

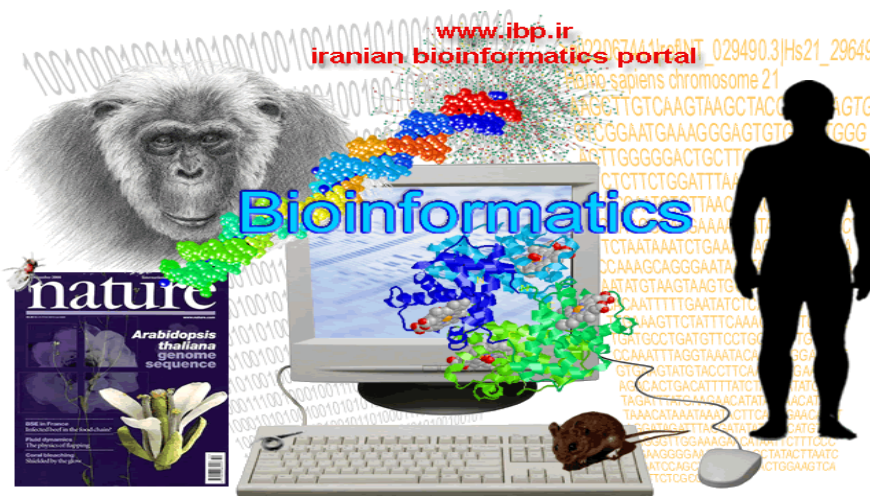
فصل اول

مقدمه ای بر بیوانفورماتیک،

پروتئومیکس و ژنومیکس

به همراه مقدمه ای بر NCBI و سازمان NIH

بیوانفورماتیک



مقدمه

درک زبان خاموش اما جذاب سلولهای زنده، تلاش بیولوژی مولکولی مدرن است. از یک الفبای ساده چهار حرفی (آدنین، گوانین، سیتوزین، تیمین) دستورالعمل فرآیندهای حیاتی شکل می گیرد که پیچیده ترین بیان و گفتار آنها "انسان" است. حجم فوق العاده زیاد داده های مولکولی و الگوهای مبهم و مرموز آنها ما را بر آن داشته است تا به نیاز مطلق به پایگاه های داده کامپیوتری و ابزارهای تحلیل داده ها پی ببریم. تلاش و مبارزه ما یافتن روش های جدید، جهت مدیریت این داده ها حجیم و پیچیده است و نیز فراهم کردن بستری جهت

دسترسی به ابزارهای کامپیوتری و تحلیل داده هاست تا میراث ژنتیکی خود را بیش تر درک کنیم و نقش آن را در سلامت و بیماری‌ها بیابیم.

تعریف: بیوانفورماتیک شاخه ای از علم است که از کامپیوتر بهره می گیرد تا اطلاعات بیولوژی مولکولی - توالی های DNA یا پروتئین‌ها- را در کامپیوتر ذخیره نماید و با ابزارهای کامپیوتری و الگوریتم های قدرتمند ریاضی آنالیز و تحلیل نماید.

نیاز علم ژنتیک به کامپیوتر: بیولوژی مولکولی و علم ژنتیک مسایلی در پیش دارند که بیوانفورماتیک می تواند با به کار بردن این اطلاعات کامپیوتری شده به حل آنها کمک نماید. حجم فوق العاده زیاد داده‌ها جهت نگهداری و مقایسه های گاهاً میلیونی رکوردها بسیار مشکل و گاهی غیر ممکن است.

یکی از کاربردهای بیوانفورماتیک تحلیل این داده‌ها جهت پی بردن به معمای تکامل هستی است. حل این معما در میلیاردها نوکلئوتید درون ژنوم موجودات زنده نهفته است.

پایگاه داده: پایگاه داده یا database یک برنامه کامپیوتری است که با روش های بسیار منظمی اطلاعات را در کامپیوتر ذخیره می کند و با سرعت فوق العاده ای در آن اطلاعات جست و جو کرده و نتیجه‌ی جستجو را ارایه می دهد.

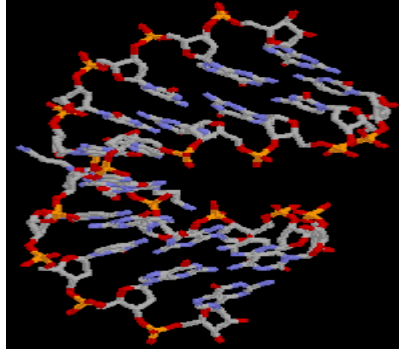
نقش پایگاه‌های داده: مشهورترین کاربرد بیوانفورماتیک در تحلیل توالی هاست. توالی های DNA مربوط به ارگانیزم‌های مختلف جهت دستیابی سریع و مقایسه آنها با یکدیگر، در پایگاه های داده ذخیره می شوند. پروژه ژنوم انسان که از سال ۱۹۹۶ تا سال ۲۰۰۳ به طول انجامید نمونه ای از تحلیل توالی هاست. در این پروژه با استفاده از کامپیوترهای بزرگ

و روش های مختلف به دست آوردن توالی ها، همه ژنوم انسان تعیین توالی گردید و درون یک پایگاه داده قرار گرفت. با کامل شدن نقشه ژنوم انسان بیوانفورماتیک در تحقیقات سرطان ها به امید رسیدن به یک درمان موفق و نهایی بسیار با اهمیت شده است.

پایگاه های داده به دو دسته اصلی و فرعی تقسیم می شوند. نتایج تجربی حاصل از تحقیقات علمی مانند توالی های نوکلئوتیدی یک ژن خاص، که در یک آزمایش تجربی به دست آمده است درون پایگاه های داده ی اصلی قرار می گیرند. این داده ها خام و بدون تحلیل هستند. نمونه ای از این نوع پایگاه ها GenBank است که توالی های نوکلئوتیدی را نگهداری می کند. این پایگاه توسط NCBI مدیریت می شود.

آشنایی با بیوانفورماتیک:

در چند دهه گذشته، پیشرفت در زمینه بیولوژی ملکولی و نیز تجهیزات موجود برای پژوهش در این زمینه موجب سرعت بخشی به زنجیره یابی پروتئین های بزرگ در ژنوم گونه های مختلف گردیده است. در واقع، بسیاری از ژنوم های باکتری همانند eukaryote های ساده (مانند *Saccharomyces cerevisiae* یا baker's yeast) به طور کامل زنجیره یابی شده اند. پروژه ژنوم انسان که برای یافتن زنجیره ۲۴ کروموزومی انسان طراحی شده است در حال حاضر در دست توسعه است. بانک های اطلاعاتی زنجیره ای رایج مانند GenBank و EMBL با نرخ توانی در حال بزرگ شدن هستند. این ازدیاد اطلاعات، ذخیره سازی مطمئن، سازمان بندی و شاخص گذاری اطلاعات زنجیره ای را می طلبد. در این راستا علم فناوری اطلاعات در بیولوژی به کار گرفته شده و رشته ای به نام **بیوانفورماتیک** را بوجود آورده است.



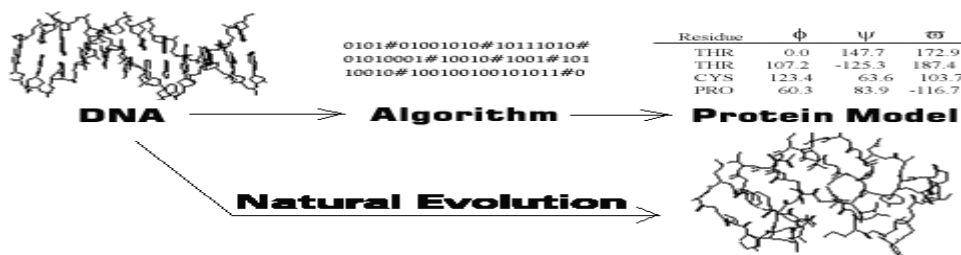
ابتدایی ترین کار در بیوانفورماتیک ایجاد و نگهداری اطلاعات بیولوژیکی در بانک های اطلاعاتی است. زنجیره اسیدهای نوکلئیک (و زنجیره پروتئین های مشتق شده از آنها) بخش عمده این بانک های اطلاعاتی را تشکیل می دهند. در حالی که ذخیره سازی و سازماندهی میلیون ها نوکلئوتید امری پیش پا افتاده به شمار می رود، طراحی یک بانک اطلاعاتی و توسعه رابطی که محققان بتوانند با آن دسترسی به داده های موجود داشته و نیز داده های جدید را وارد کند امری در دست اقدام است.

بیشترین کاری که در بیوانفورماتیک انجام می شود تحلیل اطلاعات زنجیره است. بیولوژی محاسبه ای (Computational Biology) نامی است که به این پردازش داده شده و موارد زیر را در بردارد:

- یافتن ژنها در زنجیره های DNA در اندام های مختلف
- توسعه متدهایی برای پیش بینی ساختار و یا عملکرد پروتئین های جدیداً کشف شده و زنجیره های RNA ساختیافته
- کلاستر بندی زنجیره های پروتئینی در خانواده زنجیره های مرتبط و توسعه مدل های پروتئینی
- مشخص کردن پروتئین های مشابه و ایجاد درختهای phylogenetic برای آزمایش کردن روابط evolutionary



پروسه تکامل، زنجیره های DNA ای را ایجاد نموده است که پروتئین هایی با عملکرد بسیار ویژه را رمزگذاری می کنند. پیش بینی ساختار سه بعدی پروتئین ها با استفاده از الگوریتم ها ممکن می شود که این الگوریتم ها از دانش فیزیک، شیمی و از همه مهمتر از آنالیز پروتئین های دیگر با زنجیره های اسید آمینه مشابه، مشتق می شوند. دیاگرام زیر پروسه ای که طی آن زنجیره های DNA برای مدل نمودن ساختار پروتئینی بکار می روند را نشان می دهد.



NCBI

در نوامبر ۱۹۸۸ بخش جدیدی به کتابخانه پزشکی ملی آمریکا (NLM) افزوده شد تا روش‌های کامپیوتری پردازش اطلاعات جهت هدایت تحقیقات Biomedical ایجاد شود. این بخش مهم، مرکز ملی اطلاعات بیوتکنولوژی یا NCBI نام گرفت.

کتابخانه ملی پزشکی آمریکا در سازمان ملی بهداشت آمریکا (NIH) قرار گرفته است. تجربه‌های موفق این کتابخانه در ایجاد و نگهداری پایگاه‌های داده در زمینه Biomedical و به عنوان بخشی از NIH باعث ایجاد یک برنامه تحقیقاتی در زمینه بیولوژی مولکولی کامپیوتری، شده است. امروزه NCBI به عنوان بزرگترین مرکز تحقیقات Biomedical در دنیا شناخته شده است. انستیتوی NCBI جهت رسیدن به اهداف خود فعالیت‌های زیر را دنبال می‌کند:

- هدایت تحقیقاتی در زمینه مسایل اصولی Biomedical در سطح مولکولی با بهره‌گیری از روش‌های ریاضی و کامپیوتری

- همکاری با انستیتوهای دیگر NIH، دانشگاه‌ها، صنعت و دیگر سازمان‌های دولتی
- ایجاد ارتباط‌های علمی به وسیله برگزار کردن همایش‌ها، کارگاه‌های آموزشی و

سلسله سخنرانی‌ها

- برنامه‌های تحصیلاتی برای دانشجویان در زمینه‌های بیولوژی کامپیوتری
- توسعه و انتشار نرم‌افزارها و پایگاه‌های داده مختلف

سازمان NIH

سازمان NIH علاوه بر کتابخانه NLM از ۲۷ انستیتو و مرکز تحقیقاتی دیگر تشکیل

شده است.



شکل ۱-۱

از جمله این انستیتوها می توان به

*انستیتوی ملی سرطان

*انستیتوی ملی تحقیقات ژنوم انسان

*انستیتوی ملی دیابت و بیماری های کلیه و گوارش، و

*انستیتوی ملی کودکان

اشاره کرد.

وب سایت سازمان NIH با آدرس www.nih.gov قابل دسترسی است. این ۲۷ انستیتو

زیر نظر بخش مرکزی NIH به نام Office of Director یا OD رهبری و مدیریت می شود.

OD مسؤول تعیین سیاست های NIH در برنامه ها، مدیریت ها و فعالیت ها است.

در سایت NIH می توان با رؤسای این سازمان ها و انستیتوها به وسیله Email ارتباط

برقرار کنید. آدرس های Email، آدرس پستی و مشخصات کامل آنان در این سایت درج شده

است.

ژنومیک یا ژنومیکس

ژنومیک شامل تجزیه و تحلیل داده‌ها و اطلاعات ژنتیکی بخصوص ژنوم موجودات است.

ژنوم توالی کل DNA موجود در سلول‌های یک جاندار است که بعنوان ماده ژنتیکی عمل می‌نماید و سبب بروز صفات وراثتی (فنوتیپ (می‌شود. با انتقال ماده وراثتی از یک نسل به نسل دیگر، صفات ارثی از یک نسل به نسل بعد منتقل می‌شود. در موجوداتی که تولید مثل جنسی دارند، ژنها از طریق سلول جنسی نر (اسپرم (و سلول جنسی ماده (اووم (به نوزاد منتقل می‌شود.

بطور خلاصه باید گفت که ژنومیک شامل توالی یابی و آنالیز ژن‌ها و رونوشت‌های آنها در یک موجود زنده است

پروتئومیک یا پروتئومیکس

پروتئومیکس دانش بررسی ساختار و عملکرد پروتئین‌ها در مقیاس بزرگ است. این واژه را به قیاس ژنومیک (به معنی دانش بررسی ژن‌ها) ساخته‌اند.

مقدمه

با تکمیل پروژه ژنوم انسان مشخص شد که مکانیسم مولکولی رفتار سلولها در شرایط مختلف را نمیتوان از روی توالی ژنهای آنها پیشگویی کرد. رفتار سلولی و تمام فعالیتهایی که در سلول انجام می‌شود بر عهده پروتئین‌ها است. در واقع برای ارتباط ژنوم با رفتار سلولها باید پروتئینهای سلولها را شناخت. به کلیه پروتئینهایی که در یک سلول در یک زمان مشخص بیان می‌شود، پروتئوم آن سلول گفته می‌شود و این پروتئوم است که فاصله بین ژنوم و مکانیسم مولکولی رفتار سلولی را پر می‌کند. برخلاف ژنوم، برای هر اورگانیسم نمیتوان یک پروتئوم واحد تعریف کرد. پروتئوم سلولهای مختلف با یکدیگر متفاوت اند. یعنی سلولها علاوه بر پروتئینهای ضروری که در همه انواع سلولها بیان می‌شوند، دارای یکسری پروتئینهای اختصاصی نیز هستند. از این رو بهتر است پروتئوم را برای هر یک از انواع سلولها تعریف نمود. با این حال پروتئوم یک نوع سلول نیز همیشه ثابت نیست. سلول در برابر شرایط مختلف محیطی و پیامهایی

که از سلولهای اطراف دریافت می کند، پروتئینهای مختلفی را بیان می کند. عبارت دیگر هر سلول تحت شرایط مختلف پروتئومهای متفاوتی دارد. بنابر این برای شناسایی مکانیسمهای مولکولی رفتار سلولی و واکنشهای زیستی، لازم است پروتئینهایی که در یک سلول بیان می شود، تغییرات آنها در شرایط مختلف، عملکرد آنها و همچنین برهمکنشهای بین پروتئینهای مختلف در یک سلول، بررسی شود. به مجموعه این بررسیها، نقشه برداری پروتئوم یا پروتئومیک، گفته می شود.

نقشه برداری پروتئوم

مطالعه پروتئوم به سادگی مطالعه ژنوم نیست. زیرا پروتئینها را نمیتوان (همانند DNA) به روشی مشابه PCR، تکثیر کرد. همچنین توالیهای پلیپتیدی نمیتوانند به توالیهای اسید آمینه ای مکمل خود متصل شوند. بنابر این برای مطالعه پروتئوم باید از ابزار و روشهای ویژه ای استفاده کرد. در پروتئومیک نه تنها کلیه پروتئینهایی که در یک سلول در یک شرایط مشخص بیان می شوند، مورد بررسی قرار میگیرند، بلکه عملکرد و رفتار پروتئینها، برهمکنشهای بین پروتئینهای مختلف، آرایشهایی که پس از ترجمه بر روی پروتئینها ایجاد می شود و نیمه عمر آنها در سلول نیز مورد بررسی قرار میگیرد. در واقع پروتئومیک از سه بخش تشکیل شده است:

۱- مشخص کردن کلیه پروتئینهایی که در سلول بیان می شود: در این بخش، کلیه پروتئینهایی که در سلول تحت یک شرایط معین (مانند، حالت استراحت، رشد، تمایز، بیماری، تأثیر دارو و...) مشخص می شود. به این ترتیب می توان پروتئینهایی که در شرایط مختلف بیان می شوند یا میزان بیان آنها تغییر می کند را شناسایی کرد و به عملکرد آنها پی برد. شناسایی این پروتئینها در تشخیص بیماری و بررسی روند پیشرفت یا بهبودی آنها و همچنین شناسایی داروهای جدید، مفید می باشد.

۲- نقشه برداری برهمکنشهای بین پروتئینی: پروتئینها در سلول بصورت منفرد عمل نمیکنند و اغلب تأثیر خود را با همکاری پروتئینهای دیگر و برهمکنش با آنها اعمال می کنند. نمونه بارز برهمکنشهای پروتئینی، در مسیرهای انتقال پیام و مسیرهای بیوسنتزی مشاهده می شود. با شناسایی این برهمکنشها، بطور کارآمدتری می توان عملکرد و رفتار پروتئینها را مشخص کرد.

۳- نقشه برداری آرایشهای پروتئینی : اغلب پروتئینها پس از ترجمه متحمل آرایشهای مختلفی مانند، گلیکوزیله شدن، متیله شدن، استیله شدن، فسفریله شدن و... می شوند. این آرایشها بر فعالیت و عملکرد پروتئینها، همچنین ساختار فضایی، پایداری و نیمه عمر آنها تأثیر می گذارد. بسیاری از داروها گروههای الکتروفیلی دارند که از طریق آنها به پروتئین هدف متصل شده و اثر خود را اعمال می کنند. شناسایی این آرایشها، تأثیر آنها بر عملکرد پروتئینها و شرایطی که منجر به این آرایشها می شود، به شناسایی رفتار و عملکرد پروتئینها کمک می کند.

منابع:

۱. ابزارهای بیوانفورماتیک نویسندگان: فاطمه هنری و حبیب اسلام زاده
برای اطلاعات بیشتر در رابطه با این کتاب میتوانید به آدرس www.ibp.ir مراجعه کنید

2. A practical guide to: bioinformatics by: mohammad ali malboobi & tahmineh lohrasebi

برای اطلاعات بیشتر در رابطه با این کتاب میتوانید به آدرس www.ibp.ir مراجعه کنید

۳. مطالب موجود در سایت پرتال بیوانفورماتیک ایرانیان از نوشته های بابک باباعباسی

منابع زیاد دیگری در تهیه این کتاب الکترونیک استفاده شده است که در بخش معرفی کتب پرتال بیوانفورماتیک ایرانیان میتوانید مطالعه کنید www.ibp.ir
