دانشگاه شهید بهشتی

دانشکده علوم ریاضی

پایان نامه ی کارشناسی ارشد

ریاضی محض

عنوان:

الگوریتمی نوین برای پیش‌بینی تراکنش بین پروتئین‌ها بر اساس ترکیب-قلمروهای لازم و کافی

استاد راهنما:

دکتر چنگیز اصلاح چی

نگارش:

منور فهلیانیان

تابستان ۱۳۸۸

۱۳۸۸/۱۱/-۶

۱۳۰۲۸۳

تاریخ .....
شماره .....
پیوست.....

دانشگاه شهید بهشتی

«بسمه تعالی»+

تهران ۱۹۸۳۹۶۳۱۱۳ اوین
تلفن: ۲۹۹۰۱

«صورتجلسه دفاع از پایان نامه دانشجویان دوره کارشناسی ارشد»

بازگشت به مجوز دفاع شمارهٔ ۲۰۰/۹۰۳/ت/د مورخ ۸۸/۳/۱۸ جلسه هیأت داوران ارزیابی پایان نامه: **خانم منور فهلیانیان** شماره شناسنامه: ۱۴۱۲۲ صادره از: **تهران** متولد: ۱۳۶۱ دانشجوی دوره کارشناسی ارشد: **ریاضی محض**

با عنوان:

الگوریتمی نوین برای پیش بینی تراکنش بین پروتئین ها بر اساس ترکیب- قلمروهای لازم و کافی

به راهنمایی:

آقای دکتر چنگیز اصلاح چی

طبق دعوت قبلی در تاریخ ۸۸/۴/۸ تشکیل گردید و بر اساس رأی هیأت داوری و با عنایت به ماده ۲۰ آئین نامه کارشناسی ارشد مورخ ۷۵/۱۰/۲۵ پایان نامه مزبور با نمره **۱۹/۲۵ (نوزده و بیست و پنج صدم)** و درجه **عالی** مورد تصویب قرار گرفت.

	مرتبه علمی	نام دانشگاه	امضاء
۱- **استاد راهنما:** آقای دکتر چنگیز اصلاح چی	دانشیار	شهید بهشتی	
۲- **داور:** آقای دکتر حمید پزشک	دانشیار	تهران	
۳- **داور:** آقای دکتر سهرابعلی یوسفی	دانشیار	شهید بهشتی	
۴- **مدیر گروه:** آقای دکتر علیرضا سالمکار	استادیار	شهید بهشتی	

الگوریتمی نوین برای پیش‌بینی تراکنش بین پروتئین‌ها بر اساس ترکیب-قلمروهای لازم و کافی

چکیده

مطالعهٔ عملکرد[۱] پروتئین‌ها جزء پژوهش‌های اصلی و مهم بیوانفورماتیک می‌باشد. مطالعه پروتئین‌هایی که از نظر دنباله و یا ساختار دارای شباهت هستند (همولوگ[۲]) اطلاعات مفیدی در مورد عملکرد پروتئین حاصل می‌کند، از آنجا که تمام پروتئین‌ها همولوگ‌های واضحی ندارند و پروتئین‌هائی که دنباله‌های شبیه به هم دارند ممکن است عملکردهای کاملاً متفاوتی داشته باشند بنابراین با مطالعه پروتئین‌ها صرفاً از نظر همولوژی موفقیت محدودی در تشخیص عمل پروتئین نصیب ما خواهد شد.

برای تشخیص اعمال ناشناخته پروتئین‌ها روش‌هایی وجود دارد که براساس هومولوژی نمی‌باشند؛ مانند *coexpression ,gen co – inheritance ,conserved gen position ,domain fusion* که یک تشابه عملکرد را بین پروتئین‌ها تشخیص می‌دهند.

یک جنبهٔ مشترک این روش‌ها این است که این اطلاعات وابسته به تراکنش پروتئین‌ها است، زیرا بیشتر فرآیندهای زیستی شامل یک یا چند تراکنش بین پروتئین‌ها می‌باشند و آشکار کردن تراکنش بین پروتئین‌ها یک روش مفید و کافی برای تعیین عملکرد پروتئین‌ها می‌باشد.

از اعمال مهم زیستی که شامل تراکنش پروتئین هستند می‌توان ماشینهای مولکولی[۳] وکمپلکس ساختاری [۴] را نام برد.

روش‌های آزمایشی برای آشکار کردن تراکنش‌های پروتئین ـ پروتئین وجود دارند، اما همراه با مشکلاتی است که به آن اشاره خواهیم کرد. به همین دلیل مسأله پیش بینی تراکنش بین پروتئین‌ها از طریق روش‌های محاسباتی مطرح می‌شود، ما در این پایان نامه به بررسی روش های مختلف جهت پیش‌بینی تراکنش بین پروتئین‌ها می‌پردازیم بدین منظور از مقالات زیر استفاده شده است:

- Jorome Wojcik and Vincent Schachter (۲۰۰۱). Protein-Protein interaction map inference using interacting domain profile pairs.(BIOINFOMATICS).

---

۱. Function
۲. Homolog
۳. Molecular machines
۴. Structural complex

- Wan Kyu Kim, Jong Park, Jung Keun Suh(۲۰۰۲)Large Scale Ststistical Prediction of Protein-Protein Interaction by Potentially Interacting Domain(PID). Genom Informatics ٤٢-٥٠.

- Dong-Soo Han, Hong-Soo Kim ,Woo-Hyuk Jang, Sung-Doke Lee and Jung-Keun Suh(۲۰۰٤)PreSPI: a domain combination based prediction system for protein protein interaction.(Nucleic Acids Research) ٦٣١٢-٦٣٢٠

- Yohan Kim, Mehment Koyuturk, Umut Topkara, Ananth Grama and Shankar Subramaniam(۲۰۰٥). Inferring functional information from domain co-evolution. (BIOINFOMATICS)

در آخر روشی نوین بنام قلمروهای واسطه را جهت تشخیص قلمروی سومی که واسطه ایی برای دو قلمروی دیگـر بـرای تراکنش پروتئین ها است بر مبنای ترکیب-قلمروهای لازم و کافی بیان خواهیم کرد. مقاله ی زیر بـر مبنـای ایـن روش استخراج شده است:

Discovering Domains Mediating Protein Interaction.

۱. مقدّمه ۱

۱. ۱ نگاهی بر زیست شناسی مولکولی ۱
۱. ۲ تعاریف و پیش‌نیازها ۱۱
۱. ۳ داده‌های آزمایشگاهی ۱۴
۱. ۳. ۱ چند روش آزمایشگاهی جهت تشخیص تراکنش بین پروتئین‌ها ۱۴
۱. ۳. ۲ کیفیت داده‌های آزمایشگاهی ۱۵

۲.روش *PID* ۱۷

۲. ۱ پیش‌درآمد ۱۷
۲. ۲ شرح روش *PID* ۱۸
۲. ۳ نتایج ۲۰

۳.روش *PCBP* ۲۴

۳. ۱ پیش درآمد ۲۴
۳. ۲ شرح روش *DCBP* ۲۵
۳. ۳ نتایج ۳۴

۴.روش *IDPP* ۳۷

۴. ۱ پیش درآمد ۳۷
۴. ۲ شرح روش *IDPP* ۳۸
۴. ۳ نتایج ۴۱

۵. روش ماتریس هم تکاملی ۴۵

۵. ۱ پیش درآمد ۴۵
۵. ۲ شرح روش ماتریس هم تکاملی ۴۶
۵. ۳ نتایج ۵۱

۶. روش قلمروهای واسطه ۵۳

۶. ۱ پیش درآمد ۵۳
۶. ۲ شرح روش قلمروهای واسطه ۵۴

۷. واژه نامه ۶۸

۸. مراجع ۷۴

۱

مقدّمه

۱.۱ ـ نگاهی بر زیست شناسی مولکولی

۱.۱.۲ ـ پروتئین

نام پروتئین که از کلمهٔ یونانی پروتوس[۱]به معنی اولیه اقتباس شده است؛ نخست توسط برزیلیوس[۲] به مولدر[۳] پیشنهاد شد. برزیلیوس این نام را به تمام مواد آلی نیتروژن دار کمپلکس که در سلول زنده یافت می‌شود اطلاق کرد.

پروتئین‌ها بخش اساسی و ستون فقرات ساختار سلول زنده را تشکیل می‌دهند،آنزیم‌ها که کلیهٔ واکنش‌های فیزیکی و شیمیایی را (که ادامه حیات سلول به آن‌ها بستگی دارد) به انجام می‌رساند؛ نیز تماماً ساختار پروتئین دارند.

برخی از پروتئین‌ها لایه محافظ موجود زنده را تشکیل می‌دهند، گروه دیگر هورمون‌ها و حاملین اکسیژن را به وجود می‌آورند.

---

۱. Protos
۲. Berzelivs
۳. Mulder

بعضی دیگر در انقباضات ماهیچه‌ای شرکت می‌کنند یا همراه با ژن‌ها در هسته سلول دیده می‌شوند و بالاخره بعضی نیز در اعمال ایمنی‌سازی، تشکیل پادزهر و دفاع بدن نقش مؤثری را به عهده دارند. در حقیقت پروتئین‌ها بخش اصلی یا در حدود سه چهارم مواد تشکیل دهندهٔ بافت‌های حیوانی را به وجود می‌آورند.

اسیدهای آمینه واحدهای ساختمانی پروتئین‌ها می‌باشند. اولین اسید آمینه در سال ۱۸۲۰ توسط *H.Braconnt* از هیدرولیز ژلاتین بدست آمد و گلایسین[۱] نام گرفت. از هیدرولیزپروتئین‌های طبیعی فقط بیست نوع اسیدآمینه بدست می‌آید که به اسیدهای آمینه استاندارد معروفند، از نظر ژنتیکی فقط برای این اسیدهای آمینه کد ژنتیکی وجود دارد.

نوع اسیدهای آمینه را نوع کدهای ژنتیکی و موقعیت این اسیدها را ترتیب قرار گرفتن کدهای ژنتیکی موجود در ژن تعیین می‌کنند.

فرمول عمومی اسیدهای آمینه به صورت $R-\underset{NH_3^+}{\overset{H}{\underset{|}{\overset{|}{C}}}}-Co\overline{o}$ است. به جز پرولین همگی دارای یک عامل کربوکسیل و یک عامل آمین آزاد می‌باشند که به کربن α متصل است، (NH_3 عامل آمین) ($-CO\overline{O}$ عامل کربوکسیل). اختلاف اسیدهای آمینه از یکدیگر در ساختمان شیمیایی بنیان R آن‌ها است اگر $R=H$ باشد ساده‌ترین اسیدآمینه یعنی:گلایسین به فرمول $NH_3-\underset{H}{\overset{H}{\underset{|}{\overset{|}{C}}}}-COOH$ بوجود می‌آید.

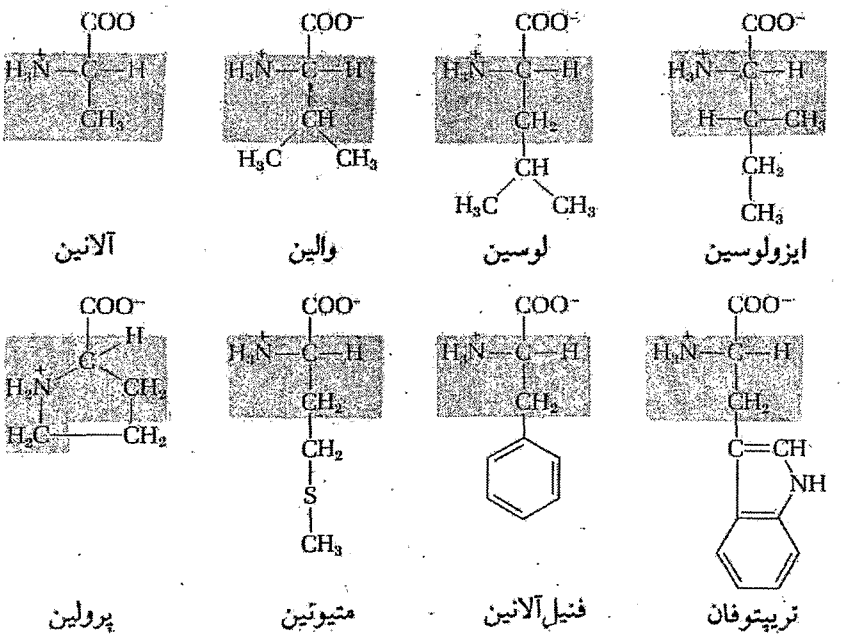
تمام اسیدهای آمینه استاندارد در گروه کربوکسیل و گروه آمین متصل به کربن α مشترک می‌باشند.

ساختار بیست اسید آمینه مختلف که معمولاً در پروتئین‌ها یافت می‌شوند در زیر نشان داده شده‌اند.

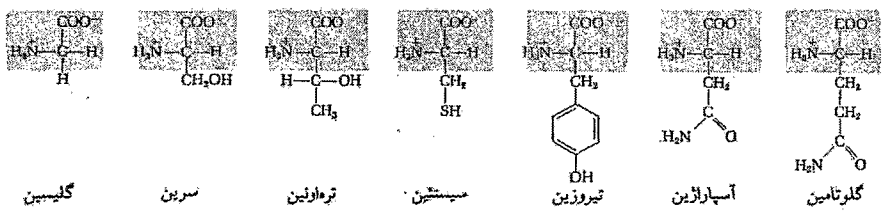
---

[۱]. Glycin

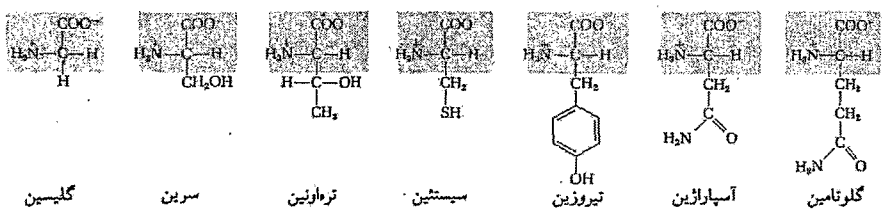
۱. ۲. ۱. ۱ ـ الف) اسید آمینه‌هایی که بنیان R در آن‌ها غیر قطبی (آبگریز) است.



۱. ۲. ۱. ۱ ـ ب) اسید آمینه‌هایی که بنیان R آن‌ها قطبی و غیر یونی است.



۱. ۲. ۱. ۱ ـ ج) اسید آمینه‌هایی که بنیان R آن‌ها قطبی و یونی است.



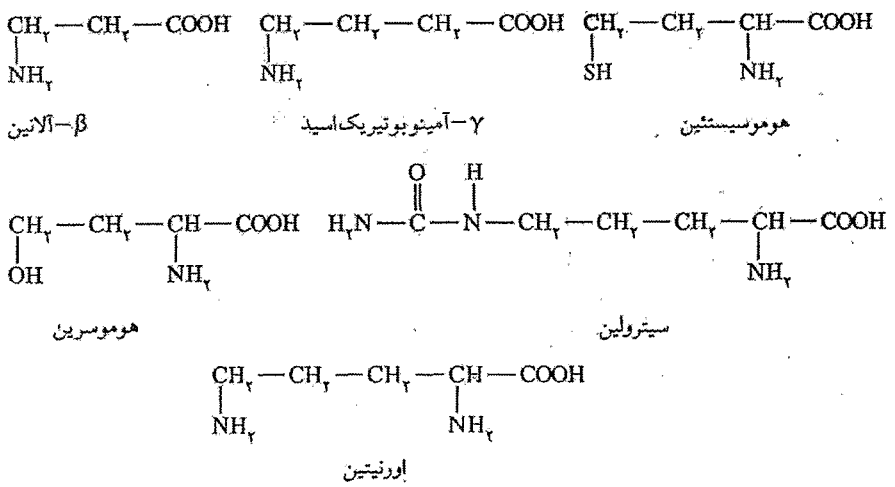
اسید آمینه‌ها را می‌توان به سه گروه مختلف تقسیم بندی کرد.

گروه اول شامل آن دسته از اسیدهای آمینه است که گروه جانبی R آن‌ها را یک گروه غیر قطبی یا آبگریز تشکیل می‌دهد،آلانین، والین،لوسین،ایزولوسین، متیونین، پرولین، فنیل آلانین و تریپتوفان را می‌توان در این گروه طبقه بندی کرد.

گروه دوم شامل آن دسته از اسیدهای آمینه است که گروه جانبی R آن‌ها یک گروه قطبی ولی غیر یونی است و با آب تولید پیوند هیدروژنی می‌کند. در نتیجه این آمینو اسیدها بیشتر در آب محلول‌اند. از این گروه می‌توان به گلایسین، سرین، تره اونین، سیستین، تیروزین، آسپاراژین و گلوتامین را نام برد.

گروه سوم شامل آن دسته از اسیدهای آمینه است که گروه جانبی R آن ها یونی است. مانند آسپارتیک اسید، گلوتامیک اسید، لیزین، آرژینین و هیستیدین.

علاوه بر بیست آمینو اسید فوق برخی از آمینو اسیدهای دیگر نیز به مقدار جزئی در ساختار پروتئین‌ها به کار می‌روند و بعضی دیگر نیز با وجودی که در ساختار پروتئین‌ها شرکت ندارند از نظر زیست شناختی دارای اهمیت فراوانی می‌باشند ساختار شیمیایی و نام این آمینو اسیدها در شکل ۲. ۲ . ۱. ۱ نشان داده شده است.



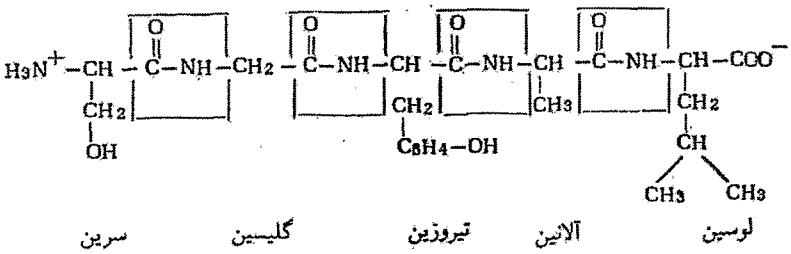
شکل : ۲. ۲ . ۱. ۱

از ترکیب دو اسید آمینه که با خارج شدن یک مولکول آب همراه است ساده‌ترین پپتید بنام دی پپتید بدست می‌آید. در این واکنش عامل کربوکسیل از یک اسید آمینه با عامل آمین از اسید آمینه دیگر ترکیب می‌شود.

پیوندی که بین این دو اسید آمینه بوجود می‌آید، پیوند پپتیدی نامیده می‌شود که در اثر شرکت عامل OH از گروه کربوکسیل یک اسید آمینه هیدروژن از عامل آمین اسیدآمینه دوم وجود می‌آید.

بدیهی است که اگر n مولکول اسید آمینه به همین روش با هم ترکیب شوند و از مجموع $(n-۱)$ مولکول آب خارج شود، محصول حاصل پلی پپتید نامیده خواهد شد. پروتئین‌ها در اثر آبکافت اسیدی یا بازی و هم چنین در اثر عمل آنزیم‌ها (پروتئازها) به پپتیدهای متفاوت تجزیه می‌شوند. به طور کلی تمام پپتیدها از ترکیب آمینو اسیدهای با اتصال پپتیدی بوجود آمده‌اند و برای نشان دادن آن‌ها قسمت آمین آزاد (N – انتهایی[۱]) را در طرف چپ صفحه و قسمت کربوکسیل آزاد (C – انتهایی[۲]) را در طرف راست قرار می‌دهند برای نام‌گذاری از سمت آمین انتهای شروع کرده و سپس آمینو اسیدهای موجود در پروتئین را به ترتیب با افزودن لفظ ایل به آخر آن‌ها نام برده و در انتها نام آمینو اسید مربوط به کربوکسیل آزاد را ذکر می‌کنیم.

برای مثال یک پنتاپپتید (سریل گلیسیل تیروزنیل آلانیل لوسین) در شکل ۳. ۲ . ۱. ۱ نشان داده شده است.



شکل: ۳. ۲ . ۱. ۱

[۱]. N-terminal
[۲].C-terminal

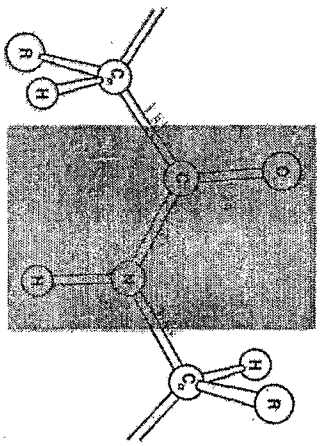
هر پروتئین پلی پپتید است ولی هر پلی پپتید الزاماً پروتئین نیست؛ زمانی پروتئین خوانده می‌شود که عاملی برای بروز یک پدیده زیستی باشد.

اگر از هیدرولیز پروتئین‌ها منحصراً اسیدآمینه بدست‌آید،پروتئین را ساده و اگر علاوه بر اسیدهای آمینه، ترکیب‌های آلی یا معدنی بدست آید؛ پروتئین را مرکب می‌گویند[۳۴].

۱ .۱ .۲ .۱ ـ پیکربندی پروتئین‌ها

یکی از ویژگی‌های مهم پروتئین‌ها ساختار سه بعدی آنها است. پروتئینی که این پیکربندی خاص را از دست می‌دهد و به صورت یک رشته در می‌آیدو یا زنجیره‌های پپتیدی آن بدون نظم و ترتیب ویژه قرار بگیرد فاقد فعالیت زیست شناختی است، به عبارت دیگر فعالیت حیاتی یک پروتئین به ساختار فضائی آن بستگی دارد و ترتیب قرار گرفتن آمینو اسیدها که خود تعیین کننده این نوع پیکر بندی است حایز اهمیت بسزائی است.

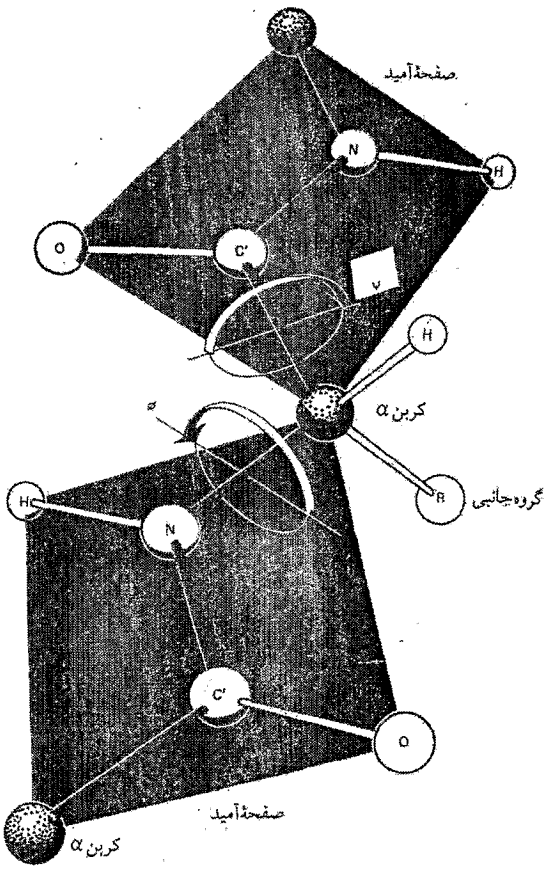
در سال ۱۹۳۰ لینوس پاوکینگ و روبرت کوری مطالعه ساختار پروتئین‌ها را با پرتوی ایکس آغاز کردند و ثابت نمودند که گروه پپتیدی ($-\overset{O}{\overset{\|}{C}}-\underset{H}{\underset{|}{N}}-$) یک گروه مسطح و محکم و هیدروژن گروه $-NH-$ همیشه نسبت به اکسیژن گروه کربونیل به شکل ترانس است(شکل : ۱. ۱. ۲ . ۱. ۱).



شکل : ۱. ۱. ۲ . ۱. ۱

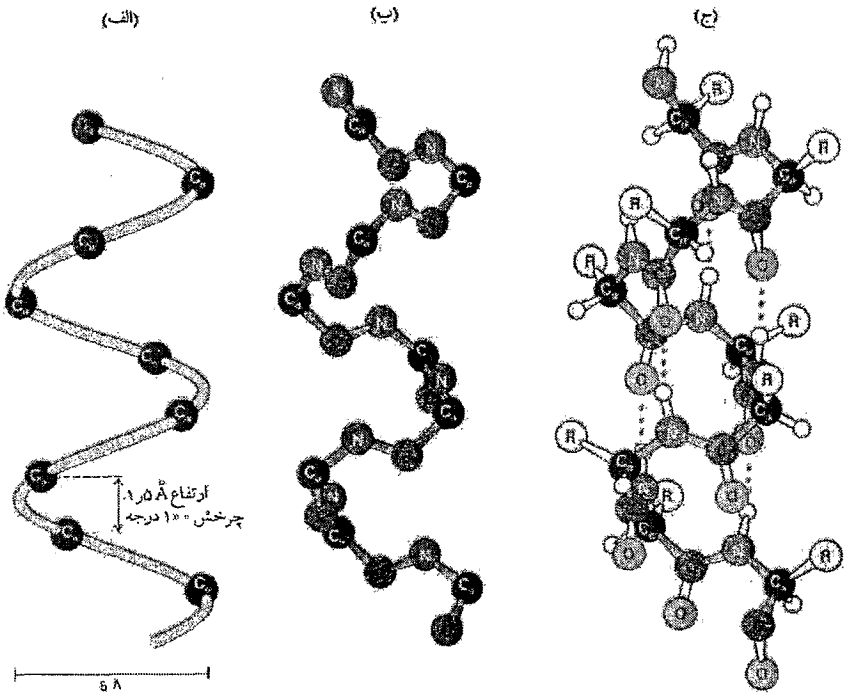
اتصال بین اتم کربن گروه کربونیل و اتم نیتروژن گروه پپتیدی $-\overset{O}{\overset{\|}{C}}-N-$ دارای خاصیت پیوندهای دوگانه است و آزادی چرخش در اطراف این پیوند وجود ندارد. بر عکس،پیوندهای بین دو اتم کربن آنها در طرفین گروه پپتیدی یک پیوند کوالانس یگانه محسوب می‌شود و در نتیجه آزادی چرخش در این دو پیوند وجود خواهد د اشت.

چرخش اطراف این دو پیوند را با زوایای $(C_\alpha - C)\psi$ و $(C_\alpha - N)\phi$ نشان می‌دهند و با اندازه‌گیری آن‌ها می‌توان پیکربندی ثانوی یک پروتئین را تعیین و تعریف کرد(شکل : ۲. ۱. ۲ . ۱. ۱).



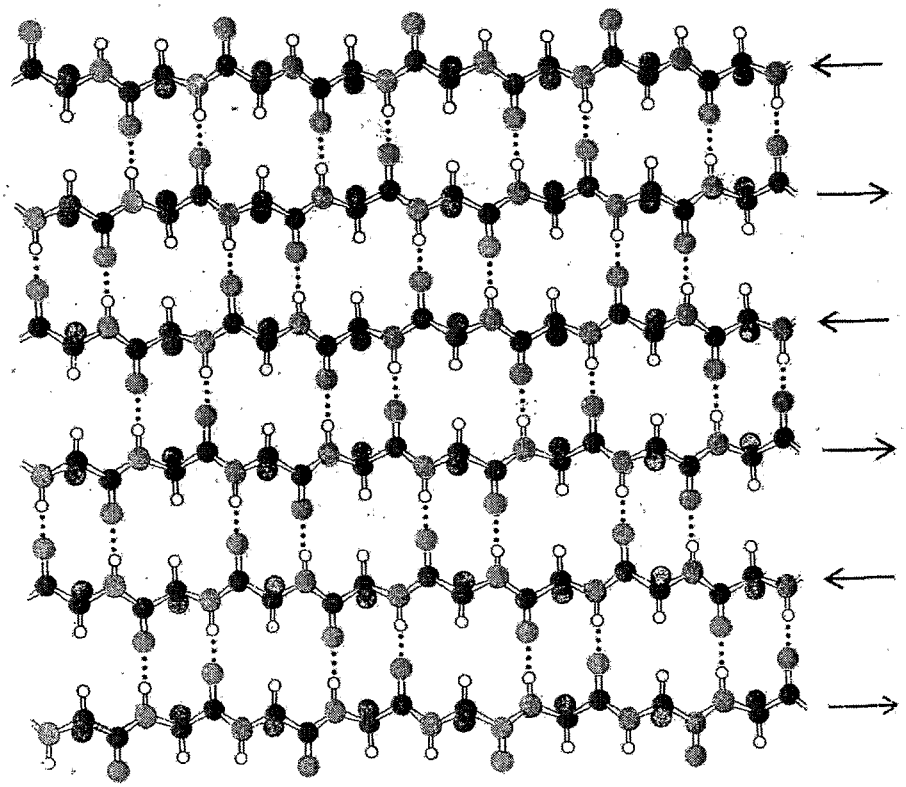
شکل : ۲. ۱. ۲ . ۱. ۱

اصولاً زوایای ψ و ϕ می‌توانند در محدودهٔ ۱۸۰ – تا ۱۸۰ + درجه تغییر کنند، لذا کلیهٔ ساختارهای ممکن برای یک پروتئین با توجه به تغییرات این زوایا، قابل پیش‌بینی است. لازم به تذکر است که به دلیل برخورد فضایی اتم‌های تشکیل دهندهٔ ستون فقرات یک پروتئین با یکدیگر و هم‌چنین برخورد فضایی زنجیرهای کناری آمینو اسیدها، بسیاری از ساختارهای پیش‌بینی شده امکان‌پذیر نمی‌باشند و تنها تعداد محدودی از آن‌ها وجود خارجی می‌یابند، برای مثال در صورتی که آمینو اسیدهای متشکله یک پروتئین از لحاظ ساختار زیاد متفاوت نباشند و زوایای ψ و ϕ به ترتیب۴۷– و ۵۷ – درجه باشند چنین پروتئین یک مارپیچ آلفا را به وجود می‌آورد (شکل : ۳. ۱. ۲ . ۱. ۱).



شکل : ۳. ۱. ۲ . ۱. ۱

برعکس در صورتی که زوایای ψ و ϕ به ترتیب ۱۳۵ + و ۱۳۹– درجه باشند چنین پروتئینی به صورت صفحات بتای چین‌دار در خواهد آمد(شکل : ۴. ۱. ۲ . ۱. ۱)[۳۴].



شکل : ۴. ۱. ۲. ۱. ۱

۵ .۱ .۲ .۱ .۱ ـ ساختار اول پروتئین‌ها

آمینو اسیدها می‌توانند به طرق مختلف باهم ترکیب شده و پپتیدهای متفاوت را تولید کنند برای مثال اگر یک هپتاپپتید که از هفت آمینو اسید تشکیل شده در نظر گرفته شود هر یک از این هفت محل را یکی از ۲۰ آمینواسید معمولی ممکن است اشغال کند که در این صورت می‌توان احتمال وجود یک میلیار نوع هپتاپپتید را تصور کرد[۳۴].

۶ .۱ .۲ .۱ .۱ـ ساختار دوم پروتئین‌ها

در اثر ترکیب آمینو اسیدها با یکدیگر رشته پلی پپتیدی حاصل می‌شود که نمی‌تواند بصورت منظم و کلاف مانند دور هم جمع شود، بلکه هر قسمت از رشته پلی‌پپتیدی دارای طرح منظمی است که دارای شکل هندسی می‌باشد، مجموع این خصوصیات را اصطلاحاً ساختار دوم پروتئین می‌نامند.

سه نوع ساختاردوم درپروتئین‌ها مشاهده می‌شود که شامل مارپیچ آلفا، صفحات چین‌دار بتا و مارپیچ کلاژن می‌باشد لازم به تذکر است که نوع آمینواسیدهای موجود در یک رشتهٔ پلی پپتیدی است که طرح ساختار ثانوی پروتئین را به یکی یا مجموعه‌ای از سه شکل ذکر شده تعیین می‌کند[۳۴].

۱.۱.۲.۱.۷ـ ساختار سوم پروتئین‌ها

باتوجه به چگونگی ساختار دوم پروتئین ومخصوصاً ایجاد تا خوردگی معکوس مولکول روی خودش (شکست بتا)قسمت های معین ومشخص در مولکول پروتئین به وجود می آید که به واحد قلمرو پروتئین شناخته شده است.اثر متقابل این واحدها بر یکدیگروهم چنین نحوهٔ قرار گرفتن شاخه های جانبی نسبت به یکدیگر وبرقراری ارتباط بین آن ها باعث میشود که ملکول پروتئین شکل سه بعدی خاصی به خود بگیرد که به ساختار سوم پروتئین معروف است[۳۴].

۱.۱.۲.۱.۸ ـ ساختار چهارم پروتئین‌ها

این ساختار فقط چگونگی قرار گرفتن رشته‌های پلی پپتیدی به یکدیگر را در فضا مشخص می‌کند. از نظر پیکربندی دو، سه، چهار و یا چندین رشته پلی‌پپتیدی در فضا می‌توانند در کنار هم قرار گرفته و بوسیله اتصالات غیرکوالانسی به یکدیگر متصل شوند[۳۴].

۱.۲ ـ تعاریف و پیش‌نیازها

۱.۲.۱ ـ تراکنش[۱]

تراکنش پروتئین[۲] یک اصطلاح جامع و فراگیر است که برای بیان رابطهٔ وسیعی بین پروتئین‌ها استفاده می‌شود.

تراکنش پروتئین‌ها می‌تواند رابطه بهم پیوستن[۳] مانند تراکنش فیزیکی بین پروتئین‌ها یا بین پروتئین‌ها و دیگر ملکول‌ها، یا رابطهٔ تنظیم کننده[۴] که یک پروتئین نسخه برداری[۵] پروتئین دیگر را تنظیم می‌کند، یا رابطه‌های فسفریلیت کردن[۶]، که یک پروتئین پروتئین دیگر را فسفریلیت می‌کند، را بیان کند.

هم چنین تراکنش پروتئین‌ها ممکن است ارتباط‌های عملکردی[۷] را بیان کند،یعنی دو پروتئین را تراکنش یافته گویند زمانی که دارای عملکردهای یکسان و یا عملکردهای مرتبط باشند[۲۵].

۱.۲.۲ ـ پروتئین‌های همولوگ[۸]

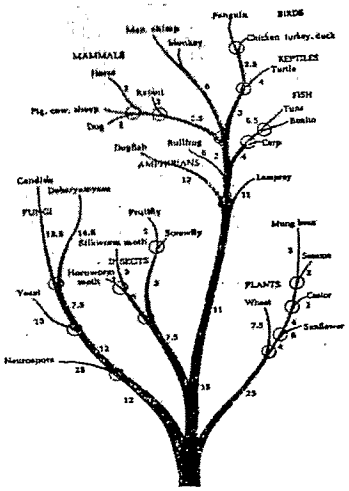
توالی یابی اسیدهای آمینه در انسولین، نه تنها راهی برای پی بردن به ترتیب قرار گرفتن اسیدهای آمینه در سایر پروتئین‌ها را گشود، بلکه نشان داد که توالی اسیدهای آمینه در پروتئین‌های همولوگ در گونه‌های مختلف یکسان می‌باشد و احتمالاًپروتئین‌هایی که بدون توجه به محل فعالیت و منشأ آن‌ها عمل یکسان در موجودات دارند، مشتق از یک نوع پروتئین اجدادی می‌باشند. به پروتئین‌هایی که نقش بیولوژیکی آنها در گونه‌های مختلف موجودات یکسان باشد پروتئین‌های همولوگ می‌گویند. هموگلوبین نمونه‌ای از این پروتئین‌ها است که در تمام مهره‌داران وجود دارد و نقش آن در انتقال اکسیژن است، طول رشته‌های پلی پپتیدی این پروتئین‌ها در گونه‌های مختلف تقریباً یکسان است. به اسیدهای آمینه

---

۱. Interaction
۲. Protein interaction
۳. Concreterlationship
۴. Regulatory relationship
۵. Transcription
۶. Phosphorilation
۷. Function association
۸. Homolog

که در پروتئین‌های همولوگ موقعیت ثابت دارند، اسیدهای آمینه غیرقابل تغییر[۱] می‌گویند، سایر اسیدهای آمینه که ممکن است از گونه‌ای به گونه دیگر تغییر یابند، اسید آمینه‌های متغییر[۲] را تشکیل می‌دهند.

بهترین نمونه برای بیان همولوگ، سیتوکروم C می‌باشد که در میتوکندری تمام جانداران نقش انتقال الکترون را بر عهده دارد. وزن مولکولی تمام سیتوکروم C های مطالعه شده تقریباً یکسان و برابر ۲۵۰۰ دالتون است. این وزن مولکولی معرف حدود ۱۰۰ واحد اسید آمینه در مولکول می‌باشد. توالی یابی اسیدهای آمینه در متجاوز از ۶۰ گونه جانوری، گیاهی و میکروبی نشان داده است که ۲۷ درصد از اسیدهای آمینه در تمام سیتوکروم Cها موقعیت یکسان دارند و اختلاف آن‌ها از یکدیگر در اسیدهای آمینه متغییر است که از گونه ای به گونه دیگر تغییر می‌کند و این تغییر در تعداد اسیدهای آمینه، رابطه مستقیم با تغییر فیلوژنتیک موجودات دارد به این معنی که هر چه فاصله فیلوژنی دو گونه از یکدیگر بیشتر شود؛ به همان نسبت درصد اختلاف در اسید آمینه‌های سیتوکروم آن‌ها نیز افزایش می‌یابد، مثلاً بین اسب و مخمر که از نظر فیلوژنی تفاوت بسیار زیاد است، ۴۸ درصد از اسیدهای آمینه سیتوکروم آن‌ها نیز متفاوت از یکدیگرند، در حالی که اختلاف سیتوکروم C مرغ و اردک از نظر فیلوژنی بسیار نزدیک به هم می‌باشند و اختلاف فقط در اسید آمینه آنها است و یا در مورد گاو، گوسفند و خوک اختلافی دیده نمی‌شود[۳۴].



شکل : ۱. ۲ . ۲. ۱

---

۱. Invarent
۲. Variable

۱.۲.۳ ـ کمپلکس[۱]

گروهی از پروتئین‌هارا که همراه هم یک وظیفه خاص سلولی را انجام می‌دهند کمپلکس می نامند، نمونه معروف از این پروتئین‌ها فتی اسید سنتتاز[۲] است که دارای هفت نوع آنزیم است ولی به صورت یک پروتئین کمپلکس مسئول بیوسنتز اسیدهای چرب می‌باشند[۳۴].

۱.۲.۴ ـ قلمرو[۳]

مفهوم قلمرو اولین بار توسط *Wetlavfer* مطرح شد و طبق تعریف او قلمرو یک قسمت از دنباله پروتئین است که عمل خاص خود را دارد و مستقل از بقیهٔ زنجیر پروتئین می باشد. هر قلمرو یک ساختار سه بعدی فشرده را تشکیل می‌دهد، یک پروتئین ممکن است چندین قلمرو داشته باشدو یک قلمرو ممکن است در چندین پروتئین دیده شود. ترکیبی از قلمروها در یک پروتئین تمام عمل پروتئین را مشخص می کند؛ تراکنش فیزیکی پروتئین‌ها[۴] نیز توسط ساختار قلمروهایشان کنترل؛ می شود. معمولاً طول هر قلمرو بین ۲۵۴ آمینو اسید تا ۵۰۰ آمینو اسیداست[۳۴].

---

۱. Complex
۲. Fattyacid synthetase
۳. Domain
۴. Physical interaction

۱.۳ ـ داده‌های آزمایشگاهی

یک مشکل اصلی در زیست‌شناسی تشخیص عملکرد پروتئین‌ها است، زمانی که دنباله پروتئینی برای اورگانیسم‌های مختلف مشخص می‌شود عملکرد پروتئین‌ها و اینکه چگونه با یکدیگر برای انجام کاری فعالیت می‌کنند به طور کامل مشخص نمی‌شود. یک منبع مهم از اطلاعات برای بیان و درک این مسأله داده‌هایی است که تراکنش پروتئین‌ها را مشخص می‌کند[۲۵] .

روش‌های آزمایشی جهت تشخیص تراکنش بین پروتئین‌ها وجود دارند که به برخی از آنها اشاره خواهیم کرد، اما این روش‌ها بسیار هزینه بر و وقت گیر می‌باشند، بطوری که تعداد پروتئین‌های انسانی در حدود ۳۵۰۰۰ تخمین زده شده است که به وضوح برای تشخیص تراکنش بین آن‌ها وقت و هزینه زیادی نیاز است[۲۵].

۱.۳.۱ ـ چند روش آزمایشگاهی جهت تشخیص تراکنش بین پروتئین‌ها

- *Co. immunoprecipitation*

در این روش پروتئین مورد نظرتوسط پادتن خاصی جدا شده و پروتئین‌های تراکنش یافته با آن، از طریق *western blotting* مشخص می‌شوند. تراکنش‌هایی که توسط این آزمایش تشخیص داده‌شده اند به عنوان تراکنش‌های واقعی تلقی می‌شوند، اگر چه این روش فقط تراکنش‌هائی را مشخص می‌کند که ذهنیت قبلی در مورد تشکیلشان وجود داشته باشد.

- *Pull – down*

این روش نوعی از روش *immunoprecipitation* محسوب می‌شود و تفاوتش در این است که به منظور تسخیر پروتئین کمپلکس‌ها به جای پادتن از لیگاند استفاده می‌شود.

- *Abel transfer*

این آزمایش جهت غربال و تأیید کردن تراکنش بین پروتئین ها می‌باشد و اطلاعاتی در مورد لایه‌هایی که تراکنش در آن اتفاق افتاده است تأمین می‌کند. هم‌چنین این روش می‌تواند تراکنش‌های زودگذر و ضعیفی که تشخیص‌شان مشکل است را آشکار کند.

- *Two-hybrid*

این روش تراکنش‌ها را بین پروتئین‌هائی که به طور مصنوعی با هم آمیخته شده‌اند، در هستهٔ مخمر مشخص می‌کند، با استفاده از این روش تعداد زیادی تراکنش بدلی تشخیص داده می‌شود، بنابراین لازم است با استفاده از *Co-immunoprecipitation* تراکنش‌ها بررسی شوند.

- *TAP*[۱]

این آزمایش دارای خروجی بالا می‌باشد و بر خلاف *two – hybrid* دقتش را می‌توان با آزمایش‌های با مقیاس کم مقایسه کرد. در این آزمایش تراکنش‌ها در محیط واقعی سلول بوسیلهٔ *co- immunoprecipition* تشخیص داده می‌شوند.

۲ .۳ .۱ ـ کیفیت داده‌های آزمایشگاهی

از مشکلات داده‌های آزمایشگاهی شلوغی[۲] و ناقص[۳] بودن آن‌ها می‌باشد، برای مثال تخمین زده شده است که حداقل نیمی از تراکنش‌های گزارش شده بوسیلهٔ روش *two – hybrid* در مخمر نادرست هستند.

یک روش برای برخورد با شلوغ بودن داده‌های تراکنش های فیزیکی[۴] نسبت دادن یک وزن به هر تراکنش می باشد که نشان دهندهٔ میزان قابل اعتماد بودن آن تراکنش است [۲۵] یعنی برای هر تراکنش بین یک زوج پروتئین u و v به صورت زیر وزن $r_{u,v}$ نسبت داده می شود:

$$r_{u,v} = 1 - \prod_{i \in E_{u,v}} (1 - r_i)^{n_{i,u,v}}$$

که r_i میزان قابل اعتماد بودن آزمایش iام و $E_{u,v}$ مجموعه تمام آزمایشاتی است که تراکنش بین u و v در آن مشاهده شده و $n_{i,u,v}$ برابر است با تعداد مرتبه ایی که تراکنش بین u و v در آزمایش iام دیده شده است[۲۴].

---

۱. Tandem affinity purification
۲. Noise.
۳. Incomplete
۴. Physical interaction data.