

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

دانشکده‌ی ریاضی، آمار و علوم کامپیوتر

دانشگاه تهران

تهران

پیشگویی ساختار سوم پروتئین‌ها در مدل آب‌گریز-آب‌دوست با استفاده از الگوریتم‌های ژنتیک

نگارش

زهرا زواره

استاد راهنما

دکتر هایداه اهرابیان

استاد مشاور

دکتر عباس نوذری دالینی

پایان‌نامه جهت دریافت درجه‌ی کارشناسی ارشد

کارشناسی ارشد علوم کامپیوتر

بهمن ۱۳۸۷

University of Tehran
College of Science
School of Mathematics, Statistics and Computer Science

Prediction of tertiary structure of proteins in Hydrophobic-Hydrophilic model using Genetic Algorithm

by

Zahra Zavareh

Under supervision of

Dr. Hayedeh Ahrabian

Co-supervisor

Dr. Abbas Nowzari Dalini

A thesis submitted to the Graduate Studies Office
In partial fulfillment of the requirements for
The degree of Master of Science
Master of Computer Science

February 2009

تقدیر و تشکر

با تشکر از اساتید محترم، خانم دکتر هایده اهرابیان و آقای دکتر نوذری دالینی به خاطر راهنمایی‌هایشان در این پایان‌نامه که مطمئناً بدون هدایت ایشان این کار به انجام نمی‌رسید.

هم‌چنین از همکاری و کمک‌های دوستان عزیزم خانم‌ها سحر اسدی، شهره تکیار و صغری میکائیل‌نژاد در گردآوری و ویرایش پایان‌نامه صمیمانه تشکر می‌کنم.

تقدیم به

پدر و مادر عزیز و فداکارم

و

برادر همیشه همراهم علیرضا

چکیده

پروتئین زنجیره‌ای متشکل از ۲۰ نوع اسیدآمینه است که با یکدیگر پیوند پپتیدی تشکیل می‌دهند. این زنجیره در فضای سه‌بعدی به‌گونه‌ای تاملی خورد که پایین‌ترین سطح انرژی را دارد. تعیین ساختار دقیق پروتئین با روش‌های آزمایشگاهی به سادگی و در زمانی محدود امکان‌پذیر نیست. در حالی که تا به حال روش‌های محاسباتی زیادی از قبیل شبکه‌های عصبی، روش تکاملی مونت کارلو، بهینه‌سازی دسته‌ی زنبوران و الگوریتم ژنتیک برای حل این مساله در فضاهاى دوبعدی و سه بعدی ارائه شده است. در اینجا مساله‌ی پیشگویی تاخوردگی پروتئین مورد بررسی قرار گرفته و یک الگوریتم ژنتیک در مدلی جدید به نام مدل سطح‌دسترسی ارائه می‌شود. این مدل قابل مقایسه با مدل آب‌گریز-آب‌دوست بوده و از یک تابع برازندگی براساس مقدار سطح‌دسترسی اسیدآمینه‌ها استفاده می‌کند تا نسبت به مدل آب‌گریز-آب‌دوست ساختاری مشابه‌تر به ساختار طبیعی پروتئین تولید کند. از آنجایی که بدست آوردن مقدار دقیق سطح‌دسترسی در بیشتر پروتئین‌ها امکان‌پذیر نمی‌باشد، در اینجا از مقادیری که با روش‌های مختلف تخمین زده می‌شوند استفاده می‌شود.

پیشگفتار

پروتئین‌ها اجزای اساسی همه‌ی سلول‌های زنده می‌باشند. برای ساخت پروتئین‌ها ریبوزوم‌ها اسیدآمینه‌ها را به هم می‌چسبانند و رشته‌های خطی تولید می‌کنند. این رشته‌ها به روش‌های گوناگونی دور یکدیگر می‌پیچند و تا می‌خورند [۳۳]، اما فقط یکی از این روش‌ها به پروتئین اجازه می‌دهد تا به درستی عمل کند. تشخیص نحوه‌ی تاخوردگی یک پروتئین به پیش‌گویی تاخوردگی پروتئین یا پیش‌گویی ساختار پروتئین معروف است. سال‌های زیادی می‌باشد که به اهمیت تاخوردگی پروتئین‌ها پی برده شده است. بررسی بیماری‌هایی مانند آلزایمر، جنون گاوی، امفیوزم ارثی و حتی خیلی از سرطان‌ها نشان می‌دهد که همه‌ی این بیماری‌های ظاهراً نامرتبط با هم نتیجه‌ی اشتباه تاخوردن پروتئین‌ها می‌باشند. همچنین بسیاری از مشکلات غیر منتظره‌ای که شرکت‌های بیوتکنولوژی هنگام تولید پروتئین‌های انسانی در باکتری با آنها مواجه می‌گردند از نادرست تاخوردن پروتئین‌ها ناشی می‌شوند. بعلاوه پروتئینی که صحیح تا نخورد برای سلول‌های مجاور مسمومیت ایجاد می‌کند. اگر یک پروتئین تاخورد را باز کنیم، دوباره به همان حالت اصلی خود تا می‌خورد و این تاخوردگی دارای پایین‌ترین سطح انرژی می‌باشد. در این پایان‌نامه عده‌ای از راه‌حل‌های بهینه برای مساله تاخوردگی پروتئین بررسی و مطالعه شده‌اند [۱۳، ۱۴، ۱۵، ۱۶]. در این روش‌ها به طور کلی یک دنباله از اسیدآمینه‌ها به عنوان ورودی دریافت می‌شود و الگوریتم‌ها سعی می‌کنند بهترین روش تاخوردن آن را که به ساختار اصلی این پروتئین نزدیکتر است به عنوان خروجی تولید می‌کند و معیار انتخاب بهترین تاخوردگی بر پایه انرژی می‌باشد. برای یافتن بهترین تاخوردگی مدل‌های مختلفی وجود دارد که در این پایان‌نامه از مدل آب‌گریز-آب‌دوست پروتئین‌ها استفاده می‌شود [۹، ۱۲] و در آخر روشی نوین با استفاده از سطح دسترسی [۲۴] هر اسیدآمینه در یک رشته‌ی پروتئین پیشنهاد شده است. ابتدا در فصل اول به تعریف مفاهیم اولیه‌ی زیستی و معرفی

الگوریتم‌های ژنتیک پرداخته می‌شود. سپس در فصل دوم الگوریتم‌های ژنتیک موجود برای حل مساله‌ی پیش‌گویی ساختار پروتئین‌ها روی شبکه‌های دوبعدی بررسی می‌شود. فصل سوم شامل الگوریتم‌های ژنتیک برای حل این مساله که روی شبکه‌های سه‌بعدی پیاده‌سازی شده‌اند را توضیح می‌دهد و در فصل آخر یک الگوریتم ژنتیک با معیاری جدید به نام سطح دسترسی ارائه خواهد شد.

فهرست مطالب

۱	تعاریف اولیه	۱
۱	مفاهیم مقدماتی مربوط به پروتئین‌ها	۱-۱
۲	اسیدآمینها	۱-۱-۱
۴	تشکیل یک پروتئین	۲-۱-۱
۵	ساختارهای پروتئین	۳-۱-۱
۱۱	تا کردن پروتئین	۴-۱-۱
۱۱	عملکرد و شکل سطح پروتئین	۵-۱-۱
۱۱	عملکردهای پروتئین	۶-۱-۱
۱۳	معرفی مدل آب‌گریز-آبدوست در پروتئین‌ها	۷-۱-۱
۱۳	سطح دسترسی در اسیدآمینها	۸-۱-۱
۱۴	الگوریتم‌های ژنتیک	۲-۱
۱۵	اجزای تشکیل دهنده‌ی الگوریتم ژنتیک	۱-۲-۱
۲۰	بررسی الگوریتم‌های موجود در فضای دوبعدی	۲
	الگوریتم ژنتیک پیشگویی ساختار پروتئین‌ها در مدل آب‌گریز-آبدوست روی شبکه‌ی	۱-۲
۲۰	مربعی دوبعدی	
۲۲	الگوریتم ژنتیک تاخوردگی پروتئین (PFGA)	۱-۱-۲
۳۱	الگوریتم ژنتیک ساختار دوم (SSGA)	۲-۱-۲

۳۳	نتایج	۳-۱-۲
	الگوریتم ژنتیک پیوندی HGA برای مدل آب‌گریز-آبدوست روی شبکه‌ای FCC در	۲-۲
۳۵	پیش‌گویی تاخوردگی پروتئین‌ها در فضای دوبعدی	
۳۶	شکل بهینه‌ی هسته‌ی H در مدل دوبعدی FCC	۱-۲-۲
۳۸	زیرساختارهای معمول برای لایه‌ی مرکب HP	۲-۲-۲
۳۹	برازندگی مقید احتمالی	۳-۲-۲
۴۰	مفاهیم پیاده‌سازی زیرساختارها	۴-۲-۲
۴۰	پیاده‌سازی	۵-۲-۲
۴۲	نتایج	۶-۲-۲
	۳ بررسی الگوریتم‌های موجود در فضای سه بعدی	۴۴
	بررسی مدل آب‌گریز-آبدوست روی شبکه‌ی مربعی سه‌بعدی در پیش‌گویی تاخوردگی	۱-۳
۴۴	پروتئین‌ها با استفاده از الگوریتم‌های ژنتیک	
۴۴	روش‌ها	۱-۱-۳
۴۶	نتایج و بحث	۲-۱-۳
	الگوریتم ژنتیک پیش‌گویی تاخوردگی پروتئین در مدل آب‌گریز-آبدوست روی شبکه‌ی	۲-۳
۴۹	سه‌بعدی FCC	
۵۱	تعاریف	۱-۲-۳
۵۳	هسته‌ی H سه‌بعدی	۲-۲-۳
۵۵	زیرساختارهایی برای ساختن لایه‌ی مرکب HP	۳-۲-۳
۵۷	حذف دوقلوها در الگوریتم ژنتیک	۴-۲-۳
۶۰	مقایسه‌ی سطح کارایی الگوریتم‌های مختلف	۵-۲-۳
	۴ الگوریتم جدید با استفاده از معیار سطح دسترسی	۶۱

۶۲	شبه‌سازی گرم کردن فلزات (SA)	۱-۴
۶۴	الگوریتم ژنتیک ارائه‌شده روی شبکه‌های مربعی و FCC	۲-۴
۶۴	تولید جمعیت اولیه	۱-۲-۴
۶۵	تابع برازندگی	۲-۲-۴
۶۶	انتخاب، تقاطع ژنی، جهش ژنتیکی و جایگزینی	۳-۲-۴
۶۷	پیاده‌سازی	۳-۴
۶۷	بررسی و تحلیل نتایج	۴-۴
۶۸	پیشنهادات	۵-۴
۷۰	مراجع	
۷۴	واژه‌نامه‌ی فارسی	
۷۸	واژه‌نامه‌ی انگلیسی	
۸۲	ضمیمه	

فصل ۱

تعاریف اولیه

در این فصل تعاریف اولیه مرتبط با این پایان نامه مطرح خواهند شد. این مفاهیم اولیه به منظور رسیدن به یک زبان مشترک اولیه مطرح شده‌اند. مفاهیم اولیه مذکور در دو بخش مفاهیم مقدماتی مربوط به پروتئین‌ها [۸, ۳۳] و الگوریتم‌های ژنتیک [۱۰, ۱۲, ۲۲, ۲۳] ارائه شده‌اند.

۱-۱ مفاهیم مقدماتی مربوط به پروتئین‌ها

مجموعه فرآیندهایی که توسط سلول‌ها جهت زنده نگه‌داشتن ارگانیسم‌ها انجام می‌شود، طبق یک برنامه‌ریزی از پیش تعیین شده است که از ابتدا در سلول قرار دارد. به این برنامه‌ریزی خصوصیات ژنتیکی سلول نیز گفته می‌شود و در هسته‌ی سلول در رشته‌ای از مولکول‌های زیستی به نام اسیدهای داکسی‌ریبونوکلئیک یا به اختصار *DNA* ذخیره شده است. از روی رشته‌ی *DNA* در فرآیندی به نام رونویسی، مولکول‌های دیگری به نام اسیدهای ریبونوکلئیک یا به اختصار *RNA* بوجود می‌آید. سپس رشته *RNA* به پروتئین ترجمه می‌شود.

مولکول‌های پروتئین عامل اصلی انجام فعل و انفعالات شیمیایی درون بدن می‌باشند. پروتئین‌ها زنجیره‌هایی متشکل از ۲۰ نوع اسیدآمینوی مختلف هستند که در شکل ۱-۱، نام اختصاری سه حرفی و نام اختصاری تک حرفی هر یک به همراه شکل سه بعدی ساختاری آن اسیدآمینو در شرایط محیط خنثی نمایش داده شده است. پروتئین‌ها از نظر تعداد و نوع اسیدآمینوهای موجود در ساختارشان از هم متمایز می‌شوند. این در حالی است که هر پروتئین می‌تواند شامل ۵۰، تا بیش از ۴۰۰۰ اسیدآمینو می‌باشد. یک سلول ساده ممکن است شامل بیش از ۱۰۰۰۰ نوع پروتئین گوناگون باشد. این گوناگونی

پروتئین‌ها برای عملکرد هر سلول در یک موجود زنده ضروری است. صرف‌نظر از آب، حدود نیمی از ترکیبات هر سلول را پروتئین تشکیل می‌دهد.

ساختار پروتئین در عملکرد آن دارای نقش مهمی می‌باشد. هر پروتئین یک ساختار منحصر به فرد دارد. پلی‌پپتیدها زنجیره‌های طولانی هستند که از به هم پیوستن اسیدآمینه تشکیل می‌شوند. زنجیره‌های پلی‌پپتیدی در طبیعت به گونه‌ای ترمی‌خورند که دارای پایین‌ترین سطح انرژی هستند و ساختار طبیعی و منحصر به فرد پروتئین‌ها را تشکیل می‌دهند. برای بحث درباره‌ی تشکیل یک پروتئین ابتدا به معرفی ساختار اسیدآمینه‌ها نیاز داریم.

۱-۱-۱ اسیدآمینه‌ها

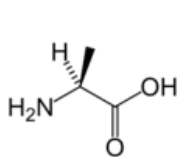
اسیدآمینه‌ها شامل کربن، هیدروژن، اکسیژن، نیتروژن و گاهی سولفور بوده و همواره دارای دو گروه اصلی آمینو NH_2 و کربوکسیل ($COOH$) هستند. هر دو گروه اصلی به کربن آلفا که یک کربن نامتقارن (کربنی که چهار گروه مختلف به آن متصل می‌باشد) است، متصلند. سومین مکان پیوندی کربن آلفا، هیدروژن است. کربن آلفا در چهارمین مکان پیوندی خود یک زنجیره‌ی جانبی یا گروه R دارد که موجب ساختار و ویژگی‌های یکتایی برای اسیدآمینه می‌شود. ۲۰ اسیدآمینه مختلف در پروتئین‌ها وجود دارد که تنها در نوع گروه R به کار رفته در ساختارشان از هم متفاوت هستند (شکل ۱-۲).

ویژگی‌های گروه R اسیدآمینه

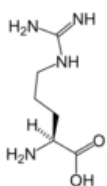
ویژگی‌های یکتای گروه‌های R اسیدآمینه‌ها، در ساختار پروتئین تشکیل شده نقش مهمی دارد، چنان‌که تعداد، نوع و دنباله‌ی پیوندی اسیدآمینه‌های به کار رفته در یک پروتئین در چگونگی ساختار آن تاثیرگذار است.

- پنج نوع از اسیدآمینه‌ها گروه R قطبی دارند و با مولکول‌های قطبی دیگر یا آب پیوندهای هیدروژنی ضعیف برقرار می‌کنند. این اسیدآمینه‌های قطبی، آب‌دوست می‌باشند.
- پنج نوع از اسیدآمینه‌ها گروه R باردار دارند که آنها هم آب‌دوست هستند. اسیدآمینه‌های باردار عموماً اسیدی هستند (گروه کربوکسیل یک H^+ از دست می‌دهد و گروه آمینو یک H^+ بدست می‌آورد).

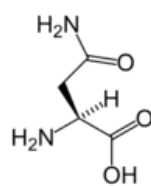
فصل ۱. تعاریف اولیه



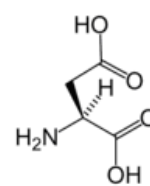
L-Alanine
(Ala / A)



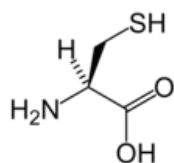
L-Arginine
(Arg / R)



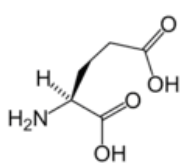
L-Asparagine
(Asn / N)



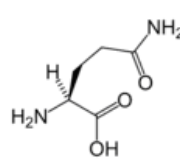
L-Aspartic acid
(Asp / D)



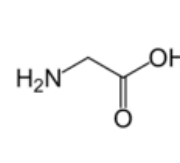
L-Cysteine
(Cys / C)



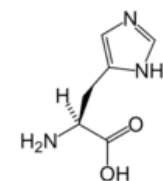
L-Glutamic acid
(Glu / E)



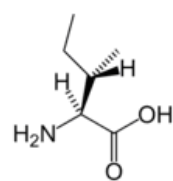
L-Glutamine
(Gln / Q)



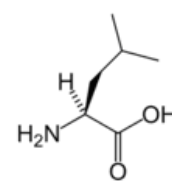
Glycine
(Gly / G)



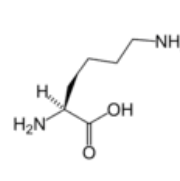
L-Histidine
(His / H)



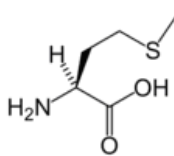
L-Isoleucine
(Ile / I)



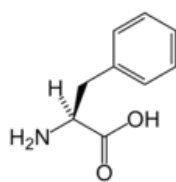
L-Leucine
(Leu / L)



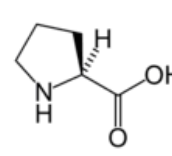
L-Lysine
(Lys / K)



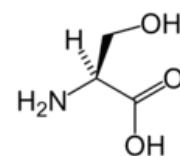
L-Methionine
(Met / M)



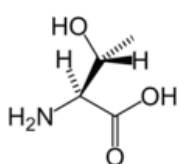
L-Phenylalanine
(Phe / F)



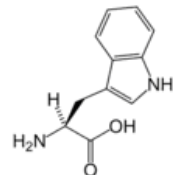
L-Proline
(Pro / P)



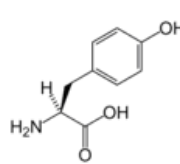
L-Serine
(Ser / S)



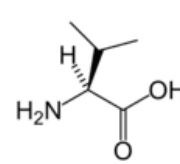
L-Threonine
(Thr / T)



L-Tryptophan
(Trp / W)

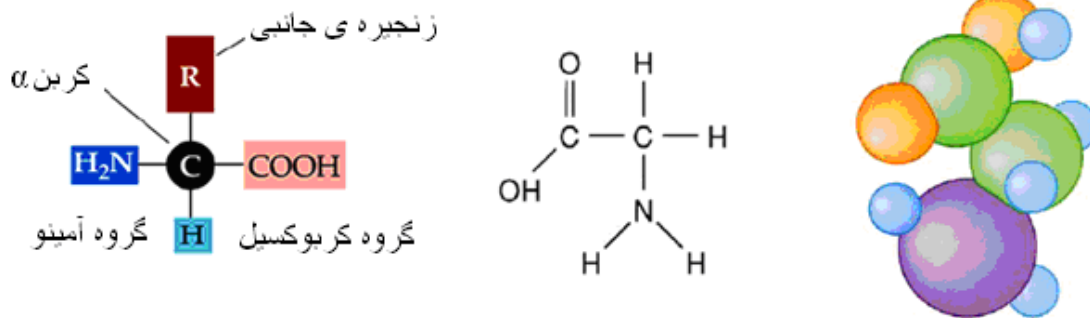


L-Tyrosine
(Tyr / Y)



L-Valine
(Val / V)

شکل ۱-۱ : فهرست و ساختار اسید آمینه‌ها.



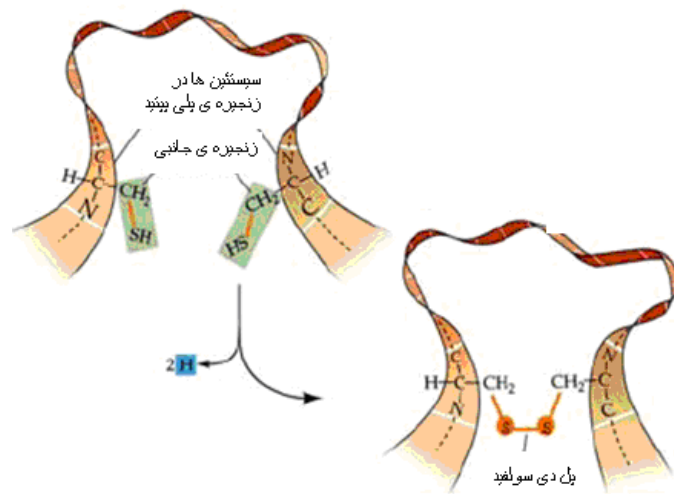
شکل ۱-۲: ساختار یک اسید آمینه.

- هفت نوع از اسید آمینه‌ها گروه R غیر قطبی دارند؛ مانند CH_2 و CH_3 . اسید آمینه‌های غیر قطبی آب‌گریز هستند و در محیط‌های آبی تمایل به جمع شدن در بخش داخلی یک پروتئین دارند.
- سه نوع اسید آمینه‌ی دیگر که هر سه آب‌گریز هستند، دارای ویژگی‌های زیر می‌باشند:
 - اسید آمینه‌ی سیستئین یک گروه سولفوهیدریل انتهایی دارد که با گروه‌های جانبی سیستئین‌های دیگر در پروتئین‌ها تشکیل پل‌های دی‌سولفید کووالانسی می‌دهد. پل‌های دی‌سولفید اهمیت زیادی در نحوه‌ی تاخوردن پروتئین‌ها دارند. سیستئین‌هایی که تشکیل پل نداده‌اند آب‌گریز هستند. (شکل ۱-۳)
 - گروه اصلی آمینو در پرولین به گونه‌ای تغییر کرده است که توانایی آن برای تشکیل پیوند هیدروژنی کاهش یافته است. همچنین گروه R حلقوی شکل، میزان چرخش کربن آلفا را محدود می‌کند، بنابراین پرولین بیشتر اوقات باعث خم شدن یا پیچ خوردن پروتئین می‌شود.
 - به دلیل کوچک بودن گروه R ($-H$) به کار رفته در ساختار گلیسین، این اسید آمینه کوچکترین اسید آمینه موجود در طبیعت می‌باشد.

- در بیشتر موارد متیونین اولین اسید آمینه در پلی‌پپتید است.
- اسید آمینه‌هایی که گروه‌های R حلقوی هیدروکربنی دارند اغلب معطر هستند.

۲-۱-۱ تشکیل یک پروتئین

۱. پروتئین در ابتدا زنجیره‌ای متشکل از اسید آمینه‌هاست که پلی‌پپتید نامیده می‌شوند.
۲. اسید آمینه‌ها با از دست دادن آب و ایجاد پیوند پپتیدی کووالانسی به هم متصل می‌شوند.



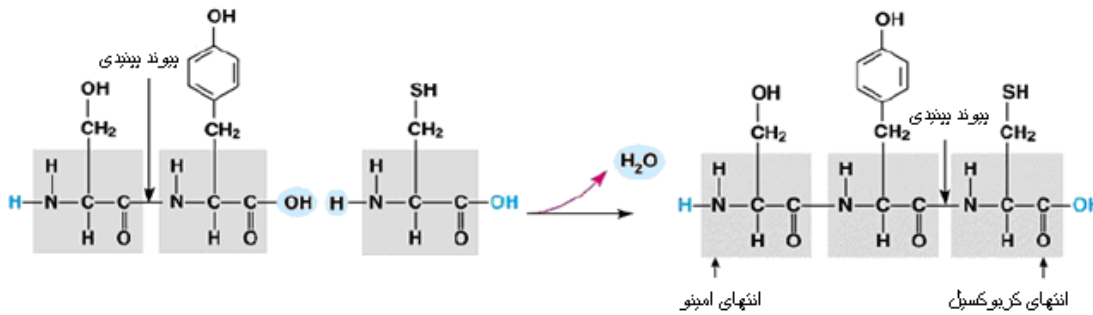
شکل ۱-۳: سیستئین و پل دی سولفید.

(شکل ۱-۴) پیوند پپتیدی بین گروه کربوکسیل از یک اسید آمینه و گروه آمینو از اسید آمینه دیگر تشکیل می شود. ابتدای یک پلی پپتید همیشه گروه آمینوی اولین اسید آمینه در پلی پپتید و انتهای آن گروه کربوکسیل آخرین اسید آمینه در زنجیره است.

- پیوند پپتیدی $C-N-C$ فاقد ویژگی چرخش آزاد در پیوندهای یگانه است و میزان تا خوردن پلی پپتید را محدود می کند.
- گروه کربوکسیل در اثر پیوند پپتیدی، قطبی منفی و گروه آمینو قطبی مثبت می شود که در طی شکل گیری مولکول پروتئین تمایل به تشکیل پیوند هیدروژنی دارند.

۳-۱-۱ ساختارهای پروتئین

اطلاعات ساختاری در یک مولکول زیستی به چهار سطح رده بندی می شود. هر سطح از اطلاعات ساختاری زیربنایی برای سطح ساختاری بعدی و درک مناسب تر از عملکرد یک مولکول زیستی است. یک مولکول زیستی ند پروتئین ساختار بسیار پیچیده ای داشته و تشریح و نگهداری اطلاعات مربوط به این ساختار بسیار دشوار است. به همین دلیل اطلاعات مربوط به ساختار مولکولهای زیستی (مولکول RNA و مولکول پروتئین) را در چند سطح بیان می کنند. رده بندی اطلاعات ساختاری به گونه ای است که ساختار اول کمترین اطلاعات از ساختار کلی مولکول زیستی و ساختار چهارم بیشترین اطلاعات را در این زمینه داراست.



شکل ۱-۴: تشکیل پیوند پپتیدی.

زنجیره‌ی پلی‌پپتید ساختار اولیه‌ی یک پروتئین را تشکیل می‌دهد. پروتئین‌ها دستخوش تحول دیگری برای رسیدن به شکل و ترکیب اصلی خود می‌شوند. پروتئین ساختارهای دوم و سوم و گاهی ساختار چهارم را پیش از رسیدن به شکل اصلی طی می‌کند. بعضی از پروتئین‌ها از چندین پلی‌پپتید ترکیب می‌شوند.

تعداد، نوع، شباهت و دنباله‌ی پیوندی اسیدآمینه‌ها در زنجیره‌ی پلی‌پپتیدی، ترکیب یا شکل نهائی آن و در نتیجه عملکرد آن را تعیین می‌کند. حتی جایگزینی یک اسیدآمینه در دنباله‌ی پیوندی یک پلی‌پپتید به طور چشمگیری شکل نهائی پروتئین و توانایی عملکرد آن را تغییر می‌دهد.

ساختار سطح پروتئین در عملکرد آن نقش بسزایی دارد. مثلاً در هموگلوبین گروه R سطح خارجی باید قطبی باشد تا بتواند گروه هم (شامل آهن) را که با مولکول‌های اکسیژن پیوند می‌دهد نگه دارد. در واقع تمام پروتئین‌ها اسیدآمینه‌های غیرقطبی خود را به درون پروتئین هدایت کرده و اسیدآمینه‌های قطبی و باردار را به سمت محیط آبی سلول می‌فرستند.

ساختار اول

ساختار اول، اجزای تشکیل دهنده رشته مولکول زیستی را مشخص می‌کند که این رشته برای مولکول‌های پروتئین یک توالی از اسیدآمینه‌ها است. برای مثال ساختار اول یک مولکول پروتئین می‌تواند به صورت زیر باشد.

AMRNFDC EIVYPERHAPLSNDCTKRQGRWA

ساختار دوم

اگرچه پیوندهای پتیدی بین اسیدآمینها در ساختار اولیه یک پروتئین تشکیل می‌شوند، پلی‌پتید قبل از رسیدن به شکل اصلی خود دستخوش تغییرات دیگری نیز می‌گردد.

ساختار دوم پروتئین شامل الگوهای تکرارشونده‌ای در طول زنجیره‌ی پلی‌پتیدی می‌باشد. هنگامی که پیوندهای پتیدی ایجاد می‌شوند، اسیدآمینها در یک ردیف قرار می‌گیرند و اتم‌های هیدروژن گروه آمینو جذب اتم اکسیژن دارای پیوند دوگانه در پیوند پتیدی شده و پیوند هیدروژنی برقرار می‌کنند. این پیوند پلی‌پتید را می‌پیچاند تا به یک ساختار دوم پایدار یا مارپیچ آلفا (شکل ۱-۵) برسد. مارپیچ آلفا بعد از هر ۴ اسید آمینه می‌پیچد. اسیدآمینهای بزرگ می‌توانند با جلوگیری از تشکیل پیوند هیدروژنی از پیچ خوردن جلوگیری کنند.

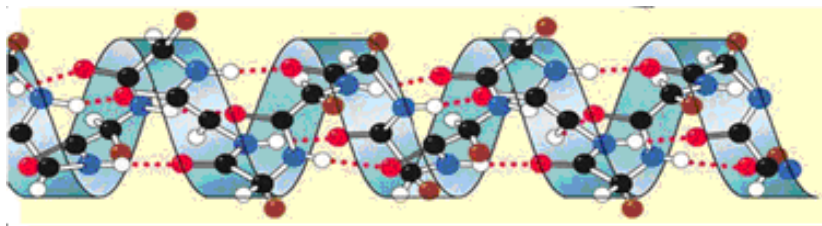
بعضی از نواحی پلی‌پتید به طور گسترده و موازی با هم قرار گرفته‌اند. مجدداً پیوندهای هیدروژنی بین گروه‌های $N-H$ و $C=O$ برقرار می‌شوند و بین زنجیره‌های مجاور صفحه‌ی چین‌داری به نام صفحه‌ی بتا (شکل ۱-۶) تشکیل می‌دهند. صفحات بتا می‌توانند بین نواحی یک پلی‌پتید یا بین پلی‌پتیدهای مختلف ایجاد شوند.

تار عنکبوت مثال خوبی برای پروتئین با صفحه‌ی بتا می‌باشد. فیبرهای آن انعطاف‌پذیر و پیچیده شده است و مایعی که آن را پوشانده باعث استحکام بیشتر آن می‌شود. فشارمایع سطح باعث می‌شود تار موج داشته و فیبرهای آن پیچ خورده بمانند (شکل ۱-۷).

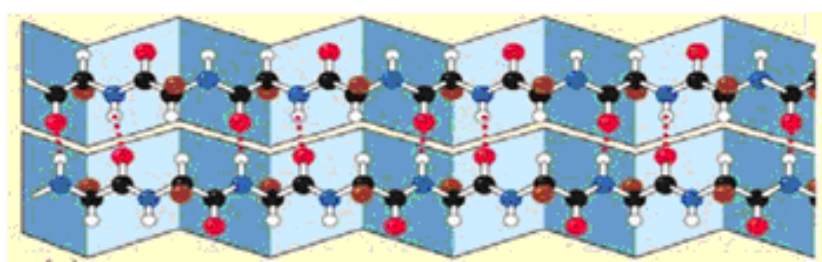
الگوهای رایجی وجود دارند که وابسته به ساختار دوم پروتئین می‌باشند. این الگوها موتیف نامیده می‌شوند. به عنوان مثال موتیف بتا آلفا بتا موجب تاخوردگی در یک پروتئین می‌شود. موتیف‌ها در تنظیم ژن‌ها مورد بحث قرار می‌گیرد. یک موتیف معمول "مارپیچ چرخش مارپیچ" یا "آلفا چرخش آلفا" می‌باشد (شکل ۱-۸).

ساختار سوم

در نتیجه‌ی پیچ و چین‌های ساختار دوم در نواحی یک پلی‌پتید، اثر متقابل بین زنجیره‌های جانبی اسیدآمینها (گروه‌های R)، موجب تاخوردگی و پیچ‌خوردگی بیشتری شده و ساختار سوم پروتئین را



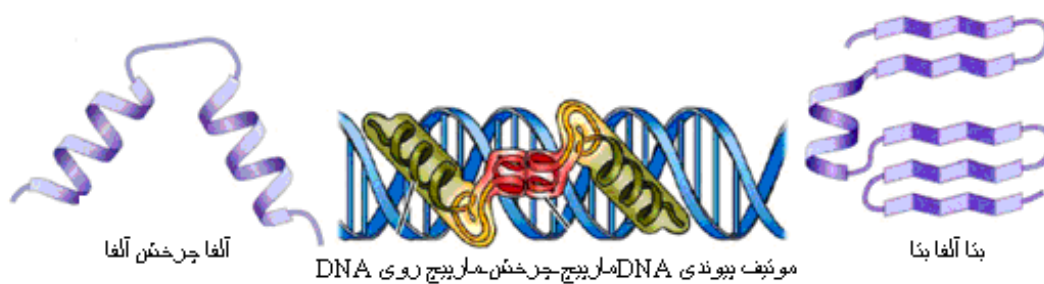
شکل ۱-۵: مارپیچ آلفا.



شکل ۱-۶: صفحه ی بتا.



شکل ۱-۷: پروتئین در تار عنکبوت.



شکل ۱-۸: موتیف‌های ساختار دوم.

ایجاد می‌کنند (شکل ۱-۹). ترکیب نهایی بیشتر پروتئین‌ها یک شکل کروی است.

- به عنوان نمونه مناطق آب‌گریز در ساختار سوم، بین گروه‌های اسیدآمینه‌هایی ایجاد گردیده که توسط زنجیره‌های جانبی غیرقطبی به سمت داخل پروتئین هدایت می‌شوند، ایجاد می‌شود. بر هم‌کنش‌های واندروالسی این اسیدآمینه‌ها را پایدار نگاه داشته و هر تغییری در یک اسیدآمینه‌ی غیرقطبی شکل نهایی پروتئین را تغییر می‌دهد.

- پیوندهای دی‌سولفید (که پیوندهای کووالانسی قوی هستند) بین مولکول‌های سیستئین نزدیک به هم، در ساختار سوم بسیار موثر است.

- پیوندهای یونی می‌توانند بین گروه‌های R باردار برقرار شده و پل‌های نمکی می‌سازند.

حوزه‌های پروتئین

دستورالعمل‌های ساختار پروتئین در مولکول‌های DNA رمز شده‌اند. در ژن (یک منطقه از DNA که دستورالعمل‌های پلی‌پپتید را رمز می‌کند) ناحیه‌هایی به نام اگزون وجود دارند. هر اگزون رمز شده در یک پروتئین به طور مستقل در یک واحد به نام حوزه (شکل ۱-۱۰) تا می‌خورد. حوزه‌های مختلف به بقیه‌ی پلی‌پپتید متصل می‌شوند. حوزه‌ها ممکن است وظایف مختلفی برای یک پروتئین معین انجام دهند. برای مثال یک حوزه از یک آنزیم ممکن است به عنوان یک عامل کمکی و حوزه‌ی دیگر به عنوان یک جایگاه فعال برای آنزیم محسوب شود.

ساختار چهارم

هرگاه دو یا چند زنجیره‌ی پلی‌پپتیدی به هم ملحق شوند، یک ساختار چهارم مانند هموگلوبین (شکل ۱-۱۱) در مولکول پروتئین می‌سازند. پروتئین‌های ساختار چهارم با یک مولکول دیگر که بیشتر اوقات یک ماده معدنی است، ترکیب می‌شوند. به عنوان مثال هموگلوبین شامل آهن است.

کلاژن یک پروتئین است که معمولاً در بافتهای پیوندی یافت می‌شود و یک مارپیچ کلاژن داشته و از پیچش زنجیره‌های پلی‌پپتیدی به دور همدیگر تشکیل می‌شود. در نتیجه فیبرهای کلاژن بسیار قوی هستند (شکل ۱-۱۱).