

УДК 57(075.2)

Біоінформатика як основний інструмент нанобіотехнології та наномедицини

С. В. Горобець, О. Ю. Горобець, Д. О. Дереча
Національний технічний університет України
«Київський політехнічний інститут»

Резюме

В даній роботі розглянуто основні досягнення та сфери застосування біоінформаційних технологій. Вказано на перспективи їх розвитку у взаємозв'язку з суміжними областями знань, такими як нанобіотехнологія, наномедицина та інші.

Ключові слова: біоінформатика, геном, база даних, ДНК, нанобіотехнологія.

Клин. информат. и Телемед.
2008. Т.4. Вып.5. с.41–49

Розвиток біоінформатики

На початку нового тисячоліття відбулась визначна подія в науковому та суспільному житті — була прочитана повна нуклеотидна послідовність генома людини, що стало справжнім тріумфом геноміки. На сьогодні, відомо декілька сотен повних послідовностей ДНК різних біологічних об'єктів починаючи з мікоплазм, вірусів, мікроорганізмів і закінчуючи людиною. Поява швидких методів розшифрування послідовності нуклеотидних пар у геномах різних організмів спричинило за собою появу великого числа генетичних текстів, у вигляді довгих символічних послідовностей. Молекула ДНК, носій генетичної інформації, являє собою біополімер, елементарною ланкою якого є нуклеотидна пара. Число таких пар у ДНК *E.coli* — 5млн, у дріжджів — 10 млн, у дрозофіли — 100 млн, у людини біля 3 млрд (3x10⁹).

Таким чином, першим фактором, що стимулював розвиток біоінформатики стала необхідність ефективного опрацювання величезних масивів інформації, отриманих експериментальними методами молекулярної біології. Біоінформатика лежить на стику молекулярно-біологічних і комп'ютерних технологій, її основні задачі можна було б визначити як розвиток та використання математичних і комп'ютерних методів для вирішення проблем молекулярної біології, першочергова серед них — це задача створення, підтримки і оновлення баз даних по молекулярній біології, інформація в які надходить з лабораторій, розсіяних по всьому світу. Починаючи

з 80-х років минулого століття були створені перші бази даних (БД) нуклеотидних послідовностей, вирішені питання зберігання, поповнення та забезпечення доступу до баз даних ДНК та РНК послідовностей та їх обробка (рис. 1). Таким чином перший етап інвентаризації всіх генів генома людини, що визначають життєдіяльність організму, був успішно пройдений.

Геноміка, протеоміка, функціональна геноміка

Однак, для вирішення сучасних задач біотехнології та медицини, а саме: створення нових лікарських препаратів, своєчасної діагностики більшості хвороб, підбору нових високоселективних ліків, а також для отримання організмів з наперед заданими властивостями недостатньо прочитати всі літери генетичного коду та визначити необхідний комплект генів, що приймає участь в життєдіяльності організму [2, 3].

У зв'язку з цим наприкінці ХХ століття вчені приступили до інвентаризації білків — основних функціональних структур організму, і в середині 90-х років в біології виник новий розділ молекулярної біології — протеоміка, яка вивчає якісний і кількісний склад білків, що синтезуються клітиною. Порівняння протеом різних клітин в нормі і при патології дозволяє

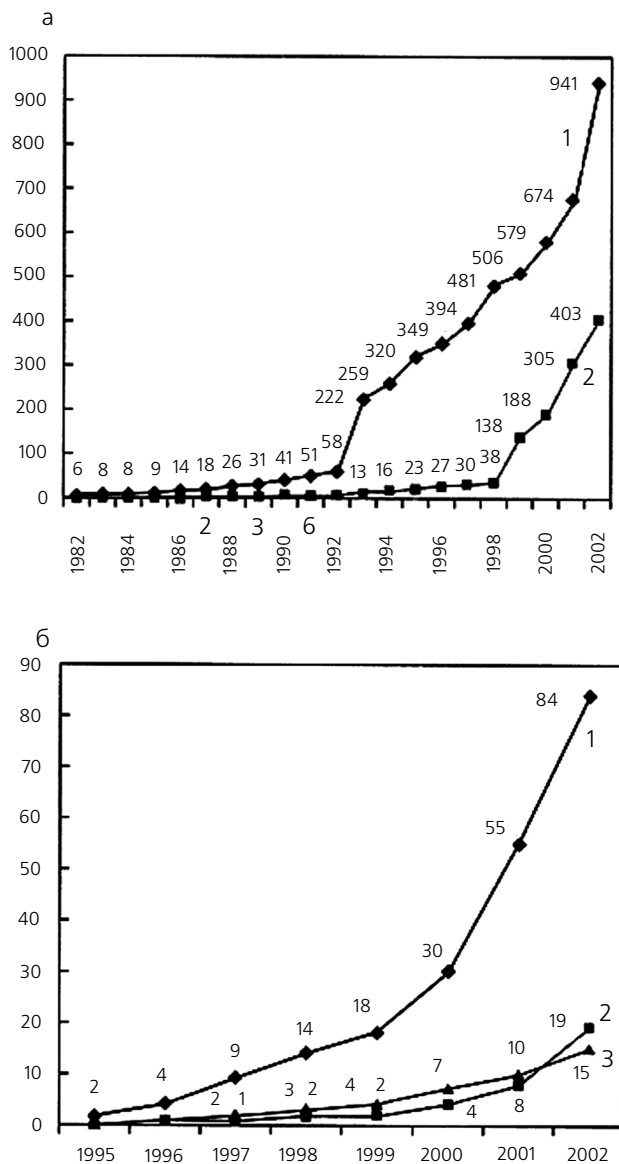


Рис. 1. Кількість повних геномів в базі даних GenBank: а — геноми вірусів (1) и органел (2); б — геноми бактерій (1), еукаріот (2) і архебактерій (3) [21].

визначити механізми, які впливають на розвиток патологічних реакцій. Результатом цього є розробка методів медичної діагностики, що базуються на змінах фізіологічного стану клітини і виражаються в пригніченні або стимуляції синтезу окремих білкових компонент клітини [2, 3, 4]. Так, в *in vitro* вже розроблені методи, які в змозі тимчасово або зовсім «виключити окремий ген» [4]. Ультрочутливість і специфічність сучасного протеомного аналізу дозволяють реєструвати розвиток патологічних процесів на ранніх етапах, без видимих симптомів захворювання, і ідентифікувати нові мішені для дії ліків, що створює основу пошуку нових патогенетичних

терапевтичних засобів. Одним з першорядних важливих медичних застосувань протеоміки — це виявлення пухлинних клітин, що «мовчать», за їх білковими профілями, порівняння клітин до і після зазначених впливів. Ці можливості створюють цілком нові перспективи як для діагностичної медицини, так і для фармацевтичної індустрії в плані створення нових лікарських препаратів.

Інший аспект протеоміки — з'ясування складу функціонально активних комплексів, що складають метаболічні ланцюги, а також взаємодії різних білків або субодиниць у складі олігомерних комплексів — це задачі функціональної протеоміки. Також функціональна

протеоміка іменується як клітинно-картована протеоміка, оскільки вона з'ясовує взаємодію білків у метаболічних шляхах.

Роль інформаційних технологій

Таким чином, другим фактором, що стимулював подальший розвиток біоінформатики, став розвиток молекулярної біології. Якщо на початковому етапі, як вже зазначалося, основним фактором розвитку біоінформатики був об'єм біологічної інформації, який настільки великий, що її обробка традиційними методами стала неможливою задачею. То розвиток геноміки, протеоміки, функціональної геноміки і т.п. є неможливим без сучасних комп'ютерних методів таких, наприклад, як співставлення генетичних даних та на основі аналізу цих даних формування теоретичних припущень, які можуть бути перевірені дослідним шляхом та дають можливість отримувати нову інформацію [2, 4, 18]. Так, досить часто існує можливість знайти схожий, вже вивчений білок, завдяки чому спрогнозувати функцію нового. У випадку, якщо схожі білки в базі даних відсутні, проводиться пошук подібних структурних елементів білка. Тому велику роль в розвитку біоінформатики мають бази даних нуклеотидних та амінокислотних послідовностей в плані не тільки створення, підтримки і оновлення баз даних по молекулярній біології, а і можливостей співставляти подібні послідовності, завдяки чому в даний час співставлення послідовностей стало звичною частиною роботи з геномами [3, 4, 8, 19, 20] і дає вагомий результат. Наприклад, завдяки застосуванню співставлення нуклеотидних послідовностей, було визначено, що вірус атипової пневмонії є продуктом злиття двох вірусів: людини і тварини, завдяки чому були визначені механізми його дії та методи нейтралізації [3, 4]. Так, у людини генів тільки в 4–5 разів більше, ніж, наприклад, у нематоди (круглий черв'як). Тому частина генів людини має споріднені гени серед відомих генів нематоди, що полегшує пошук нових генів людини. Функції невідомих генів нематоди вивчати легше ніж аналогічних генів людини, в них легко здійснювати мутації, виводити їх з дії одночасно відстежуючи зміни, що відбуваються в організмі.

Нанобіотехнології, наномедицина

Третім основним фактором розвитку біоінформатики стало те, що на сьогоднішній день передній фронт біотехнологічних та медичних досліджень перемістився з мікро- до наномасштабів та нанотехнологій [1], що знов таки в значній мірі пов'язано з розвитком геноміки і протеоміки [2] та сприяло розвитку нових інструментів та методів роботи з найбільш важливими біологічними макромолекулами, що потрапляють в область наномасштабів – ДНК, РНК та білків. Раніше для вивчення цих молекул та їх функцій використовувалась велика кількість хімічних реакцій. Проблема цього підходу полягає в тому, що реакції відбуваються стохастично, і тому можна вимірювати лише середні величини. Для того, щоб контролювати ці системи точно, необхідні інструменти, що взаємодіють з цими об'єктами на наномасштабі в їх природному середовищі [5]. Тому для розв'язання цих та багатьох інших задач нещодавно виникла нова область досліджень – нанобіотехнологія. Біотехнологія і нанотехнологія – це дві найбільш обіцяючи технології 21 сторіччя. Нанотехнологія – наука про наномасштабні структури, яка має справу зі створенням, дослідженням та використанню систем, що в 1000 разів менші, ніж компоненти, що на сьогоднішній день використовуються в області мікроелектроніки. Іншими словами нанотехнологія – це область науки, що стосується роботи з матеріалами та приладами на наномасштабному рівні. Нанобіотехнологія – це маніпуляція хімічними та фізичними властивостями речовини на молекулярному рівні. Нанобіотехнологія змінює спосіб нашого мислення, вона розмиває межі між фізикою, хімією та біологією. В промові всесвітнього відомого Нобелівського лауреата Річарда Феймана «На дні є багато кімнат» в 1959 році привертається особлива увага до цієї концепції – Якщо не вистачає розуму, то для зручності, розділимо цей всесвіт на частини, фізику, біологію, геологію, астрономію, і т.д. – Пам'ятайте, що природа їх не знає [1].

Очікується, що нанобіотехнологія створить інновації та зіграє важливу роль в різних біомедичних застосуваннях, не тільки в постачанні ліків, генній терапії, та генній інженерії [6], а також в формуванні зображень молекул, біомаркерів та біосенсорів. Транспорт ліків до мішеней і методи ранньої діагностики патологій є пріоритетними

областями наукових досліджень, в яких нанотехнологія гратиме основну роль [3]. Річард Смали, лауреат Нобелівської премії по хімії 1996 року проголосив про вступ в нову золоту еру пов'язану з становленням нової дисципліни, відомої, як наномедицина [7]. В широкому розумінні наномедицина – це процес діагностики, лікування, профілактики хвороб і травматичних випадків, розробка знеболювальних засобів, збереження і покращення здоров'я людини з використанням молекулярних (нано-) приладів і молекулярних знань про тіло людини.

Використання нанобіотехнологій включає в себе створення матеріалів і приладів, що спроектовані для взаємодії з тілами на субклітинному масштабі з високим рівнем селективності. Нанобіотехнологія пропонує багато революційних можливостей в боротьбі проти раку, нейродегенеративних захворювань, інфекцій та інших хвороб.

Так Національний інститут США «The National Institutes of Health Bioengineering Consortium» (BECON) проводив в 2000 році симпозиум «Нанонауки та технології формують біомедичні дослідження» [8]. Вісім областей нанонауки і нанотехнології були серед напрямків роботи конференції і вважається, що вони мають найбільше відношення до досліджень по біомедицині. Ці області включають в себе синтез і використання наноструктур, застосування нанотехнологій в терапії, біологічні наноструктури, електронно-біологічні поверхні розділу, прилади для ранньої діагностики хвороб, пристрої для вивчення окремих молекул, розробку сенсорів для космонавтів, військових, розробка біорідин та магнітокерованих сорбентів (для роботи з клітинами, макромолекулами, такими як ДНК, РНК, білки та ін.), нанофільтрацію, генетично модифіковану їжу та ін.

Області досліджень нанобіотехнології можуть знайти застосування в проєктуванні фармацевтичних засобів, як функції генотипу пацієнта, і в застосуванні хімічних речовин для стимуляції виробництва, як функції генотипу рослини. Необхідно взяти до уваги, що нанобіотехнологія не є окремою дисципліною, а є точкою перетину традиційних наук, таких як фізика, біологія, хімія, матеріалознавство, інформатика та ін., що несуть разом колективне знання і експертизу, необхідну для розвитку цих новітніх технологій.

Зокрема, фізичні методи, такі як атомна силова мікроскопія (АСМ), скануюча тунельна мікроскопія (СТМ), електрофорез, магнітофорез, імунмагнітна сепарація, оптичні та магнітні

нанопінцети та ін., використовуються як інструменти для наноманіпуляцій біологічними об'єктами. Наприклад, наноманіпуляція біологічними об'єктами застосовується в таких галузях як:

- *Молекулярна хірургія.* За допомогою фізичної маніпуляції можна розрізати ДНК в будь-якому бажаному місці, тобто молекулярна хірургія за допомогою АСМ, лазерів та ензимів може вилючати та модифікувати специфічні гени [9, 10].
- *Молекулярні мотори.* Формування зображень окремих молекул та маніпуляція ними допомагає зрозуміти, як хімічні реакції генерують рух і сили в молекулярних машинах, і це дасть можливість побудувати високоефективні штучні наномашини [1]. Останні роботи американських вчених дозволили створити двуюго крокуючу молекулу (Університет Каліфорнії у Риверсайді) та навіть молекулярний «автомобіль», що котиться (Техаський університет Райса).
- *Лабораторія на чіпі.* Методи такі, як електрофорез, магнітофорез, електрообертання та оптичне розсіювання можуть бути інтегровані в пристрої лабораторії на чіпі, що дозволяють працювати з нуклеїновими кислотами, частками вірусних розмірів і забезпечують побудову потужних діагностичних приладів незвичайно малих масштабів. Великих досягнень в цій області досягла, зокрема, фірма Affymetrix, ДНК-чіп якої вміщується на половині долоні.
- *Секвенування ДНК.* Звичайне секвенування не забезпечує точність розділу ДНК на окремі фрагменти і забирає дуже багато часу для довгих ДНК. З використанням точних маніпуляцій, наприклад, з використанням магнітних міток, фрагменти ДНК можуть вирізатися по порядку точно та швидко [12].
- *Розмірна сепарація ДНК.* Методи типу глобулярної трансформації і серії дипольних уловлювачів [13] можуть використовуватися для розмірної сепарації ДНК значно швидше, ніж гелеві методи для великих молекул. Дипольні уловлювачі, поміщені біля входу каналу капілярного електрофорезу, можуть значно покращити роздільну здатність цього методу [14]. Це може знайти широке застосування у сучасних мікрорідинних аналітичних установках.
- *Клітинна інженерія.* Може використовуватися для введення штучних хромосом в живі клітини і виробляти антитіла, які можуть використовуватися для лікування раку та інших хвороб [15, 16].
- *Нановиробництво.* АСМ може використовуватися для вибору та пересадження ДНК на силіконову поверхню. Це може використовуватися в нановиробництві сенсорів та біоелектронних пристроїв [17].

Задачі біоінформатики

Перехід від біотехнології до нанобіотехнології, тобто від мікро- до нанорівня, та відповідно подальше вдосконалення методів дослідження молекулярно-біологічних даних, розвиток протеоміки, функціональної протеоміки, метаболіки та ін. поставило перед біоінформатикою наступні задачі [2]:

- 1). Аналіз геномів, виділення з їх складу окремих генів, їх екзон-інтронної структури, сигнальних послідовностей і т.п.
- 2). Передбачення функцій генів і експресованих ними продуктів.
- 3). Виявлення генів – потенційних мішеней дії нових ліків.
- 4). Оцінка ролі окремих складових амінокислотної послідовності у функціонуванні білка.
- 5). Побудова молекулярних моделей білків і нуклеїнових кислот, виходячи з їх послідовностей.
- 6). Дослідження механізму функціонування макромолекул, виходячи з їх молекулярних моделей.
- 7). Комп'ютерне конструювання ліків, засноване на раціональному виборі генів-мішеней і молекулярних моделей їх білкових продуктів.

Таким чином, біоінформатика – область науки, що розробляє і застосовує технології інформатики для аналізу, систематизації молекулярно-біологічних даних, моделювання процесів, що відбуваються на молекулярному рівні з метою виявлення структур, функцій та взаємодії макромолекул (ДНК, РНК, білків) з подальшим використанням цих знань при створенні нових лікарських препаратів та нановиробів для діагностики і лікування, а також отримання організмів з наперед заданими властивостями [2].

Інформаційні ресурси з біоінформатики, пошукові системи, системи обробки інформації

На поточний момент відомо сотні повних геномів різних організмів, інформація про які знаходиться в БД, існують

сотні web- систем, на яких ці бази даних доступні для огляду та пошуку даних по молекулярній біології і близьким до неї дисциплінам. Кожна з них має свій формат даних, різну кількість надлишкової інформації, зв'язок з іншими базами даних [23, 24].

Найбільшими базами даних нуклеотидних та амінокислотних послідовностей є (EMBL, GenBank, DDBJ, Ген-експрес, SWISS-PROT, PIR) вони поповнюються дослідниками через Internet. Надалі ці дані проходять перевірку. Іншим джерелом інформації для них є наукова література [4, 45, 46]. Більшість баз даних, що містять подібну біологічну інформацію постійно синхронізуються, наприклад, бази даних нуклеотидних послідовностей EMBL, GenBank, DDBJ [29].

Поряд з загальними базами даних, що містять інформацію щодо нуклеотидних і амінокислотних послідовностей, існує багато спеціалізованих інформаційних ресурсів, які містять отримані за допомогою комп'ютерної обробки результати теоретичних прогнозів, що відносяться до певних предметних областей, спеціалізовані бази даних по регуляторних ділянках нуклеотидних послідовностей, бази даних по експресії генів, бази даних послідовностей РНК, структурі білків, мутаціям, метаболічним шляхам, трансгенним організмам, науковій літературі, програмному забезпеченню та по самим базам даних [25, 26, 27, 49].

Мабуть і в майбутньому поряд з загальними базами даних будуть створюватись спеціалізовані інформаційні ресурси, що мають відношення до певних предметних областей [52]. Перелік найбільш відомих біоінформаційних ресурсів наведено в табл.1.

Програмне забезпечення баз даних повинно задовільняти наступним функціональним вимогам [27, 32, 33, 50]:

- Необхідно, щоб об'єм баз даних був практично не обмеженим (тобто обмежений лише параметрами апаратних засобів).
- База даних повинна бути достатньо гнучкою для забезпечення проходження процесу перебудови по мірі її заповнення, так як попереднє проектування детальної структури бази даних по молекулярній біології є неможливим.
- Бази даних повинні бути інтегровані з різними міжнародними базами даних, постійно оновлюватися, синхронізуватися з аналогічними базами даних та мати системи перевірки даних, що надходять.
- Бази даних повинні підтримувати не лише стандартні мультимедійні формати, але й ряд спеціальних гіпермедіа-сервовищ (просторові структури молекул, хімічні структурні формули та ін.).

• Експлуатація та поповнення баз даних через комп'ютерні мережі має бути легко доступним та зрозумілим для користувачів, які не мають комп'ютерної підготовки (біологи, медики).

Всі існуючі бази даних надають можливість роботи з ними через Internet та практично усі вони використовують стандартні методики пошуку [50], наприклад, Genbank надає можливість роботи з одною з двох пошукових систем [29, 32, 33]:

- Entrez (пошук по назві, номеру, організму, автору і т. ін. [53]). Забезпечує доступ до амінокислотних і нуклеотидних послідовностей, їх тривимірних структур, а також до повних секвенованих геномів, надає графічне відображення генів. Практично для кожної послідовності можна підібрати подібні послідовності та вже розраховані і визначені дво- та тривимірні структури, що мають відношення до даної послідовності.
- Blast (пошук за подібністю) – порівнює надану інформацію з послідовностями, що вже є в базі для пошуку подібних.

Робота з БД може проходити в двох режимах:

- Offline-інтерфейс – спочатку з мережі Internet на локальний комп'ютер скачується частина бази даних, потім з цією частиною проводиться подальша робота.
- Режим клієнт-сервер – на локальному комп'ютері встановлюється програма математичної обробки нуклеотидних послідовностей або послідовностей амінокислот, далі дана програма з'єднується з сервером бази даних і обробляє інформацію без скачування останньої на локальний комп'ютер.

В той же час, інтенсивно розвиваються системи обробки інформації та пошукові системи, що збирають і обробляють інформацію відповідно до запитів користувачів [7, 47].

Як було відзначено, на поточний момент відомо сотні повних геномів різних організмів, більшість з яких не може бути детально вивчена в дослідній лабораторії [42, 44], так як пряме вивчення займає надто багато часу. В той же час від генетичної вивченості виду залежать усіх в розвитку біотехнології, медицини, фармації та інших пов'язаних з ними галузей. Генетична вивченість виду на сучасному рівні визначається насиченістю його геномної карти тобто кількістю генетичних маркерів (генів і нуклеотидних послідовностей), чіс положення (локалізація) визначено в хромосомі або локусі хромосоми, визначеністю інтрон-екзонних (кодуючи чи не кодуючи) областей генів, амінокислотних та просторових структур білків і т.д.

Табл. 1. Перелік найбільш відомих біоінформаційних ресурсів [27].

№	НАЗВА БАЗИ ДАНИХ	ХАРАКТЕРИСТИКА
НУКЛЕОТИДНІ ПОСЛІДОВНОСТІ		
1	GENBANK (National Centre for Biotechnology Information, Los-Alamos, USA) http://www.ncbi.nlm.nih.gov/ Заснована в 1982 р.	База даних генетичних послідовностей, містить анотовану колекцію всіх загально-доступних послідовностей ДНК, РНК та білків разом з літературними посиланнями. Поповнюється раз на 2 місяці. Станом на жовтень 1999 р. містить 3,841,163,011 нуклеотидних пар в 4,864,570 записах послідовностей, на вересень 2002 р. – 20,017,246,707 нуклеотидних пар в 8,293,265 записах послідовностей, на лютий 2004р. – 37,893,844,733 нуклеотидних пар в 32,549,400 записах послідовностей. Щоденно синхронізується з DDBJ та EMBL.
2	DDBJ (Center for Information Biology (CIB), DNA DATA Bank of Japan) http://www.ddbj.nig.ac.jp/ Заснована в 1984 р.	База даних генетичних послідовностей. Збір інформації проводиться в першу чергу у японських вчених та з літератури. 75% зібраних послідовностей являють собою частково секвеновані фрагменти ДНК з декількох сотень експресованих генів, так званих EST (Expressed Sequence Tags).
3	EMBL (European Molecular Biology Laboratory) Gezylberg, Germany http://www.ebi.ac.uk/embl/ Заснована в 1982 р.	База даних нуклеотидних послідовностей. Поповнюється безпосередньо авторами, що визначили первинну структуру фрагмента ДНК чи РНК, окрім того містить, інформацію щодо кожного фрагменту включаючи літературні посилання, перехресні посилання на документи з інших баз даних.
4	ГЕНЭКСПРЕСС (ВИНИТИ-ИМГ Россия) ftp://ftp.infobiogen.fr/pub/db/Genexpress	База даних генетичних послідовностей. Містить генетичні послідовності по 10,000 генів людини. Оновлюється раз на два місяці.
БІЛКОВІ ПОСЛІДОВНОСТІ ТА 3D СТРУКТУРИ		
5	SWISS-PROT (Department of Medical Biochemistry of the University) Switzerland, Geneva http://www.expasy.ch/sprot/sprot-top.html	Містить анотовані амінокислотні послідовності трансльовані з нуклеотидних послідовностей EMBL; адаптовані послідовності з PIR; а також послідовності опубліковані в літературі і надіслані безпосередньо авторами. Містить високоякісні анотації без збиткової інформації, посилання на споріднені бази даних (EMBL, Prosite, PDB). Кожна анотація містить опис функції білка, його доменної структури, особливостей пост-трансляційної модифікації. Оновлюється щотижня. Для академічних користувачів є безкоштовною.
6	PIR (Protein Information Resource, National Biomedical Research Foundation, USA) http://www-nbrf.georgetown.edu/pirwww/dbinfo/pirpsd.html	Містить інформацію щодо білків для яких відомі нуклеотидні послідовності. Пошук організовано як по таксономії так і гомології. Має низький рівень збиткової інформації. Поповнюється щотижня.
7	MMDB (Molecular Modelling Database, Georgetown University, USA) http://www.ncbi.nlm.nih.gov/Structure/	Містить просторові структури білків, визначені дослідним шляхом (рентгеноструктурною кристало-графією та ЯМР-спектроскопією), надає ін формацію про біологічну функцію та механізми, що з нею пов'язані; еволюційну історію і взаємозв'язок між макромолекулами. Входить до складу PDB, що містить також теоретичні моделі. Всі структури цієї бази даних мають первинні структури в NCBI. Оновлюється щоденно.
8	ENZYME http://www.expasy.ch/enzyme/	Містить інформацію щодо номенклатури ферментів, і описує всі типи білків, яким присвоєно номер ЕС (Enzyme Commission). Пошук реалізовано по ЕС-номеру, класам ферментів, хімічним компонентам, по кофакторам, по назвам хвороб, пов'язаних з ферментом. Версія 25.0 за липень 1999 року містить 3705 одиниці інформації.
СПЕЦІАЛІЗОВАНІ БІОІНФОРМАЦІЙНІ РЕСУРСИ		
9	ОМІМ – ГЕНЕТИЧНИХ ХВОРОБ (John Hopkins University, USA) http://www3.ncbi.nlm.nih.gov/omim	База даних по генах людини та пов'язаними з ними хворобами. Підтримується NCBI. Оновлюється щоденно.
10	МУТАЦІЙ БІЛКІВ - PMD (Protein Mutant Database, Japan) http://www.genome.ad.jp/htbin/www_bfind?pmd	База даних призначена для досліджень в області білкової інженерії. Одиницею інформації є не білок а одна стаття присвчена мутації в білках. Версія за вересень 2002р. містить 26651 статтю.
11	КАТАЛОГ ГЕНІВ – GENES KEGG (Kioto, Japan) http://www.genome.ad.jp/kegg/	База даних по систематичному аналізу функцій генів Складається з шести баз даних – метаболічних шляхів (PATHWAY), генів (GENES), лігандів (LIGAND), дослідних даних по експресії генів (EXPRESSION и BRITE), білків (SSDB). Забезпечує можливість роботи з усіма крупними світовими інформаційними ресурсами. Оновлюється щоденно.
12	БАЗА ДАНИХ ПО БАЗАМ ДАНИХ – DBCAT (Databases Catalog) http://www.infobiogen.fr/services/dbca	Містить анотації більш ніж по 500 біологічним базам даних. В тому числі короткий опис, призначення, авторів бази та її адресу.

Основна складність у процесі виділення генів полягає у визначенні екзон-інтронної структури генів еукаріот, оскільки кодуючі та некодуючі ділянки, однозначно не виділяються. Ділянки кодуючого гена (екзони) чергуються з некодуючими фрагментами (інтронами), число екзонів та їх довжина в генах не визначені. По мірі картування генів еукаріот (особливо ссавців) стало зрозуміло, що гени, однакові по еволюційному походженню і функціям, які вони виконують, (тобто гомологічні гени) часто виявляються зв'язаними з тими самими гомологічними генами в різних видах. На підставі цього передбачають ймовірний участок локалізації генів у одних видів, якщо відомо з якими генами гомологічні гени зв'язані в інших видах, тобто проводять «порівняльне картування». Таким чином існує можливість описати фізіологію організму шляхом комп'ютерного аналізу його генома, шляхом знаходження подібних структурних одиниць в геномах вже вивчених організмів [2, 3, 8, 22]. В даний час найбільш ефективним методом визначення біологічної функції гена є пошук схожих на нього нуклеотидних послідовностей в існуючих базах даних. З математичної точки зору поставлена проблема має відношення до задачі розпізнавання.

Вивчення просторової структури і моделювання молекул

Як вже було відмічено, для розуміння функціонування макромолекул необхідно знати не тільки їх первинну структуру, а й просторову структуру та їх динаміку. У той час як розшифровування первинної структури молекул ДНК людини завершена, вивчення її просторової структури і моделювання знаходиться ще на самому початку. Передбачення просторової структури білків по амінокислотній послідовності, тобто фактично по послідовності нуклеотидних пар у ДНК, є однією з центральних задач біоінформатики. Важливість розв'язання цієї задачі полягає ще й у тому, що число відомих первинних білкових структур, встановлених по відомих нуклеотидним послідовностям ДНК, набагато перевершує число експериментально підтверд-

жених просторових білкових структур. Формально, володіючи інформацією про взаємодію між окремими атомами в поліпептидному ланцюзі з відомою первинною структурою і оточуючими таким ланцюг молекулами розчинника, шляхом мінімізації вільної енергії всієї системи можна було б знайти просторову структуру. Але навіть для невеликого білка це задача пошуку глобального мінімуму функції десятків тисяч змінних, тому точне рішення подібної задачі за допомогою будь-якого суперкомп'ютера є цілком безперспективним [12]. В даний час для розв'язання проблеми фолдинга розроблена велика кількість наближених методів [11]. Відома просторова структура білка, гомологічного досліджуваному, вибирається у якості початкового наближення, а потім проводиться її уточнення [12]. Зрозуміло, що такі задачі під силу тільки сучасним суперкомп'ютерам, так як час і пам'ять, які використовуються в алгоритмах вирівнювання, зростають як квадрат довжини первинної структури. Так, для моделювання фрагментів ДНК, що містить 15 витків подвійної спіралі (150 пар основ), потрібно враховувати різні конфігурації як окремих нуклеотидів, що входять до складу ДНК, так і великого числа (білше сотні) молекул розчинника, що оточує макромолекулу. Такі обчислювальні експерименти з молекулами потребують величезних обчислювальних потужностей. Подібні задачі на сучасних комп'ютерах розраховуються днями і місяцями [43], і тільки при розпаралелювання обчислень і використанні суперкомп'ютерів, існує можливість істотно прискорити розрахунки, що дозволить вивчити просторову організацію великих фрагментів ДНК (сотні пар основ), включаючи надспіралізацію ДНК, механізми зв'язування ліків і інших біологічно-активних сполук із ДНК, а також комплексоутворення білок-ДНК [4, 34, 54, 55, 56]. При цьому одна з основних цілей дослідників полягає в розробці технології візуалізації, які дозволили б перевести великі масиви інформації у більш зручні для сприйняття форми, ніж строки коду чи потоки випадкових, на перший погляд, послідовностей літер.

Основним прямим методом визначення просторової структури біологічних макромолекул (білків, нуклеїнових кислот) та їх комплексів, при атомній розподільній здатності є рентгеноструктурний аналіз (РСА) та ЯМР. Процедура розшифровування структури цим методом є складним і дорогим процесом, який включає в себе: а) виділення й очищення білка; б) кристалізацію очищеного білка; в) рентгеноструктурний експеримент; г) комп'ютерне розшифровування структури білка. Але й при використанні прямих

методів комп'ютерна частина є необхідною складовою процесу розшифровування структури макромолекул оскільки дані, отримані в рентгенівському експерименті, містять лише частину інформації, необхідної для реконструкції просторової структури молекули білка [61–69]. Це складна обчислювальна задача, яка зводиться до пошуку локального мінімуму в просторі 104–106 змінних. Застосування вище зазначених підходів зіштовхується з особливо значними труднощами при роботі з великими макромолекулярними комплексами, які особливо цікаві для біологів та медиків. Ось чому поряд з експериментальними методами, визначення структури білка комп'ютерними методами по структурі білка, гомологічного досліджуваному, не має альтернативи на даний час. Зрозуміло, що така задача ставить серйозні вимоги до потужності комп'ютерів.

Суперкомп'ютерні технології в біоінформатиці

В зв'язку з вищевикладеним, суперкомп'ютерні технології для біологічних досліджень є необхідним та досить перспективним напрямком. Так в 2004 році було створено суперкомп'ютер здатний виконувати 100 трлн. операцій на секунду – в 80 разів більше в порівнянні з системами, що використовувались фірмою Celera Genomics [37] для запису генома людини. З часом вчені сподіваються створити суперкомп'ютер, що здатен виконувати 1000 трлн. операцій на секунду. За оцінкою Блейка, віце-президента компанії Compaq Computer по технічних обчисленнях, в найближчий час ринок технічних обчислень виросте майже в двічі, причому сегмент суперкомп'ютерів для досліджень в області генетики ростиме швидше за інші. Суперкомп'ютер, що розробляється Compaq Computer, Celera Genomics та Національною лабораторією в Сандії (шт. Нью-Мексико), міститиме від 10 до 20 тис. процесорів Alpha. Одною з головних задач Compaq є розробка оптимізованої до потреб геноміки програмної архітектури, що буде однаково ефективно працювати як на суперкомп'ютері зі 100 так і з 10 000 процесорів [37, 38, 59, 60].

До тих пір доки біологи не скористались суперкомп'ютерними технологіями для досліджень в області геному людини, сферою найбільш складних задач для обчислювальної техніки була ядерна

фізика. Зараз, при розробці алгоритмів для ведення генетичних досліджень, біологи повертаються до досліду фізиків-ядерщиків. Крег Вентер та інші вчені говорять про те, що нестача дослідників, котрі володіють навичками в області біоінформатики є фактором, що стримує дослідження в області геноміки. Таким чином, біоінформатика – складна дисципліна, що потребує знань в області фізики, математики, біології та інформатики [2, 20, 57].

При розробці систем візуалізації для геноміки і алгоритмів обробки генетичних даних на паралельних системах дослідники Сандії скористаються навичками, отриманими при створенні способів моделювання ядерних вибухів. «У нас в співтоваристві ядерщиків вважається, що нічого складніше, ніж розробка ядерного озброєння, не існує, – сказав Білл Кэмп, директор лабораторії в Сандії, – визнаю, геноміка складніше» [35, 37].

Біотехнологічні компанії конкурують між собою в пошуках методів обробки генетичної інформації. Так комп'ютерний алгоритм багатоваріантного аналізу активності генів, розроблений в американському Національному центрі дослідження генома людини, дозволяє вченим аналізувати взаємодію різних генів, в результаті якої виникає захворювання. Так лише близько 2% захворювань викликано дефектами генів та хромосом, а 98% захворювань викликано порушенням експресії нормальних генів, порушенням їх функцій. Тому алгоритм багатоваріантного аналізу активності генів може застосовуватись при дослідженні будь-якого захворювання [39]. Обмежуючим фактором розвитку є обчислювальна потужність комп'ютера, що застосовується при обробці генетичних даних за допомогою певного алгоритму. Тому фірми NuTec та IBM планують не тільки створити суперкомп'ютер для багатоваріантного аналізу активності генів, але й надати університетам, що працюють в області біотехнології, компаніям та фармацевтичним фірмам доступ через Internet до цього суперкомп'ютера та можливість використовувати даний алгоритм [40, 41].

Стан та перспективи розвитку біоінформатики

В останній час кількість центрів біоінформатики виросла в декілька разів

і складає близько двох сотень по всьому світу. За кордоном біоінформатика перетворилась на область науки, що швидко розвивається [2]. Проекти в області біоінформатики активно фінансуються фармацевтичними та біологічними фірмами та науковими фондами США, Японії та країн Євросоюзу. В той же час на Україні на поточний момент відсутня будь-яка координація робіт в даному напрямку, нема державної підтримки наукових проектів. Користувачами ж подібних ресурсів є: [29]

- науково-дослідні інститути та ВУЗи біологічного, біотехнологічного, медичного та хімічного профілю;
- виконавці проектів по грантам МОН та НАН;
- біотехнологічні та фармацевтичні фірми.

Вже зараз звертають на себе увагу гігантські темпи росту біотехнологічних продуктів: у 1999 р. розроблялося 1200 нових молекул для фармацевтичних цілей, із них більш 50% відносилось до біотехнологічних продуктів. Ріст біотехнологічного ринку можна оцінити, порівнявши такі показники: якщо в 1991 р. загальний обсяг продажу біотехнологічної продукції складав 2 млрд дол., у 2000 р. – 16,5 млрд, то до 2008 р. прогнозується обсяг до 75 млрд дол. Таких темпів росту не знає жодна з існуючих галузей промисловості. Більш 70% усіх досліджень і практичних результатів у біотехнології пов'язано з одержанням фармацевтичних і біомедицинських препаратів. Технологічний вибух у пост-геномній ері спричинив за собою істотні зміни в інфраструктурі біотехнологічної індустрії. Сьогодні багато компаній займаються науковими дослідженнями: розробкою лікарських засобів, скринінгом геномів, відкриттям нових мішеней. До них підключилися приладобудівні фірми, що забезпечують технічний прорив у сучасній біотехнології. У цей процес утягуються біохімічні компанії, що упроваджують високо-стандартизовані реактиви і комплексні системи аналізу.

Висновки

Таким чином, наступна революція, що очікується в високих технологіях, включає злиття нано- біо- та інформаційних технологій. Біоінформатика та нанобіотехнологія є найбільш наукоємними розділами технології та науки, в яких вітчизняні вчені можуть працювати на світовому рівні [7, 29]. Але для цього необхідно створення вітчизняного цент-

ру біоінформатики та його інтеграція з світовими біоінформаційними ресурсами.

Задачі центру:

- Наукова робота (проведення наукових досліджень та розробка нових інформаційних технологій).
- Навчальна робота (організація навчального процесу студентів, аспірантів, докторантів та викладацького складу; створення курсів підвищення кваліфікації; проведення конференцій, семінарів і т. ін.
- Надання послуг в сфері біоінформатики (методичних, інформаційних, наукових та ін.).

Для того щоб Україна не залишилась на узбіччі процесу розвитку біоінформатики в світі і змогла брати участь в сучасних біотехнологічних проектах, необхідно підсилити природні людські ресурси найбільш сучасними високотехнологічними засобами. Сучасна наука потребує все більших інформаційних ресурсів. Самі комп'ютери не здатні вирішувати наукові задачі, однак вирішувати такі задачі без комп'ютерів вже неможливо. Тому створення комп'ютерного центру біоінформатики, забезпечить інтеграцію та виконання на сучасному рівні робіт в області генетичної та протеомної інформації починаючи з розробки теоретичних методів і закінчуючи практичним застосуванням результатів фундаментальних досліджень в біології та медицині.

Література

1. Salata O. V. Application of nanopractices in biology and medicine // *Journal of nanobiotechnology*. – 2004, – P. 95–101.
2. Арчаков А. И. Геномика, протеомика и биоинформатика — науки XXI столетия // *Фармацевтический вестник*, №9 (208), 13 марта 2001 г.
3. Арчаков А. И., Геномика, протеомика, биоинформатика. // *Медицинская газета* № 26, 11 апреля 2001 г.
4. Свердлов Е. Д. Фрэнсис Крик в его прогнозе на 2000 год был почти абсолютно прав // *Биоорганическая химия* том 26, № 10, – 2000, – С. 761–766.
5. Шабарова З. А., Богданов А. А. Химия нуклеиновых кислот и их компонентов. – М.: Химия, 1978.
6. Баранов В. Медицина на пороге революции // (url) <http://nauka.relis.ru/08/0009/08009008.htm>
7. Арчаков А. И., Поройков В. В., Белкина Н. В., Гусев С. А., Дубанов А. В., Иванов А. С., Лагунин А. А., Лисица А. В., Скворцов В. С., Соболев Б. Н. Новые технологии в биомедицине: Биоинформатика. // (url) <http://www.mirrebenka.h10.ru/genomics.htm>

8. Мисуров Д. Нано-биотехнологические исследования находят применение в медицине. // В МИРЕ НАУКИ, № 8, – 2005.
9. I. Safarik, M. Safarikova, Magnetic nanoparticles and biosciences // Monatshefte fur Chemie, – 2002, – P. 737–759.
10. A. Mehta, M. Rief, and et al. Single-molecule biomechanics with optical methods, Science, vol. 283, – 1999, – P. 1689–1695.
11. Mahtab R., Rojers J. P., Mnrphy C. J. Protein sized quantum dot can distinguish between «straight», «bent» and «kink oligonucleotides» // J. Am. Soc. – 1995, – P. 9099–9100.
12. Cote R., D. Rothwell, J. Palotay, R. Beckett and L. Brochu, Eds. SNOMED International: The Systematized Nomenclature of Human and Veterinary Medicine. Northfield, IL: College of American Pathologists, – 1993.
13. Molday R. S., MacCenzie D. Immuno-specific ferromagnetic iron dextran reagents for the labeling and magnetic separation of cells // J immunol Mae-hods, – 1982, – P. 353–367.
14. Ma J., Wong H., Kong L. B. Peng K. W. Biometric processing of nanocrystallite bioactive apatite coating on titanium // nanotechnology, – 2003, – P. 619–623.
15. de la Isla A., Brostow W., Bujard B., Estevez M. et al Nanohybrid scratch resistant coating for teeth and bone viscoelasticity manifested in tribology // Mat Resr Innovat – 2003, – P. 110–114.
16. G. V. Shivashankar and A. Libhaber Single DNA molecule grafting and manipulation using a combined atomic force microscope and an optical tweezer, Appl. Phys. Lett., vol. 71, – 1997, – P. 27–3729.
17. Sahoo K. S. and Labhasetwar V. Nanotech Approaches to Drug Delivery and Imaging, DDT Vol. 8, – 2003, – P. 1112–1120.
18. Гельфанд М. С., Миронов А. А. Вычислительная биология на рубеже десятилетий // Молекулярная биология. – 1999, С. 969–984.
19. Экспериментальные методы исследования белков и нуклеиновых кислот // Под ред. М. А. Прокофьева. – М.: Изд-во МГУ, 1985.
20. М. С. Гельфанд, В. А. Любецки. Биоинформатика от эксперимента к компьютерному анализу и снова к эксперименту. Вестник РАН, том 73, № 11, – 2003 С. 987–994.
21. Koonin E. V., Galperin M. Y. Sequence-Evolution-Function: Computational approachers in comparative genomics. Kluwer Academic Press, – 2003.
22. Bryan Bergeron Bioinformatics Computing. Prentice Hall PTR, – 2002, – 439p.
23. Balas, E., S. Boren and G. Brown, Eds. Information Technology Strategies from the United States and the European Union. Amsterdam: IOS Press, – 2000.
24. Altschul, S., M. Boguski, W. Gish and J. Wootton. Issues in Searching Molecular Sequence Databases. // Nature Genetics – 1994, – P. 119–129.
25. Jagota A., Data Analysis and Classification for Bioinformatics. Santa Cruz, CA: Arun Jagota, – 2000.
26. Mack, R., Y. Ravin and R. Byrd. Knowledge Portals and the Emerging Digital Knowledge Workplace. // IBM Systems Journal – 2001, – P. 814–830.
27. Объединенный центр вычислительной биологии и биоинформатики. // Институт математических проблем биологии РАН. (url) <http://www.jcibi.ru>.
28. Арчаков А. И., Поройков В. В., Белкина Н. В., Гусев С. А., Дубанов А. В., Иванов А. С., Лагунин А. А., Лисица А. В., Скворцов В. С., Соболев Б. Н. Новые технологии в биомедицине: Биоинформатика. // (url) <http://www.mirrebenka.h10.ru/genomics.htm>
29. (url) [http://severin.ru/laboratories/ petrosyan_grp.php](http://severin.ru/laboratories/petrosyan_grp.php), – 2004
30. Celko, J. Data & Databases: Concepts in Practice. San Francisco, CA: Morgan-Kaufmann, – 1999.
31. Binstock, A. and J. Rex. Practical Algorithms for Programmers. New York: Addison-Wesley, – 1995.
32. Benson, D., I. Karsch-Mizrachi, D. Lipman, J. Ostell, B. Rapp and D. Wheeler. «GenBank.» Nucleic Acids Research, – 2000, – P 15–18.
33. Компьютерный анализ генетических текстов // А. А. Александров, Н. Н. Александров, М. Ю. Бородавский И ДР. М.: Наука, – 2002.
34. Bergeron, B. and R. Rouse. Cognitive Aspects of Modeling and Simulating Complex Biological Systems. Structuring Biological Systems: A Computer Modeling Approach (S. Iyengar, Ed.). Boca Raton, FL: CRC Press, – 1992, – P. 125–166.
35. Marti-Renom, M., A. Stuart, A. Fiser, R. Sanchez, F. Melo and A. Sali. Comparative Protein Structure Modeling of Genes and Genomes. // Annual Review of Biophysics and Biomolecular Structure, – 2000, – P. 291–325.
36. (url) <http://compaq.com>.
37. (url) <http://www.celera.com>.
38. Sahoo K. S. and Labhasetwar V. Nanotech Approaches to Drug Delivery and Imaging, DDT Vol. 8, No. 24, – 2003, – P. 1112–1120.
39. (url) <http://nutec.com>.
40. (url) <http://www.ibm.com>.
41. D. Bakewell, M. Hughes, J. Milner, and H. Morgan, Dielectrophoretic manipulation of Avidin and DNA, Proc of the Int. Conf. of the IEEE Engg. In Medicine and Biology Soc., Vol. 20, N. 2, – 1998, – P. 1079–1082.
43. Banatao R. Visualization. Bioinformatics Methods and Techniques // Stanford University, Stanford Center for Professional Development, – 2002.
44. Vaxevanis, A. and B. Ouellette, Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins. New York: John Wiley & Sons, – 2001.
45. Jagota, A. Data Analysis and Classification for Bioinformatics. Santa Cruz, CA: Arun Jagota, – 2000.
46. Garfinkel S. Database Nation. Sebastopol, CA: O'Reilly & Associates, – 1997.
47. Ratledge C. and Kristiansen B. Eds. Basic Biotechnology. New York: Cambridge University Press, – 2001.
48. Smith D. Java for the World Wide Web. Berkley, CA: Peachpit Press, – 1998.
49. Agarwal P. and States D. Comparative Accuracy of Methods for Protein Similarity Search. Bioinformatics, – 1988, – P. 40–47.
50. Altschul, S., M. Boguski, W. Gish and J. Wootton. Issues in Searching Molecular Sequence Databases. // Nature Genetics, – 1994, – P. 119–129.
51. Binstock A. and Rex J. Practical Algorithms for Programmers. New York: Addison-Wesley, – 1995.
52. Cote, R., D. Rothwell, J. Palotay, R. Beckett and L. Brochu, Eds. SNOMED International: The Systematized Nomenclature of Human and Veterinary Medicine. Northfield, IL: College of American Pathologists, – 1993.
53. Van Ginneken, A. The Structure of Data in Medical Records. // Yearbook of Medical Informatics (J. Van Bommel and A. McCay, Eds.). Stuttgart, Germany, Schattauer, – 1995, – P. 61–70.
54. Altman R. Pairwise & Multiple Sequence Alignment, Bioinformatics Methods and Techniques. Stanford University, Stanford Center for Professional Development, – 2002.
55. Altschul, S., W. Gish, W. Miller, E. Myers and D. Lipman. Basic Local Alignment Search Tool. // Journal of Molecular Biology, N 215, – 1990, – P. 403–410.
56. Apostolico, A. and R. Giancarlo. Sequence Alignment in Molecular Biology. Journal of Computational Biology, N 5, – 1998, – P. 173–196.
57. Cheng B. Protein Structure Analysis. // Bioinformatics Methods and Techniques. Stanford University, Stanford Center for Professional Development, – 2002.
58. Eddy S. Hidden Markov Models. // Current Opinion in Structural Biology, N 6, – 1996, – P. 361–365.
59. Schuler G. Sequence Alignment and Database Searching. // Meth Biochem Anal, N 39, – 1998, – P. 145–171.
60. Thompson, J., F. Plewniak and O. Poch. A Comprehensive Comparison of Multiple Sequence Alignment Programs. Nucleic Acids Research, – Vol 27, – 1999, – P. 2682–2690.
61. Попов Е. М. Успехи химии, Т 64, – 1995, – С. 143–1182.
62. Зоркий П. М., Зоркая О. Н. Ординарная органическая кристаллохимия // Журн. структур. химии, Т 39, № 1, – 1998, – С. 126–151.
63. Зоркий П. М. Структура органического кристалла. // соросовский образовательный журнал Т 7, № 11, – 2001, – С. 53–58.
64. Лахно В. Д. Биоинформатика и высокопроизводительные вычисления / В. Д. Лахно // Вестн. РФФИ., – 2000, – С. 38–45.
65. Копылов А. М. Рентгеноструктурный анализ рибосом – статика динамики / А. М. Копылов // Биохимия. – 2002. – Т. 67, вып. 3. – С. 449–460.

66. Копылов А. М. Рентгеноструктурный анализ РНК – белковых комплексов / А. М. Копылов // Биохимия., – Т.61., – 1996, – С. 1911–1916.
67. Лунин В. Ю. Рентгеноструктурный анализ белков: пробл. расшифровки структуры: Докл. на III всесоюз. совещ. «Математические методы для исследования полимеров», (Пушино, 2123 июня 1983г.): Препринт / В. Ю. Лунин. – Пушино, – 1983, – 19с.
68. Никифоров В. Г. Структурно-функциональные исследования РНК-полимеры (1962-2001) / В. Г. Никифоров // Молекулярная биология., – 2002, – С. 167–207.
69. Компьютеры и суперкомпьютеры в биологии Под. ред. В. Д. Лажно и М. Н. Устинова Москва–Ижевск: Институт компьютерных исследований, – 2002, 528 с.

Биоинформатика как основной инструмент нанобиотехнологии и наномедицины

С. В. Горобец, О. Ю. Горобец, Д. О. Дереча

Национальный технический университет Украины «Киевский политехнический институт»

Резюме

В данной работе рассмотрены основные достижения и сферы применения биоинформационных технологий. Указано на перспективы их развития совместно с параллельными областями знаний, такими как нанобиотехнология, наномедицина и др.

Ключевые слова: биоинформатика, геном, база данных, ДНК, нанобиотехнология.

Bioinformatics as the basic tool for nanobiotechnologies and nanomedicine (review)

S. V. Gorobets, O. Yu. Gorobets, D. O. Derecha

National Technical University of Ukraine «Kiev Polytechnical Institute»

Abstract

In the current work the basic achievements and spheres of application of bioinformatics technologies are considered. It is specified prospects of their development together with other fields of knowledge, such as, nanobiotechnology, nanomedicine etc.

Keywords: Bioinformatics, genom, data base, DNA, nanobiotechnology.

Переписка

С. В. Горобець, О. Ю. Горобець, Д. О. Дереча

Національний технічний університет України «Київський політехнічний інститут»

кафедра біоінформатики
пр-т Перемоги, 37, корпус 4, к. 273
Київ, 56
Україна
тел.: + 380 (44) 454 99 37
эл. почта: dderecha@yandex.ru,
gorob@mail.kar.net