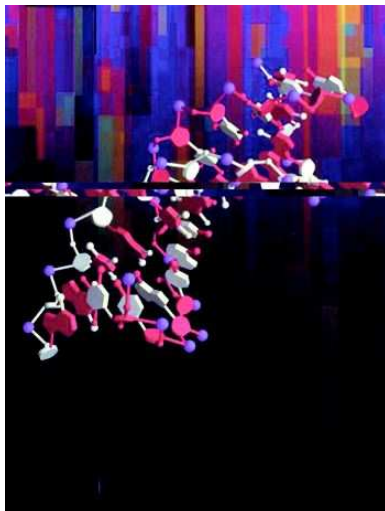


# برآیند علوم زیستی و محاسباتی بیوانفورماتیک ابزاری برای تحلیل حیات در سطح مولکولی

مهدی صارمی فر



وظیفه اصلی بیوانفورماتیک طراحی سیستم های کامپیوتری و مدل های ریاضی برای نگهداری، مدیریت و تحلیل داده های بیولوژیکی است.

ابزارهای بیوانفورماتیک در آینده قادر خواهند بود که با دانستن یک توالی ژنی، نحوه بروز صفت در سلول زنده را توصیف کنند.

بیوانفورماتیک دانشی بین رشته ای است. نقطه اتصال ریاضیات، علوم محاسباتی و علوم زیستی. این شاخه از علم با توجه به گسترش روزافزون حیطه های علمی و درگیر شدن آنها با هم و طرح مسائل مشترک بین شاخه هایی از علم که پیش از این چندان ربطی به هم نداشتند از یک طرف و گسترش توجه به زیست شناسی و به خصوص زیست شناسی مولکولی در دو دهه اخیر، بسیار مورد توجه پژوهشگران ریاضی و زیست شناسان قرار گرفته است. اگر بخواهیم موضوع علم بیوانفورماتیک را در یک جمله خلاصه کنیم، می توان گفت: «طراحی سیستم های کامپیوتری و مدل های ریاضی برای نگهداری، مدیریت و تحلیل مجموعه عظیمی از داده های بیولوژیکی و همچنین ارائه دانش زیستی» موضوع بیوانفورماتیک است. محتوای این «مجموعه عظیم داده ای» چیست؟ بیوانفورماتیک با چه مسائلی درگیر است؟ چه جایگاهی در کنار علوم دیگر دارد؟

با توجه به گسترش روزافزون و توجه به این شاخه در کشورمان در این گزارش نگاهی داریم به کلیات بیوانفورماتیک.

## • توالی ها

اطلاعات وراثتی هر سلول در مولکول های طولی به نام «دیوکسی ریبونوکلیک اسید» یا DNA نگهداری می شود. DNA پلیمری است که از چهار مونومر ساخته شده است. این مونومرها که باز هستند به اختصار A, T, C, و G نامیده می شوند. به این ترتیب برای نمایش مولکول DNA دانشمندان از یک رشته طولانی از این چهار حرف استفاده می کنند. به این رشته طولانی، توالی می گویند. در واقع DNA زنجیره ای دوگانه و به هم تابیده است، که هر دو مونومر از هر زنجیره با مونومری از زنجیره مقابل، در هم جفت می شوند، به این صورت که A با T و G با C جفت می شوند. پس با دانستن توالی یک زنجیره، توالی متقابل به دست می آید. در درون هسته هر سلول، مولکول DNA چندین بار تاب می خورد و برای استحکام بیشتر، به دور گلوله هایی از چربی حلقه می زند. به این مجموعه کروموزوم می گویند. برای مثال در بدن انسان ۴۶ کروموزوم وجود دارد که همه خصلت های انسانی در آنها کد شده اند.

اما مولکول DNA به تنهایی قادر به کنترل فعالیت های حیاتی موجود زنده نیست. بلکه هر کدام از قسمت های این دنباله باید طی فرآیندی که روخوانی و ترجمه نام دارد، در درون اندامی از سلول به نام ریبوزوم به یک آنزیم یا پروتئین معادل آن تبدیل شوند. این آنزیم یا پروتئین است که برای مثال وارد جریان خون می شود و به وظایف محوله فرد می پردازد. به هر قسمتی از یک مولکول DNA که یک پروتئین را کد می کند، یک ژن می گویند. یک مولکول DNA حاوی هزاران ژن است که هر کدام وظیفه ای خاص خود دارند، به این ترتیب جریان اطلاعات که توصیف کننده رفتار و عملکرد هر سلول زنده است، در DNA کنترل می شود. پروتئین ها نیز زنجیره ای از آمینواسیدها هستند. بیش از بیست نوع آمینواسید وجود دارد که با نسبت دادن یک علامت به هر کدام، توالی پروتئینی نشان داده می شود. هر سه تا باز روی DNA یک آمینواسید را کد می کنند. به این ترتیب از روی یک توالی ژنی، می توان توالی پروتئینی خاص آن را به دست آورد. به این ترتیب نخستین و مهمترین شباهت زیست شناسی مولکولی و علوم کامپیوتر نمایان شد: توالی ها، داده هایی دیجیتالی هستند. قسمت عمده ای از داده هایی که در مدل ها، الگوریتم ها و پایگاه های داده ای (Data base) بیوانفورماتیک مورد تحلیل قرار می گیرند، به توالی ها اختصاص دارد.

## • پایگاه های داده

جریان اطلاعاتی را که منجر به بروز یک واکنش خاص توسط یک سلول زنده یا یک موجود چند سلولی می شود را در دیاگرام زیر می توان خلاصه کرد:

توالی ژنی، توالی RNA، توالی پروتئین، ساختار پروتئین، عملکرد پروتئین، بروز صفت در سلول زنده.

براین اساس، هدف نهایی بیوانفورماتیک این است که با در دست داشتن توالی DNA های یک سلول یا یک موجود زنده، تمام خصوصیات و رفتارهای آن را پیش بینی کند. یکی از محرک های اصلی برای فعالیت روزافزون در زمینه بیوانفورماتیک پروژه های ژنوم و پروتئوم بوده اند. پروژه های ژنوم، با هدف تعیین کد توالی DNA های موجودات زنده تعریف شده اند که مهمترین و بزرگ ترین آنها پروژه ژنوم انسانی است. این پروژه که یک برنامه بزرگ جهانی بود و دانشمندان از سراسر جهان در آن شرکت داشتند، وظیفه اش شناخت کامل DNA انسان است. این پروژه یکی از چهار پروژه عظیم جهانی است که البته به لطف ساخت دستگاهی به نام PCR که به این فرآیند سرعت بخشید، در سال ۲۰۰۲ و سه سال پیش از موعد پانزده ساله پیش بینی شده اولیه، به پایان رسید. پایگاه های داده ای اینترنتی مخصوص این کار توالی های به دست آمده را در اختیار پژوهشگران قرار می دهند. این پایگاه ها کاملاً عمومی هستند و با وارد کردن هر توالی دلخواه می توان تمام توالی های مشابه به همراه مجموعه کاملی از اطلاعات مربوط به آنها را استخراج کرد. این پایگاه های بیوانفورماتیک حجم زیادی از اطلاعات ژنتیکی را در خود حفظ کرده اند و در طول مدت کوتاهی که از

راه اندازیشان می گذرد، به مهمترین ابزار پژوهشی در زیست شناسی مولکولی مبدل گشته اند.

سرعت رشد اطلاعات موجود در پایگاه های داده ای به صورت نمایی رشد می یابد. طوری که برای مثال در پایگاه Gen Bank هر ۱۴ ماه حجم اطلاعات دو برابر می شود. به طور مشابه برای پروتئین هم پروژه پروتئوم تعریف شده است. که البته در اینجا حجم کار به طرز بسیار وحشتناکی بالاتر است. در عین حال روش های خوبی مثل روش های به کار رفته در پروژه ژنوم در دست نیست ضمن اینکه بسیاری از پروتئین ها و آنزیم ها، ناشناخته مانده اند. هدف این پروژه ها تعیین توالی پروتئین ها و شکل ساختاری سه بعدی آنها است. به خصوص این آخری که نقش اساسی در عملکردهای پروتئین ها دارد. اگر پروژه ژنوم در طی یک فرآیند ۱۲ ساله به پایان رسید ما با توجه به حجم کار پروژه پروتئوم، به نظر می رسد که حدود یک قرن برای به پایان رساندن این پروژه که با همکاری اکثر مراکز پژوهشی جهان در حال انجام است، لازم باشد.

#### \* زمینه های مهم بیوانفورماتیک

##### ۱- تحلیل توالی های ژنوم

هدف اولیه بیوانفورماتیک طراحی روش های استخراج، نگهداری، پردازش و تحلیل تعداد بسیار زیادی از توالی ها بود. رسیدن به این هدف، برای محققان علوم زیستی دستاورد عظیمی به شمار می رود. به طور کلی در طی چند سال اخیر، کاوش در این پایگاه های داده ای برای پژوهشگران زیست شناسی مولکولی به یک فعالیت روزمره و نیاز حیاتی مبدل شده است. برای مثال فرض کنید که توالی قسمتی از یک DNA در آزمایشگاه به دست آمده است.

نخستین سؤالی که به ذهن می رسد این است که آیا این توالی در برگیرنده یک ژن هست یا نه؟ در صورت مثبت بودن جواب، این ژن در کجای زنجیره DNA اصلی قرار دارد و نهایتاً آنزیمی را که کد می کند چه نقشی در سلول یا در فرآیندهای حیاتی ایفا می کند؟ در غیاب بیوانفورماتیک و ابزارهای آن، ما وقت لازم است تا یک تیم تحقیقاتی به حدس های اولیه ای درباره پاسخ سؤالات فوق برسد. در حالی که تنها با یک کامپیوتر شخصی متصل به این پایگاه های داده ای ظرف چند دقیقه می توان به جواب قطعی یا حدس هایی محکم رسید.

سرعت بالای روش های تعیین توالی با روش های کامپیوتری و مدل های ریاضی در طراحی تراشه های DNA به دست آمده اند. دستگاه های فوق پیشرفته مجهز به تراشه های DNA قادر هستند ضمن تعیین توالی همزمان هزاران قطعه نوکلئوتیدی آنها را به طور خودکار در پایگاه های داده ای به ثبت برسانند.

##### ۲- پیش بینی ساختار سه بعدی (ساختار سوم و چهارم) پروتئین

کارکرد مولکول های عظیم پروتئین به شدت به شکل فضایی و ساختار سه بعدی آنها بستگی دارد. از طرفی همان گونه که دیدیم ژن ها نیز از طریق عملکرد پروتئین هایی که می سازند، نقش خود را اعمال می کنند. بنابراین شناخت کامل ماهیت و وظیفه ژن ها، منوط به دانستن اطلاعات کافی درباره پروتئین ها است. ولی پروژه های پروتئوم با وجود این اهمیت حیاتی، به کندی پیش می روند.

دلایل این کندی پیشرفت، هزینه های زیاد و کندی روند تعیین توالی پروتئین ها و مشکل بودن تعیین ساختار سه بعدی آنها در آزمایشگاه است. با توجه به سرعت بالای روند کار در پروژه های ژنوم، حل مسائل پروتئینی مهمترین چالش حال حاضر بیوانفورماتیک به حساب می آید.

دو اصل اساسی برای تعیین ساختار سه بعدی پروتئین از روی توالی آن وجود دارد که هر کدام روش جداگانه ای را برای حل مسئله ساختار پیشنهاد می کنند:

× پروتئین هایی که توالی نسبتاً مشابهی دارند، شکل فضایی شبیه به هم پیدا می کنند: جست و جو برای یافتن توالی های مشابه.

× شکل فضایی مولکول به نحوی است که به حداقل انرژی برسد: استفاده از قوانین شیمی، فیزیک و ترمودینامیک.

##### ۳- تحلیل کارکردی در سطح ژنوم

ابزارهای تحلیل کلان داده های زیستی، روش کار پژوهش های مهندسی ژنتیک، داروسازی و زیست شناسی را دگرگون کرده اند. فناوری جدید بیوانفورماتیکی امکانات جدید و بسیار قوی را فراهم ساخته است؛ مثل بررسی همزمان میزان فعالیت هزاران ژن در سلول، تحلیل نحوه تعامل تعداد زیادی پروتئین و تحلیل خصوصیات هزاران سلول جهش یافته در آن واحد. این مسائل با به کارگیری روش های آماری پیشرفته و کلاستر بندی حل شده اند. دانش مربوط به این بخش تحت عنوان «ژنوم شناسی کارکردی» به یکی از فعال ترین زمینه های تحقیقی در بیوانفورماتیک مبدل شده است.

از دستاوردهای مهم در این زمینه می توان به پیش بینی نقش و کارکرد ژن ها در سلول بدون نیاز به تحلیل داده های پروتئینی اشاره کرد.

##### ۴- ایجاد و مدیریت پایگاه های داده ای

صرف نظر از نوع داده های تولید شده در زیست شناسی مولکولی و نحوه تحلیل و تفسیر آنها، باید این داده ها را از طریق پایگاه هایی در اختیار پژوهشگران قرار داد. اما نحوه این ارائه هم مشکلات خاص خود را پیش رو دارد؛ مثل نحوه حصول اطمینان از درستی داده های ثبت شده و چگونگی نمایش مفید داده ها برای کاربران. از این جهت اداره کنندگان پایگاه های بزرگ بیوانفورماتیکی، چالش هایی بیش از یک مهندس پیش رو دارند.

##### ۵- مدل سازی ریاضی فرآیندهای حیات

استفاده کنندگان ابزارها و داده های بیوانفورماتیکی محدود به متخصصان زیست شناسی مولکولی نمی شود. گروهی که اخیراً به اهمیت بیوانفورماتیک پی برده اند، فیزیولوژیست ها هستند. آنها با استفاده از حجم عظیم داده های ژنومی و پروتئومی در تلاشند تا راه شبیه سازی فرآیندهای بیوشیمیایی سلول های زنده را هموار سازند. تلاش محققان این است که فرآیندهای خاص سلولی را شبیه سازی کرده و با یک پارچه سازی آنها به یک سلول کامل برسند که در این صورت یکی از هدف های مهم بیوانفورماتیک علوم زیستی محقق خواهد شد؛ یعنی درک کامل ساز و کار ارگانیسم های زنده در سطح مولکولی.

در خاتمه باید یادآور شد که اهمیت بیوانفورماتیک تنها در سرعت بخشیدن به کارهای آزمایشگاهی نیست بلکه گسترش این شاخه علمی و طرح و پاسخگویی به سؤالات جدید افق های نوینی را پیش روی زیست شناسان گشوده است.

منابع:

۱- [www.stratagene.com](http://www.stratagene.com)

۲- [www.pubmed.com](http://www.pubmed.com)

۳- [www.bioinformatic.org](http://www.bioinformatic.org)

۴- M. Campbell, L. Heyer; Genomics, Proteomics, and Bioinformatics