradi R., Billeter M., Wuthrich K. MOLMOL: a program for display and analysis of macromolecular structures // J. Mol. Graph.— 1996.— V. 147— № 1.— P. 51–55.
27. Sayle R. A., Milner-White E. J. RASMOL: biomolecular graphics for all // Trends Biochem. Sci.— 1995.— V. 20.— № 9, — P. 374.
28. Hubbard T. J., Lesk A. M. Modeling protein structures // Computer modeling in molecular biology / Eds. by J. Goodfellow.— VCH Weinheim, 1995.— P. 10–33.
29. Johnson M. S., Srinivasan N., Sowdhamini R., Blundell T. L. Knowledge-based protein modeling // Crit. Rev. Biochem. Mol. Biol.— 1994.— V. 29.— № 1,— P. 1–68.
30. Jones D. T. Progress in protein structure prediction // Curr. Opin. Struct. Biol.— 1997.— V. 7.— № 3.— P. 377–387.
31. Guex N., Diemand A., Peitsch M. C. Protein modelling for all // Trends Bio. Sci.— 1999,— V. 24.— № 9.— P. 364–367.
32. May A. C., Blundell T. L. Automated comparative modelling of protein structures // Curr. Opin. Biotechnol.— 1994.— V. 5.— № 4.— P. 355–360.
33. Russell R. B., Sternberg M. J. Structure prediction. How good are we? // Curr Biol.— 1995.— V. 5.— № 5.— P. 488–490.
34. Sali A., Kuriyan J. Challenges at the frontiers of structural biology // Trends Cell. Biol.— 1999.— V. 9.— № 12.— P. 20–24.
35. Sali A. 100,000 protein structures for the biologist // Nat. Struct. Biol.— 1998.— V. 5.— № 12.— P. 1029–1032.
36. Sali A. Comparative protein modeling by satisfaction of spatial restraints // Mol. Med. Today— 1995.— V. 1.— № 6 — P. 270–277.
37. Sali A., Overington J. P. Derivation of rules for comparative protein modeling from a database of protein structure alignments // Protein Sci.— 1994.— V. 3.— № 9.— P. 1582–1596.
38. Todd A. E., Orengo C. A., Thornton J. M. Evolution of protein function, from a structural perspective // Curr. Opin. Chem. Biol.— 1999.— V. 3.— № 5.— P. 548–556.
39. Hilbert M., Bahm G., Jaenicke R. Structural relationships of homologous proteins as a fundamental principle in homology modeling // Proteins.— 1993.— V. 17.— № 2.— P. 138–151.
40. Sanchez R, Badretdinov A. Y, Feyfant E., Sali A. Homology protein structure modeling // Cur. Opin. Biotechnol.— 1998.— V. 9.— № 4.— P. 383–389.
41. Sanchez R, Sali A. Advances in comparative protein-structure modelling // Curr. Opin. Struct. Biol.— 1997.— V. 7.— № 2.— P. 206–214.
42. Rodriguez R., Chinea G., Lopes N., Pons T, Vriend G. Homology modelling, model and software evaluation: three related resources // Bioinformatics — 1998.— V. 14.— № 6 — P. 523–528.
43. Sali A. Modeling mutations and homologous proteins // Curr. Opin. Biotechnol.— 1995.— V. 6.— № 4.— P. 437–451.
44. Cardozo T, Totrov M. M, Abagyan R. A. Homology modeling by the ICM method // Proteins Struct. Funct. Genet.— 1995.— V. 23.— № 3.— P. 403–414.
45. Peitsch M. C. Protein modeling by email // BioTechnology.— 1995.— V. 13,— P. 658.
46. Peitsch M. C, Guex N. SWISS-MODEL and the SWISS-PDB VIEWER: an environment for comparative protein modelling // Electrophoresis.— 1997.— V. 18.— № 15.— P. 2714–2723.
47. Peitsch M. C. PROMOD and SWISS-MODEL: Internet-based tools for automated comparative protein modelling // Biochem. Soc. Trans.— 1996.— V. 24.— № 1.— P. 274–279.

48. Taylor W. R. Protein structure modelling from remote sequence similarity // J. Biotechnol.— 1994.— V. 35.— № 2–3.— P. 281–291.
49. Taylor W. R. Remotely related sequences and structures: analysis and predictive modelling // Trends Biotechnol.— 1994.— V. 12.— № 5.— P. 154–158.
50. Elofsson A., Fischer D., Rice D. W., he Grand S. M., Eisenberg D. A study of combined structure/sequence profiles // Fold. Des.— 1996.— V. 1.— № 6.— P. 451–461.
51. Claessens M., Van Cutsem E., Tasters I., WodakS. Modelling the polypeptide backbone with 'spare parts' from known protein structures // Protein Eng. — 1989— V. 2.— № 5— P. 335–345.
52. Sternberg M. J., Bates P. A., Kelley L. A., MacCallum R. M. Progress in protein structure prediction: assessment of CASP3 // Curr. Opin. Struct. Biol.—1999.—V. 9.— № 3.— P. 368–373.
53. Kelley L. A., MacCallum R. M., Sternberg M. J. E. Enhanced genome annotation using structural profiles in the program 3D-PSSM // Mol. Biol.— 2000.— V. 29.— № 2.— P. 501–522.
54. Vriend G. WHAT IF: a molecular modelling and drug design program//J. Moï. Graph.— 1990.—V. 8.— № 1.—P. 52–56.
55. Froinowitz M. HyperChem: a software package for computational chemistry and molecular modeling // Biotechniques.— 1993.—V. 14.— № 6.— P. 1010–1013.
56. Hooft R. W. W., Sander C, Vriend G. Verification of protein structures: side-chain planarity // J. Appl. Cryst.— 1996.— V. 29 — P. 714–716.
57. Vriend G., Sander C. Quality control of protein models: directional atomic contact analysis // J. Appl. Cryst.— 1993.— V. 26.— P. 47–60.
58. Hooft R. W. W, Vriend G., Sander C, Abola E. E. Errors in protein structures//Nature.— 1996—V. 381.— № 6580.— P. 272.
59. Hooft R. W., Sander C, Vriend G. Objectively judging the quality of a protein structure from a Ramachandran plot // Comput. Appl. Biosci.— 1997 — V. 13.— № 4.— P. 425–430.
60. Zvelebil M. J., Thornton J. M. Peptide-protein interactions: an overview // Q. Rev. Biophys.— 1993.— V. 26.— № 3.— P. 333–363.
61. Ramachandran G. N., Ramakrishnan C, Sasisekharan V. Stereochemistry of polypeptide chain conformations // J. Mol. Biol.— 1963.— V. 7.— P. 95–99.
62. Hooft R. W. W., Sander C, Vriend G. Positioning hydrogen atoms by optimizing hydrogen bond networks in protein structures // Proteins.— 1996.— V. 26.— № 4.— P. 363–376.
63. Голуб А. Г., Одинец К. А., Ныпорко А. Ю., Корнелюк А. И. Моделирование пространственной структуры СООН-концевого цитокин-подобного модуля цитоплазматической тирозил-тРНК синтазы млекопитающих // Биополимеры и клетка.—2000.—Т. 16.— № 6.— С. 515–524.
64. Dinner A. R, Sali A., Smith L. J., Dobson C. M., Karplus M. Understanding protein folding via free-energy surfaces from theory and experiment // Trends Biochem. Sci.— 2000.— V. 25.— № 7.— P. 331–339.
65. Karplus M., Sali A. Theoretical studies of protein folding and unfolding//Curr. Opin. Struct. Biol.— 1995.—V. 5.— № 1.— P. 58–73.
66. Sternberg M. J., Gabb H. A., Jackson R. M. Predictive docking of protein-protein and protein-DNA complexes // Curr. Opin. Struct. Biol.— 1998.— V. 8.— № 2.— P. 250–256.
67. Luscombe N. M., Austin S. E., Berman H. M., Thornton J. M. An overview of the structures of protein-DNA complexes // Genome Biol.— 2000.— V. 1.— № 1.— P. REVIEWS001.
68. Parkinson G., Voitechovsky J., Clowney L., Bruenger A. T., Berman H. New parameters for the refinement of nucleic acid-containing structures // Acta Crystallogr.— 1996.— V. D52.— P. 57–64.
69. Hornak V, Dvorsky R, Sturdik E. Receptor-ligand interaction and molecular modelling // Gen. Physiol. Biophys.— 1999.— V. 18 — № 3,— P. 231–248.
70. Zeng J. Mini-review: computational structure-based design of inhibitors that target protein surfaces // Comb. Chem. High Throughput. Screen — 2000.— V. 3.— № 5.— P. 355–362.

Odynets K. A., Kornelyuk A. I.

METHODS OF ANALYSIS AND MODELING OF PROTEIN SPATIAL STRUCTURE

In the review the methods of computational analysis and modeling of protein spatial structure are presented. The utilization of Internet servers and local computational programs is described. Different approaches to the analysis of homologous amino acid sequences, vizualization od 3D structures of proteins, homology modeling of 3D structures, minimization of free energy of protein structure, molecular dynamics, macromolecular interactions are discussed. More than 110 Web-addresses and programs are presented in the review.