بسم الله الرحمن الرحيم

١٠٨٢٣٢

دانشکده مهندسی

پایان نامه کارشناسی ارشد در رشته مهندسی کامپیوتر (نرم افزار)

بررسی کاربردهای کشف قوانین وابستگی در کشف بر هم کنش های پروتئین – پروتئین

توسط

نوشین عمرانیان

۱۳۸۷/۹/۲۳

استاد راهنما

دکتر محمد هادی صدرالدینی

شهریور ماه ۱۳۸۷

۱۰۸۲۳۲

به نام خدا

بررسی کاربردهای کشف قوانین وابستگی در کشف بر هم کنش های پروتئین - پروتئین

به وسیله‌ی

نوشین عمرانیان

پایان نامه

ارائه شده به تحصیلات تکمیلی دانشگاه به عنوان بخشی
از فعالیت های تحصیلی لازم برای اخذ درجه کارشناسی ارشد

در رشته‌ی

مهندسی کامپیوتر (نرم افزار)

از دانشگاه شیراز

شیراز

جمهوری اسلامی ایران

ارزیابی شده توسط کمیته پایان نامه با درجه: عالی

دکتر محمد هادی صدرالدینی، استادیار بخش مهندسی و علوم کامپیوتر (رییس کمیته).....

دکتر علی نیازی، استادیار بخش بیوتکنولوژی کشاورزی.....

دکتر منصور ذوالقدری جهرمی، دانش یار بخش مهندسی و علوم کامپیوتر.....

شهریور ۱۳۸۷

تقدیم به تمامی معلمان،

پدر و مادر مهربانم

و

همسرم که در تمامی مراحل همراهم بود...

سپاسگزاری...

شکر و سپاس الطاف بیکران آفریننده آسمانها و زمین و انسان را، که مـرا نعمـت وجـود ارزانـی داشت و امانتی نزد پدر و مادری مهربانم کرد تا زندگی را با درک معنای شگرف آن تجربه کنم.
پدری که گرمای وجودش، سرمای روزگار را از یادم برد و عصاره محبتی به نام مادر، که مهرش و دعای خیرش، همراه و معنابخش زندگیم و بخشنده شور و امید در وجودم است.
سپاس خدای را که مرا خواهرانی از جنس محبت و شور، همراهم کرد، تا طی مسیر زندگیم، بـا همراهیشان لذت بخش تر گـردد. مـریم کـه همـاره اخـلاق خـوش و بزرگـواریش و سـارا کـه شادابی‌اش، زندگی را شوری دو چندان می بخشد.
و همسری که آمدا از جنس عشق و محبت و با اراده محکم خود، اعتماد به نفسی دوباره به من بخشید و دریچه های دیگری از جنس نور، محبت، خوشبختی و پیشرفت را بر روی من گشود.
مهربانا شکر...
و به نمایندگی از تمامی معلمان زندگیم، از آغاز تا کنون، از اسـتاد راهنمـای گرانقـدرم جنـاب آقای دکتر صدرالدینی و اساتید مشاورم، جناب آقای دکتـر نیـازی و دکتـر ذوالقـدری و دکتـر بوستانی و نماینده محترم تحصیلات تکمیلی، جناب آقای دکتر گشتاسبی کمال تشکر را دارم.
همچنین از همه دوستان همکلاسی و کارکنان بخش علوم و مهندسی کامپیوتر نیز کـه در ایـن مدت یار و همراه من بودند، سپاس گزارم...

نوشین عمرانیان
۱۳۸۷/۶/۳۰

چکیده

بررسی کاربردهای کشف قوانین وابستگی در کشف بر هم کنش های پروتئین – پروتئین

به وسیله‌ی

نوشین عمرانیان

بر هم کنش های پروتئین–پروتئین از واکنش های اصلی بیوشیمیایی می باشند، که در بافت های زنده رخ می‌دهند و به دلیل این که فرایندهای بیولوژیکی از طریق آنها مشخص می‌شوند، نقش بسیار مهمی را در موجودات ایفا می‌کنند. بنابراین شرح جامعی از برهم‌کنش های پروتئین–پروتئین، به طور قابل توجهی در فهم رویدادهای بیولوژیکی به ما کمک می‌کند.

اخیرا بر هم کنش‌های پروتئین–پروتئین، به وسیله‌ی روش مخمر دو هیبریدی مورد آزمایش قرار می گیرند. در نتیجه، مقدار زیادی اطلاعات درباره‌ی بر هم کنش‌های پروتئین–پروتئین، روی هم انباشته شده است. با این وجود، اطلاعات یا دانش کلی در ارتباط با برهم کنش ها، چندان زیاد نمی باشد. تکنیکی که اطلاعات یا دانش مفید پنهان را، از میان حجم زیاد داده، استخراج می کند، داده کاوی نام دارد. این تکنیک توجه بسیار زیادی را به خود جلب کرده و تحقیقات زیادی نیز در زمینه‌ی، به کارگیری داده‌کاوی در بیوانفورماتیک پیش از این انجام شده است. به همین منظور در این مطالعه به بررسی و تحقیق و توسعه یک سیستم، برای کشف دانش یا اطلاعات سودمندِ مربوط به بر هم‌کنش های پروتئین–پروتئین، از داده های روی هم انباشته شده‌ی بر هم کنشی پروتئین–پروتئین، با استفاده از الگوریتم "کشف قوانین وابستگی" که یکی از روش های رایج و مشهور داده‌کاوی می باشد، پرداخته شده است.

یکی از الگوریتم های معروف در زمینه‌ی کشف قوانین وابستگی، الگوریتم کشف قوانین وابستگی مبتنی بر گراف، به نام "ایجاد محموعه اقلام تکراری جهت دار" می باشد. در این مطالعه، از این الگوریتم به دلیل داشتن سرعت پردازش بالا، نسبت به الگوریتم های ساده دیگر، استفاده شده است. این روش، تنها یک بار از روی پایگاه داده حجیم بر هم کنشی خوانده و با استفاده از تکنیکی که در این مطالعه برای فشرده سازی داده ها، جهت کم کردن حافظه مصرفی به کار رفته، داده ها را به صورت فشرده، ذخیره می‌کند. همچنین برای بالاتر بردن سرعت، عملیات را بر روی داده‌های فشرده شده، اعمال می‌کند.

از آنجاییکه مهمترین مطلب برای بیولوژیست ها، کیفیت قوانین وابستگی می باشد، از تکنیک حذف قوانین درجه دو، برای بالاتر بردن کیفیت قوانینِ تولید شده، استفاده شده است. همچنین به جهت افزایش داده های بر هم کنشی و تغییر و به‌روزرسانی این داده ها، به معرفی تکنیکی برای پشتیبانی از قوانین وابستگی کشف شده، پرداخته شده است. نتایج به دست آمده، نشان می دهد که این روش، سرعت پردازش را در زمان به روزرسانی داده های بر هم کنشی، کاهش می‌دهد. همچنین نتایج نشان می‌دهد که، حافظه مصرفی روش پیشنهادی در این مطالعه، بسیار کمتر و سرعت پردازش داده ها، بیشتر از الگوریتم های ساده‌ی قبلی می باشد. به علاوه مهمترین نتیجه به دست آمده از این تحقیق، کاهش تعداد و افزایش کیفیت قوانین وابستگی استخراج شده، نسبت به روش های قبلی می‌باشد.

فهرست مطالب

عنوان	صفحه
فصل اول: مقدمه و کلیات	۱
۱-۱- پروتئین	۲
۱-۲- ساختار پروتئین ها	۳
۱-۳- سنتز پروتئین	۳
۱-۳-۱- مراحل سنتز پروتئین	۴
۱-۳-۱-۱- نسخه برداری	۴
۱-۳-۱-۲- ترجمه	۵
۱-۴- بر هم کنش های پروتئین- پروتئین	۸
۱-۵- دلایل اهمیت و روش های مطالعه بر هم کنش پروتئین – پروتئین	۹
۱-۶- سیستم دو هیبریدی در مخمر	۱۰
۱-۷- بیوانفورماتیک	۱۲
۱-۸- داده کاوی	۱۳
۱-۹- مراحل داده کاوی	۱۶
۱-۱۰- روش های داده کاوی	۱۷
۱-۱۱- کاوش قواعد وابستگی	۱۸
۱-۱۲- تعریف مسئله کاوش قواعد وابستگی	۱۹
۱-۱۳- الگوریتم ای. پریوری	۲۰
۱-۱۴- بهینه سازی های انجام شده بر روی الگوریتم ای. پریوری	۲۲
۱-۱۵- شمارش پویای مجموعه اقلام	۲۳
۱-۱۶- روش اف. پی. گروت	۲۳
۱-۱۷- ساختمان داده Trie	۲۴
۱-۱۸- روش مبتنی بر گراف	۲۵
۱-۱۹- کشف قوانین وابستگی اولیه DLG	۲۷
۱-۱۹-۱- ایجاد گراف وابستگی	۲۷
۱-۱۹-۲- ایجاد الگوی وابستگی اولیه	۲۹

۱-۲۰- پشتیبانی و نگهداری از قوانین وابستگی کشف شده ۳۰

فصل دوم: مروری بر تحقیقات انجام شده ۳۲

۲-۱- کشف برهم کنش پروتئین-پروتئین ۳۳
۲-۲- الگوریتم کشف قوانین وابستگی ۳۸

فصل سوم: روش انجام کار ۴۲

۳-۱- تعریف داده ها ۴۳
۳-۱-۱- داده های برهم کنشی ۴۴
۳-۱-۲- خصوصیات پروتئین های مخمر ۴۵
۳-۲- آماده سازی داده برای کاوش ۴۷
۳-۳- قوانین مورد توجه برای استخراج ۴۹
۳-۴- معیار سنجش قوانین استخراج شده ۵۱
۳-۴-۱- درصد پشتیبان ۵۱
۳-۴-۲- درصد اطمینان ۵۲
۳- ۵- کشف قوانین وابستگی ۵۵
۳- ۵-۱- الگوریتم کشف قوانین وابستگی مبتنی بر گراف ۵۵
۳-۵-۱-۱- فاز ایجاد مجموعه های تک قلمی تکراری ۵۶
۳-۵-۱-۲- فاز ساختن گراف ۵۷
۳-۵-۱-۳- فاز تولید مجموعه اقلام تکراری ۵۷
۳-۶- بهینه سازی حافظه در راستای استفاده از داده های برهم کنشی ۶۱
۳-۶-۱- فشرده سازی داده ۶۱
۳-۶-۲- استفاده از DFS به جای BFS ۶۳
۳-۷- بهینه سازی سرعت در راستای استفاده از داده های برهم کنشی ۶۴
۳-۸- بهینه سازی کیفیت در راستای استفاده از داده های برهم کنشی ۶۵
۳-۹- نگهداری و پشتیبانی از قوانین وابستگی کشف شده ۶۶

فصل چهارم - نتایج، بحث و پیشنهادات ۶۹

۴-۱-فرضیات تحقیق ۷۰
۴-۲- الگوریتم های پیاده سازی شده ۷۱

۴-۳- قوانین کشف شده ۷۲
۴-۴- بستر سخت افزاری ۷۲
۴-۵- بررسی نتایج ۷۳
۴-۵-۱- بررسی نتایج از دیدگاه حافظه اشغال شده ۷۳
۴-۵-۲- بررسی نتایج از دیدگاه سرعت اجرایی ۷۴
۴-۵-۳- بررسی نتایج از دیدگاه کیفیت ۷۶
۴-۵-۴- بررسی نتایج از دیدگاه به روز رسانی داده های برهم کنشی ۷۶
۴-۶- نتیجه گیری ۷۷
۴-۷- کار آینده ۷۷

فصل پنجم – فهرست منابع ۷۹

واژه نامه ۸۳

فهرست جدول ها

شماره و عنوان **صفحه**

جدول ۱ - برخی از فعالیت های زیستی پروتئین ها ۲
جدول ۲ - یک پایگاه داده نمونه ۱۸
جدول ۳ - نمادها و اسامی متغیر های استفاده شده در شکل ۵ ۲۱
جدول ۴ - پایگاه داده‌ای از تراکنش‌ها ۲۸
جدول ۵ - پایگاه داده تراکنشی ۵۶
جدول۶ - بردارهای بیتی اقلام تکراری در جدول ۵ ۵۶
جدول ۷ - اطلاعات مربوط به جدول ۵ که برای الگوریتم *DLG ذخیره شده است. ۵۹
جدول ۸ - پایگاه داده‌ای از تراکنش‌ها ۶۲
جدول ۹ - میزان حافظه اشغال شده برای حالت استخراج قوانین فیلتر نشده ۷۳
جدول۱۰ - میزان حافظه اشغال شده برای حالت استخراج قوانین فیلتر شده ۷۳
جدول ۱۱ - پیاده سازی حالت استخراج قوانین فیلتر نشده ۷۴
جدول ۱۲ - پیاده سازی برای حالت استخراج قوانین فیلتر شده ۷۴

فهرست شکل ها

شماره و عنوان **صفحه**

شکل ۱ – ساختار شماتیک ار. ان. ای. انتقال دهنده ۶
شکل ۲ – ساختار کلی سنتز پروتئین در هسته داران ۷
شکل ۳ – نمای کلی بر هم کنش دو پروتئین ۸
شکل ۴ – سیستم دو هیبریدی در مخمر ۱۱
شکل ۵ - الگوریتم ای. پریوری ۲۱
شکل ۶ : تولید مجموعه اقلام کاندید در الگوریتم ای. پریوری ۲۲
شکل ۷ - مثالی از یک Trie ۲۵
شکل ۸ – گراف وابستگی بدست آمده از فاز ایجاد گراف ۲۸
شکل ۹ – بستر کاری استفاده شده توسط اویاما، برای کشف قوانین وابستگی مربوط به بر هم کنش پروتئین-پروتئین. داده‌ها برای کاوش از روی داده‌خای بر هم کنشی و پایگاه داده‌های ژنومی، ساخته می‌شوند. ۴۳
شکل ۱۰ – تراکنش‌های مبتنی بر پروتئین، برای استخراج قوانین بر هم کنشی دو پروتئین، مناسب نیستند. بنابراین از بر هم کنش‌های مبتنی بر بر هم کنش استفاده می‌شود. ۴۷
شکل۱۱ - کپی برداری مجدد از بر هم کنش ها، با جابجایی ویژگی‌های پروتئین-های سمت راست و چپ. ۴۹
شکل ۱۲ – اگر چه سه قانون اول، شامل سه بر هم کنش مثبت می‌شوند، اما برهم کنش نوع ۱، امتیاز بالاتری نسبت به بقیه دارد. ۵۴
شکل ۱۳ – با افزایش تعداد پروتئین‌هایی که به برهم‌کنش‌های مثبت، متصل نیستند، امتیاز کاهش مییابد. ۵۴
شکل ۱۴ – گراف وابستگی بدست آمده از فاز ایجاد گراف. ۵۷

فصل اول

مقدمه و کلیات

فصل ۱ - مقدمه و کلیات

۱-۱- پروتئین

پروتئین ها فراوان ترین پلی مرهای زیستی در طبیعت هستند، که بیش از ۵۰٪ از وزن خشک سلول ها را تشکیل می دهند. در سلول های موجودات زنده، صدها نوع پروتئین مختلف وجود دارد، که از نظر ساختار و عمل متفاوت‌اند. این درشت مولکول ها در تمام سلول ها و همه بخش های یک سلول مشاهده می شوند و بدین ترتیب در سیستم های زیستی از گوناگونی گسترده ای برخودارند. این گوناگونی، اعمال مختلفی را به آن ها نسبت داده است. جدول ۱ فعالیت های مهم پروتئین ها را در سلول نشان می دهد.

جدول ۱ - برخی از فعالیت های زیستی پروتئین ها (ربانی چادگانی، ۱۳۷۸)

فعالیت	مثال
آنزیمی	الکل دهیدروژنازها در تخمیر الکلی
انتقالی	هموگلوبین در انتقال کسیژن خون
ذخیره ای	فریتین ذخیره کننده آهن
ساختاری	کلاژن در بافت همبند
حرکتی	میوزین در رشته های ضخیم بافت ماهیچه ای
حفاظتی	فیبرینوژن در لخته شدن خون
تنظیمی	انسولین در تنظیم قند خون- هورمون های رشد

پروتئین ها درشت مولکول هایی هستند، که از اتصال کووالانسی یک سری واحدهای ساختاری، به نام اسیدهای آمینه ساخته شده‌اند. تعداد، ماهیت و طرز قرار گرفتن این واحدها در زنجیره‌ی پروتئینی، ساختار خاص و رفتار ویژه آن ها را تعیین می کند. ۲۰ نوع اسید آمینه مختلف وجود دارد که، در ساختار پروتئین ها قرار می گیرند.

اتصال بین اسیدهای آمینه متوالی، منجر به ایجاد زنجیره پروتئینی می گردد.

۱-۲- ساختار پروتئین ها

پروتئین ها بسیار متنوع اند و هر یک از آن ها فعالیت ویژه ای دارند. تنوع پروتئین ها به ترتیب قرار گرفتن اسیدهای آمینه در مولکول پروتئین و ساختارهای ویژه موکول وابسته است. به طور کلی، برای پروتئین ها چهار سطح ساختاری در نظر گرفته می شود، که شامل ساختارهای اول تا چهارم است. همه پروتئین ها ساختارهای اول تا سوم را دارند ولی ساختار چهارم در تعداد محدودی از پروتئین ها مشاهده می شود. قرار گیری خطی اسیدهای آمینه، منجر به ایجاد ساختار اول پروتئین ها می گردد.

ایجاد پیوند بین هیدروژنی اسیدهای آمینه نزدیک به هم در زنجیره پروتئینی، منجر به ایجاد ساختار دوم می گردد. عموماً ۲ نوع ساختار منظمِ مارپیچ آلفا و صفحات بتا، ساختار دوم را در پروتئین ها تشکیل می دهند.

در ساختار سوم پروتئینی، که ساختار کروی شکل است، پیوند بین اسید آمینه های دور از هم (علاوه بر پیوندهای هیدروژنی، پیوندهای دی سولفیدی و غیره نیز دیده می شود)، منجر به قرار گیری قسمت های آبگریز در درون و قسمت های آب دوست، در سمت خارج کره می‌گردند.

با آنکه اکثر پروتئین ها در ساختار خود، منحصراً شامل یک زنجیره پلی پپتیدی هستند، برخی از آن ها از تعداد دو یا چند زنجیره پلی پپتیدی ساخته شده اند، که این حالت ساختار چهارم در پروتئین ها را نشان می دهد.

به طور خلاصه می توان گفت که ساختار اول، ترتیب و ردیف اسیدهای آمینه‌ی سازنده پروتئین را نشان می دهد. ساختار دوم، شامل دو ساختار منظمِ مارپیچ آلفا و صفحات بتاست، ساختار سوم درباره چگونگی پیچ و تاب خوردن زنجیره‌ی پروتئینی و فشرده شدن آن ها بحث می کند. ساختار چهارم هم در ارتباط با وضعیتِ زیر واحدهای پروتئین های مجاور هم و ایجاد یک مولکول فعال با تمامی ویژگی های آنهاست (ربانی چادگانی، ۱۳۷۸).

۱-۳- سنتز پروتئین

دی. ان. ای.[1] به طور مستقیم در سنتز پروتئین دخالت ندارد، بلکه بیشتر شبیه مدیری عمل می کند، که انجام کارهای لازم را به گروهی از کارکنان محول می نماید. زمانی که سلول

[1] - Deoxyribonucleic Acid (DNA)

به یک پروتئین خاص احتیاج داشته باشد، ناحیه رمزگذار این پروتئین در دی. ان. ای. شناسایی می شود و به صورت یک نوع اسید نوکلوئیک دیگری به نام آر. ان. ای.[1] کپی برداری می شود. این آر. ان. ای. که از روی قسمت کوچکی از دی. ان. ای. کپی برداری شده است، به عنوان الگو و به طور مستقیم در سنتز پروتئین دخالت دارد. در هر ثانیه هزاران تبدیل اطلاعات از دی. ان. ای. به پروتئین در هر سلول بدن ما در حال رخ دادن است. بنابراین در هر سلول بدن ما، جریان اطلاعات ژنتیکی از دی. ان. ای. به آر. ان. ای. و از آنجا به پروتئین می باشد. تمام سلول ها از باکتری ها گرفته تا سلول های بدن انسان، اطلاعات ژنتیکی را بدین طریق بیان می کنند. این فرآیند کلی به دلیل اهمیت، تحت عنوان اصل مرکزی در زیست شناسی مولکولی نام نهاده شده است. (ربانی چادگانی، ۱۳۷۸)

۱-۳-۱- مراحل سنتز پروتئین

۱-۳-۱-۱- نسخه برداری

سنتز پروتئین، طی دو گام اصلی انجام می گیرد، که یکی نسخه برداری[2] و دیگری ترجمه[3] است. نخستین گام سنتز پروتئین، نسخه برداری بخشی از مولکول دی. ان. ای. به مولکولی از آر. ان. ای. پیام رسان است. این فرایند را آنزیم آر. ان. ای. پلیمراز انجام می دهد. این آنزیم، به توالی نوکلئوتیدی ویژه ای از دی. ان. ای. به نام پروموتر[4] (راه انداز) می پیوندد، راه انداز در ناحیه ای جلوتر از (در فرادست) ژنی قرار دارد، که باید نسخه برداری شود. تعدادی آنزیم، باز شدن موضعی مارپیچ دی. ان. ای. را بر می انگیزد و به آر. ان. ای. پلیمراز امکان می دهند، که نسخه برداری یکی از دو رشته ی دی. ان. ای. را آغاز کند. در هر ژن، فقط یکی از دو رشته‌ی دی. ان. ای. نقش الگو ایفا می کند و به آر. ان. ای. پیام رسان، نسخه برداری می شود. این رشته‌ی دی. ان. ای. را اصطلاحاً رشته‌ی مخالف مفهوم، می نامند. رشته‌ای که نسخه برداری نمی شود (و توالی آن همانند توالی آر. ان. ای. است، جز آنکه به جای باز U، باز A دارد) رشته‌ی مفهوم، نامیده می شود. در ژن های دیگری از همان مولکول دی. ان. ای.، ممکن است رشته‌ی دیگر مولکول (رشته مقابل الگو) به عنوان الگوی ساخته شدن آر. ان. ای. به کار رود. اما در یک ژن، آنزیم آر. ان. ای. پلیمراز، برای نسخه برداری مولکول آر. ان. ای.، در یک امتداد پیش می رود. یعنی از یک رشته‌ی دی. ان. ای. به رشته‌ی دیگر آن، نمی جهد. هنگامی که

---

[1] - Ribonucleic Acid (RNA)
[2] - Transcription
[3] - Translation
[4] - Promoter

آر. ان. ای. پلیمراز، به انتهای یک ژن ساختاری می‌رسد و در آنجا به نوعی«توالی نوکلئوتیدی پایانگر» برخورد می‌کند، نسخه برداری پایان می‌پذیرد. در بعضی از ژن های باکتریایی، نوعی پروتئین کمکی به توالی پایانگر متصل می شود و از این طریق به جدا شدن آر. ان. ای. پلیمراز از دی. ان. ای.، کمک می کند. در یوکاریوت ها، مکانیسم پایان یافتن نسخه برداری هنوز مجهول است.

در سلول های یوکاریوتی، نسخه های اولیه‌ی آر. ان. ای. پیام رسان، پیش از آنکه به صورت مولکول های آر. ان. ای. پیام رسان بالغ، هسته را ترک کنند، مورد «پردازش» قرار می‌ گیرند. بیشتر نسخه‌های اولیه‌ی آر. ان. ای. پیام رسانِ یوکاریوتی، ابتدا به صورت موزائیکی از منطقه‌های کدکننده و منطقه‌های غیرکدکننده هستند (منطقه های کدکننده را «اگزون[1]» و منطقه های غیرکدکننده را «اینترون[2]» می نامند). پیش از آنکه آر. ان. ای. پیام رسان، هسته سلول را ترک کند و به صورت آر. ان. ای. پیام رسان بالغِ سیتوپلاسمی در آید، باید اینترون‌های آن به دقت حذف و اگزون ها به هم وصل شوند. افزون بر آن، یک نوکلئوتید گوانین دارِ غیر عادی (به نام کلاهک) به انتهای ′۵ آر. ان. ای. پیام رسان هسته ای و یک رشته از نوکلئوتیدهای آدنین دار (به نام دمِ پلی –A) به انتهای ′۳ آن متصل می شود. اما در سلول‌های پروکاریوتی، غشای هسته وجود ندارد و آر. ان. ای. پیام رسان، مورد پردازش قرار نمی‌گیرد. به استثنای آرکئوباکتریهای ابتدایی، ژن‌های باکتری‌ها اینترون ندارند. بنابراین، در باکتری‌ها، حتی پیش از کامل شدن نسخه برداریِ مولکول آر. ان. ای. پیام رسان از دی. ان. ای.، ترجمه‌ی پیامِ این مولکول به پروتئین، می تواند آغاز شود.

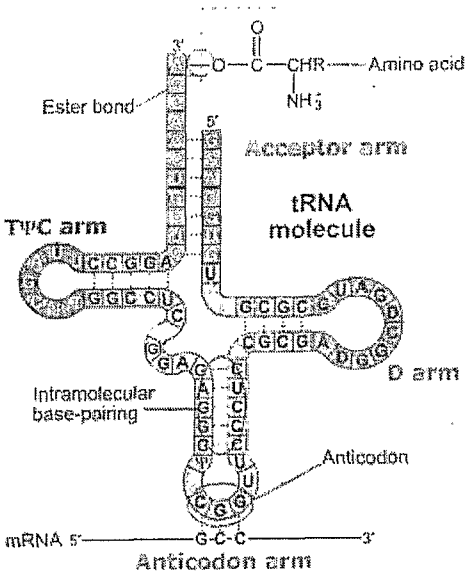
۱-۳-۱-۲- ترجمه

در جریان دومین گام سنتز پروتئین، ریبوزوم ها و کمپلکس‌های آر. ان. ای. انتقال دهنده متیونین (به نام متیونیل آر. ان. ای. انتقال دهنده «باردار»)، نزدیک انتهای ′۵ مولکول آر. ان. ای. پیام‌رسان، در نخستین کدون شروع، یا کدون آغاز (′۳ AUG ′۵) متصل می‌شوند و ترجمه‌ی توالی ریبونوکلئوتیدی این آر. ان. ای. پیام رسان را به توالی آمینواسیدیِ مولکول پروتئین، آغاز می‌کنند. هر ریبوزوم از ۳ نوع ملوکول آر. ان. ای. ریبوزومی و در حدود ۵۰ نوع پروتئین، ساخته شده است. هر نوع آمینواسید، دست کم به وسیله‌ی یک نوع مولکول آر. ان. ای. انتقال دهنده، شناسایی می شود. از آنجا که کد ژنتیک می تواند برای یک اسید آمینه به صورت چندگانه باشد، عملاً بیش از ۲۰ نوع آر. ان. ای. انتقال دهنده، در سنتز

---

[1] - Exon
[2] - Intron

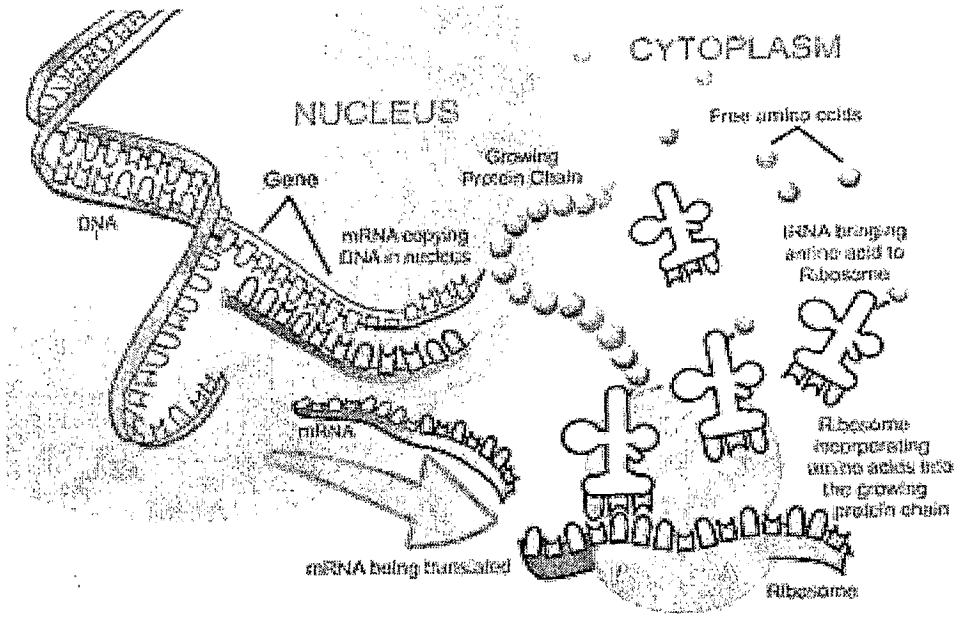
پروتئین دخالت دارند. هر آمینواسیدی با کمک آنزیم معینی، یعنی آمینواسیل سنتتاز، (از پایانه‌ی کربوکسیل خود) به انتهای ′۳ مولکول آر. ان. ای. انتقال دهنده‌ی ویژه‌ی خود متصل می شود، که به اصطلاح ظرفیت آن «پر می شود» (شکل ۱). از این رو، دست کم بیست نوع آنزیم سنتتاز در سلول وجود دارد. آر. ان. ای. انتقال دهنده با ظرفیت «پر شده» را فعال شده یا باردار می‌نامند. نزدیک وسط مولکول آر. ان. ای. انتقال دهنده، حلقه‌ای از بازهای جفت نشده وجود دارد، که محتوی نوعی توالی سه نوکلئوتیدی به نام آنتی کدون (ضد رمزینه) است. گمان می رود که در جریان پروتئین سازی، بخش های دیگر مولکول آر. ان. ای. انتقال دهنده، با بازهای مکمل خود در مولکول آر. ان. ای. ریبوزومِ ریبوزوم جفت می شوند، یا به صورت جایگاه های شناسایی برای آمینواسیل سنتتاز ویژه‌ای عمل می کنند.



شکل ۱ – ساختار شماتیک ار. ان. ای. انتقال دهنده

ترجمه‌ی همه‌ی مولکول‌های پروتئین با کدون شروع که ′۳ AUG′۵ است و آمینواسید متیونین را تعیین می کند، آغاز می شود (در تمامی بحث بعدی به شکل ۲ مراجعه شود). برای آر. ان. ای. انتقال دهنده‌ی فعال شده، دو جایگاه روی ریبوزوم وجود دارد. یکی جایگاه پپتیدیل (جایگاه P) و دیگری جایگاه آمینواسیل (جایگاه A) است. آر. ان. ای. انتقال دهنده‌ی متصل به متیونینِ آغازیِ فعال شده (احتمالاً با گذشتن از جایگاه A) وارد جایگاه P می شود. آنتی کدون ′۵ UAC′۳ مولکول آر. ان. ای. انتقال دهنده، با کدون مکمل خود در مولکول آر. ان. ای. پیام رسان، یعنی ′۳ AUG′۵ جفت می شود. ریبوزوم مانند نوعی لرزه میز عمل می‌کند و در جریانِ ترجمه، همه‌ی اجزای واکنش کننده را، در ردیف درستی نگه می دارد. آر. ان. ای. انتقال دهنده‌ی فعال شده‌ی دیگری (مثلاً حامل ترئونین) به جایگاه A (باز هم از طریق

جفت شدن بازهای کدون- آنتی کدون، به نحوِ ویژه‌ای) وارد می‌شود. بین دو آمینو اسیدِ مجاورِ هم، یک پیوند پپتیدی تحت تاثیر نوعی پروتئین آنزیمیِ جزء ریبوزوم، به نام پپتیدیل ترانسفراز تشکیل می شود. هنگامی که پیوند پپتیدی تشکیل می‌شود، پیوند امینواسیلی کـه میتـونین را به آر. ان. ای. انتقال دهنده‌اش متصل می‌کـرده اسـت، مـی‌گـسلد. متیونیـل آر. ان. ای. انتقـال دهنده که اینک «بی بار» شده است، جایگـاه P را تـرک مـی‌کنـد (و معمـولاً بـارِ دیگـر فعـال می‌شود). در این هنگام ریبوزوم به اندازه‌ی سه نوکلئوتیـد، در امتـداد مولکـول آر. ان. ای. پیـام رسان جا به جا می شود و در نتیجه کدون آزاد دیگری در جایگاه خالی A قرار مـی گیـرد و در همین حال، آر. ان. ای. انتقال دهنده‌ی حامل آمینواسید ترئـونین[1] را (کـه اکنـون بـه یـک دی پپتید متصل است) از جایگاه A به جایگاه P منتقل می کند. اکنون آر. ان. ای. انتقال دهنـده‌ی فعال شده‌ی سومی (مثلاً، حامل فنیل آلانـین) وارد جایگـاه A مـی شـود. در ایـن حالـت بـین دومین و سومین آمینواسید، پیوند پپتیدی تشکیل می‌شود و دومین آر. ان. ای. انتقال دهنـده، جایگاه P را ترک می‌کند. جابه‌جا شدن ریبوزوم در امتداد آر. ان. ای. پیام رسان، دومین کـدون را برای آرژینین در جایگاه A آزاد می‌سـازد، در حـالی کـه آر. ان. ای. انتقـال دهنـده‌ی حامـل فنیل آلانین [2] را (که اکنون به یک تری پپتید متصل است) از جایگـاه A بـه جایگـاه P منتقـل می‌کند و بر همین قیاس. سرانجام این سیستم به یک یا چند گروه بی معنـی یـا کـدون پایـان (UAG, UGA یا UAA) می رسد و در نتیجه، زنجیـره‌ی پلـی پپتیـدی از آخـرین آر. ان. ای. انتقال دهنده، آخرین آر. ان. ای. انتقال دهنده از ریبوزوم، و ریبـوزوم از آر. ان. ای. پیـام رسـان جدا می‌شوند.



[1] - Ttheronin
[2] - Phenylalanin

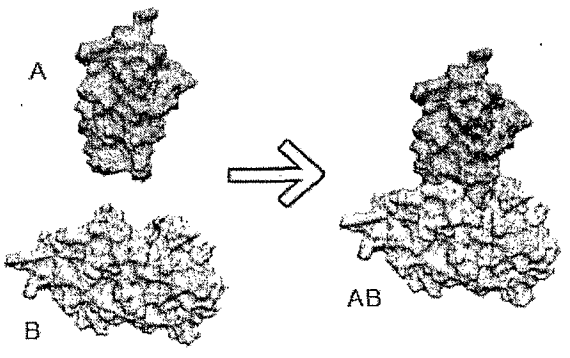
شکل ۲ – ساختار کلی سنتز پروتئین در هسته داران

بدین ترتیب، انتهای ′۵ مولکول آر. ان. ای. پیام رسان با پایانه‌ی آمینو اسیدیِ زنجیره‌ی پلی پپتید، و انتهای ′۳ این مولکول با پایانه‌ی کربوکسیلِ این زنجیره، مطابقت دارند (صبور، ۱۳۸۱).

شرح جریان سنتز پروتئین که اینک از نظر گذشت، چیزی جز بیان کلیات این فرایند نیست. بعضی از جنبه های مهم این فرایند، به نحو متفاوتی در باکتری ها و در سلول های یوکاریوتی انجام می گیرند، که جزئیات آن ها احتیاج به مطالعه تخصصی در این موجودات دارد.

۱-۴- بر هم کنش های پروتئین- پروتئین

اصولاً پروتئین ها برای ایفای نقش های خود در داخل سلول، به صورت مجموعه های پیچیده‌ای در می‌آیند و می‌توانند با ترکیب شدن با سایر پروتئین ها، بر سهولت انجام واکنش های سلولی تاثیر به سزایی داشته باشند (شکل ۳). در بعضی موارد، پروتئین های مشابه با هم ترکیب شده و نقش جدیدی را به عهده می گیرند و در بعضی حالات هم ترکیب چند پروتئین غیر مشابه، منجر به ایجاد نقش جدید و مهمی توسط آن مجموعه خواهد شد. ارتباطات بین پروتئین ها می تواند از بر هم کنش دامنه - دامنه[1] برقرار شود و در بعضی موارد هم تقابل از طریق بر هم کنش دامنه – پپتید[2] انجام می‌گیرد. در هر صورت آنچه مهم است، ایجاد ارتباط بین مجموعه‌ای از پروتئین ها، جهت انجام بهینه نقش های سلولی است.



شکل ۳ – نمای کلی بر هم کنش دو پروتئین

[1] - Domain-Domain Interaction
[2] - Domain-Peptid Interaction

۱-۵- دلایل اهمیت و روش های مطالعه بر هم کنش پروتئین-پروتئین

عمل بیولوژیکی بسیاری از پلیمرهای زیستی، به پیوند شدنِ ملکول‌های کوچک و یا یون ها مربوط می‌شود. برای انجام واکنش آنزیمی، باید سوبسترا به آنزیم متصل شود. در واقع اغلب اعمال زیستی ناشی از برهم کنش مولکول‌های کوچکی همانند متابولیت‌ها و تنظیم کننده‌ها با ناحیه خاصی از ماکرومولکول می باشند. گروهی از پروتئین ها که در تنظیم سلولی و انتقالات غشاء نقش دارند، عمل خود را در صورت تشکیل کمپلکس، انجام می دهند. بنابراین بسیاری از عملکردها در سلول بوسیله پروتئین ها انجام می شود و اختلال در ایجاد کمپلکس و بر هم کنش بین پروتئین ها، بیماری‌های مختلفی را بوجود می‌آورد. بنابراین داشتن اطلاعاتی در مورد نحوه بر هم کنشِ پروتئین ها با پروتئین ها و مولکول‌های کوچک، گام موثری در جهت طراحی داروهای جدید می باشد. از طرفی کافی نبودن ابزار تجربی جهت مطالعه مولکولی بر هم کنش پروتئین ها، نقش بیوانفورماتیک را در این بخش نمایان می سازد.

از کاربردهای مهم بر هم کنش پروتئین ها، می‌توان به نقش‌های مهم در ساختار اندامک های داخل سلولی، سیستم انتقال مواد درون سلولی، بسته بندی کروماتین، انتقال پیام، تنظیم بیان ژن و بسیاری دیگر از فرآیندهای بیولوژیکی مهمِ دیگر اشاره کرد. همچنین، مطالعات بر هم کنشی پروتئینی نقش مهمی را در مطالعه بیماری های فراوانی چون آلزایمر و کروتسفیلد-ژاکوب، که ناشی از اختلال در بر هم کنش بین پروتئین هاست، را دارا می‌باشد. به دلایل ذکر شده و اهمیت موضوع بر هم کنش پروتئین ها در نمو موجودات و بسیاری از بیماری ها، سال های زیادی است که تحقیقات گسترده‌ای در این مورد صورت پذیرفته است.

روش های متعددی برای مطالعه این موضوع ابداع شده است، از جمله این روش ها می توان به سیستم مخمر دو هیبریدی[1] ، روش آنالیزی اولتراسانتریفیوژ[2] ، روش افتراق نوری[3] و غیره اشاره کرد. با توجه با اینکه داده های مورد استفاده قرار گرفته شده در این تحقیق، عمدتا با روش مخمر دو هیبریدی استخراج شده‌اند، به طور خلاصه به توضیح این روش می پردازیم (Fu, 2004 & Jones, 1996).

---

[1] - Yeast two hybrids system
[2] - Analytical ultracentrifugation method
[3] - Light scattering method