

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

٩٥٧٥٥

دانشگاه تهران
پردیس علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

پیش بینی ساختار دوم پروتئین به کمک شبکه های عصبی

نگارش:

امیر لکی زاده

استاد راهنما:

دکتر هایده اهرابیان

استاد مشاور:

دکتر عباس نوذری دالینی

پایان نامه برای دریافت درجه کارشناسی ارشد

در رشته علوم کامپیوتر

۱۳۸۷ / ۲ / ۲۸

بهمن ۱۳۸۶

ب

۹۵۷۵۵



بنام خدا
دانشگاه تهران

پرديس علوم
دانشکده ریاضی، آمار و علوم کامپیوتر

گواهی دفاع از پایان نامه کارشناسی ارشد

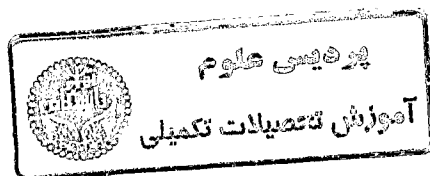
هیات داوران پایان نامه کارشناسی ارشد آقای امیر لکی زاده
در رشته علوم کامپیوتر
با عنوان: پیش بینی ساختار دوم پروتئین با کمک شبکه عصبی
را در تاریخ ۸۶/۱۱/۲۷

به عدد به حروف
با نمره نهایی: ۱۹/۲۵ نوزده و بیست و پنج صدم

و درجه: عالی ارزیابی نمود.

ردیف	مشخصات هیات داوران	نام و نام خانوادگی	مرتبۀ دانشگاهی	دانشگاه یا موسسه	امضاء
۱	استاد راجعاً:	دکتر هایده اهرابیان	دانشیار	دانشگاه تهران	
۲	استاد مشاور داخلی:	دکتر عباس نوذری دالینی	استادیار	دانشگاه تهران	
۳	استاد داور خارجی:	دکتر مهدی صادقی	استادیار	دانشگاه تهران IBB	
۴	نماینده کمیته تصمیمات تکمیلی دانشکده ریاضی، آمار و علوم کامپیوتر:	دکتر هایده اهرابیان	دانشیار	دانشگاه تهران	

تذکره: این برگه پس از تکمیل توسط هیات داوران در نخستین صفحه پایان نامه درج می گردد.



بسمه تعالی



اداره کل تحصیلات تکمیلی

جمهوری اسلامی ایران

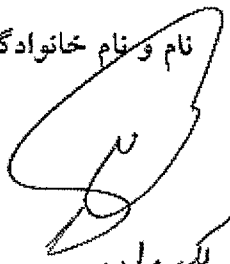
دانشگاه تهران

تعهد نامه اصالت اثر

اینجانب امیر لکی زاده متعهد می شوم که مطالب مندرج در این پایان نامه حاصل کار پژوهشی اینجانب بوده است و به دستاوردهای پژوهشی دیگران که در این پژوهش از آن ها استفاده شده است، مطابق مقررات، ارجاع و در فهرست منابع و مأخذ ذکر گردیده است. این پایان نامه قبلاً برای احراز هیچ مدرک هم سطح یا بالاتر ارائه نشده است و در صورت اثبات تخلف در هر زمان، مدرک تحصیلی صادر شده توسط دانشگاه از اعتبار ساقط خواهد شد. کلیه حقوق مادی و معنوی این اثر متعلق به پردیس علوم دانشگاه تهران می باشد.

نام و نام خانوادگی دانشجو

امضاء


امیر لکی زاده

تقدیم به

پدر و مادر عزیز و مهربانم

و

همسرم فاطمه

پیشگفتار

مواد سازنده بدن موجودات زنده به دو گروه مواد معدنی و مواد آلی تقسیم می شوند. مواد معدنی از قبیل آب، دی اکسید کربن و اکسیژن در طبیعت یافت می شوند. بدن موجودات زنده مواد معدنی را به مواد معدنی تبدیل می کند. چهار گروه عمده مواد آلی بدن موجودات زنده عبارتند از هیدرات های کربن، لیپیدها، اسیدهای نوکلئیک و پروتئین ها. دو گروه عمده اسیدهای نوکلئیک عبارتند از DNA و RNA هر دو از اجتماع نوکلئوتیدها تشکیل می شوند. اندوختن و انتقال اطلاعات ژنتیکی بوسیله DNA انجام می گیرد. از روی DNA و به کمک RNA، پروتئین ساخته می شود. پروتئین ها خود از اسیدهای آمینه تشکیل می شوند و هر سه نوکلئوتید در کنار هم یک اسیدآمینه را کد می کند [۳۲]. پروتئین ها نقش کلیدی در بسیاری از فعالیت های حیاتی موجودات زنده بر عهده دارند. آنزیم ها که نقش راه انداز و تسریع کننده واکنش های درون سلول را بر عهده دارند، خود نوعی پروتئین به شمار می روند. با توجه به این اهمیت، تولید پروتئین و تغییر ساختار آن برای نیل به اهدافی از قبیل طراحی داروها، ارسال دارو به بافت هدف و تولید جایگزین های زیستی برای از بین بردن ترکیبات شیمیایی خطرناک محیط و غیره، از اهداف مطالعه پروتئین ها در علوم بیوانفورماتیک و مهندسی پروتئین می باشد [۱۷، ۱۸، ۳۰، ۴۳].

در این راه رسیدن به این هدف آگاهی از عملکرد و طرز کار یک پروتئین با توالی آمینو اسیدی کشف شده، اهمیت فوق العاده ای می یابد. اما عملکرد یک پروتئین به ساختار فضایی یا همان ساختار سه بعدی آن بستگی دارد. برای یک پروتئین به طور کلی سه نوع ساختار تعریف می شود. ساختار نوع اول پروتئین همان توالی آمینو اسیدی آن می باشد. ساختار دوم پروتئین ساختاری دو بعدی می باشد و به طور عمده از قطعات مارپیچی به نام مارپیچ آلفا^۱ و صفحات تاخورده بتا^۲ و پیچ^۳ تشکیل می شود و منظور از ساختار سوم پروتئین، ساختار سه بعدی و فضایی آن می باشد. به علت اینکه ساختار فضایی یک پروتئین، پایه اصلی در راه کشف عملکرد آن می باشد و از طرفی بدست آوردن آرایش فضایی در شرایط آزمایشگاهی و به کمک روش هایی از قبیل کریستالوگرافی با اشعه X، زمان بر و گران می باشد، روش های محاسباتی برای پیش بینی ساختار پروتئین اهمیت فوق العاده ای می یابد. البته با توجه به مشکلات پیش بینی ساختار سوم، در عمل سعی می شود که ساختار دوم پروتئین پیش بینی شود و آن گاه به کمک ساختار دوم، ساختار فضایی پروتئین پیش بینی شود [۱۷، ۱۸، ۳۰، ۴۳].

از دهه ۱۹۷۰ تاکنون روش های پیش بینی ساختار دوم بسیاری ارائه شده است بطوری که درگذشت زمان با استفاده از الگوریتم ها و ایده های جدید، دقت پیش بینی به میزان قابل توجهی افزایش یافته است [۴۰]. به طور کلی روش های پیش بینی ساختار دوم تا به امروز را می توان در قالب سه نسل دسته بندی کرد. روش های نسل اول که در زمره اولین تلاش ها بود، روش هایی آماری بودند و سعی

alpha helix	۱
beta sheet	۲
Coil	۳

می کردند که برای هر دنباله آمینو اسیدی جدید با کمک پارامترهای بدست آمده از تعداد معدودی پروتئین با ساختار شناخته شده تا آن زمان، ساختار دوم آن را صرفاً بر اساس همان پروتئین و مستقل از خانواده پروتئینی آن پیش بینی کنند. این روش ها در عمل توانستند با دقت حدود ۵۰ درصد ساختار دوم را پیش بینی کنند [۹،۲۱]. با افزایش تعداد پروتئین های با ساختار معلوم و واقعی تر شدن پارامترها، روش های نسل دوم بوجود آمدند و توانستند به دقت حدود ۶۰ درصد دست یابند. ولی این روش ها هم چنان از اطلاعات تکاملی پروتئین استفاده نمی کردند [۲۰،۲۲،۲۳]. از اوایل دهه ۹۰ روش های نسل سوم ارائه شد. که می کوشیدند که از اطلاعات تکاملی یک پروتئین استفاده کنند. این اطلاعات به کمک انطباق چندگانه^۴ پروتئین مطلوب با مجموعه پروتئین هایی که شباهت آمینو اسیدی با آن دارند بدست می آید. ایده این کار هم اینست که پروتئین های با شباهت آمینو اسیدی به احتمال فراوان دارای ساختار سه بعدی شبیه به هم نیز می باشند هر چند که عکس این مطلب صحیح نمی باشد و اکثر پروتئین های با ساختار فضایی مشابه دارای تنها ۱۰ درصد شباهت آمینو اسیدی می باشند. این روش ها توانستند به دقت های حدود ۷۰ درصد دست یابند. [۲]، [۵،۲۶،۳۷] و [۱۲].

در سال های اخیر از تکنیک های ماشین یادگیری از قبیل شبکه های عصبی^۵ و ماشین های بردار پشتیبان^۶ به طور گسترده ای در روش های پیش بینی ساختار دوم استفاده می شود و توانسته اند به نتایج قابل قبولی دست یابند به گونه ای که روش های نسل سوم که از شبکه های عصبی به عنوان ابزار محاسباتی استفاده می کنند توانسته اند که ساختار دوم را با دقت تقریبی ۸۰ درصد پیش بینی کنند. علت این موفقیت به ماهیت مسئله پیش بینی، مانند پراکنده بودن و نویزی بودن اطلاعات در دسترس از ساختارهای معلوم و هم چنین ویژگی های این تکنیک ها از قبیل امکان یادگیری و توانایی تعمیم آن ها برمی گردد [۲،۱۲].

در این پایان نامه در ابتدا در فصل اول مروری بر مفاهیم پایه زیست شناسی صورت می گیرد. سپس در فصل دوم شبکه عصبی به عنوان یک مدل محاسباتی معرفی می شود. در این فصل انگیزه ایجاد این مدل محاسباتی و ویژگی های آن و نحوه بکارگیری آن بیان می شود. در فصل سوم، ابتدا مطالبی درباره اهمیت پیش بینی ساختار دوم پروتئین و کاربرد های آن ذکر می شود سپس تاریخچه ای از روش های پیش بینی و سیر تکاملی آن ها بیان می شود و در ادامه دو پارامتر و معیار برای بررسی دقت روش های پیش بینی معرفی می گردد [۴۵] سپس روش PHD که جزو اولین روش های نسل سوم بر پایه شبکه عصبی می باشد، بررسی می شود [۳۷] در ادامه دو مورد از جدیدترین روش های پیش بینی ساختار دوم بررسی می شوند. این روش ها عبارتند از SABLE و SPINE و [۲،۱۲]. هر دو این روش ها از نوع روش های نسل سوم می باشند بدین معنی که از اطلاعات تکاملی موجود در جداول پروفایل استفاده می کنند. جدول پروفایل یک جدول امتیاز دهی بر اساس مکان می باشد و به کمک انطباق چندگانه بدست می آید. هر دو این روش ها از شبکه های عصبی، به عنوان یک مدل محاسباتی با قابلیت یادگیری استفاده می کنند و ساختمان شبکه ها در هر دو مشابه روش PHD می باشد با این تفاوت که روش SABLE در هنگام پیش بینی ساختار دوم یک اسید آمینه از اطلاعات سطح دسترسی

multiple alignment	۴
Neural Networks	۵
Support Vector Machine	۶

آن اسیدآمینة استفاده می کند . منظور از سطح دسترسی یک اسیدآمینة ، درصد نسبت مساحت اشغال شده توسط یک اسیدآمینة به کل مساحت سطح خارجی پروتئین در ساختار فضایی آن می باشد . معمولاً اسیدآمینة ها را بر حسب سطح دسترسی آن ها به سه کلاس کاملاً بیرونی، نسبتاً بیرونی و کاملاً بیرونی و یا دو کلاس درونی و بیرونی دسته بندی می کنند . روش SABLE نشان می دهد که دخالت دادن سطح دسترسی می تواند به میزان قابل قبولی دقت پیش بینی را افزایش دهد .

روش SPINE با اینکه از روش SABLE جدیدتر می باشد ولی توانسته است بدون دخالت دادن اطلاعات سطح دسترسی به موفقیت بسیار خوبی یعنی دقت بالای ۸۰ درصد در پیش بینی ساختارهای دوم دست یابد . البته SPINE علاوه بر پیش بینی ساختار دوم ، پیش بینی سطح دسترسی را هم به طور مستقل انجام می دهد . در انتهای فصل سوم ، نحوه ساخت جدول پروفایل به روش PSI-BLAST و سپس برنامه DSSP معرفی می گردد . DSSP برنامه ای است که به ازای هر دنباله آمینو اسیدی از یک پروتئین، یک فایل حاوی مشخصات مختلف اسیدآمینة های پروتئین از قبیل ساختار دوم هر اسیدآمینة و سطح دسترسی آن ها و زنجیره ای که این اسیدآمینة به آن تعلق دارد و پاره ای مشخصات دیگر تولید می کند .

در فصل چهارم این پایان نامه ، یک روش جدید از نوع روش های نسل سوم و بر پایه شبکه های عصبی ارائه شده است . این روش توانسته است با دخالت دادن عدد تماس^۷ هر اسید آمینة ، دقت پیش بینی ساختار دوم پروتئین را افزایش دهد . عدد تماس برای هر اسید آمینة در یک پروتئین تعریف می شود و بیانگر تعداد پیوندهای هر اسید آمینة با اسید آمینة های دیگر در دنباله آمینو اسیدی پروتئین می باشد . مطالعات آزمایشگاهی نشان می دهد که رابطه معنی داری میان ساختار دومی که اسید آمینة به آن تعلق دارد با عدد تماس اسید آمینة وجود دارد ، به طوریکه با داشتن عدد تماس اسید آمینة ، به طور دقیق تری می توان ساختار دوم اسید آمینة را پیش بینی کرد و نتایج عملی حاصل از این روش ، صحت این موضوع را نشان می دهد . از نقاط قوت این روش صرف نظر از استفاده از مقادیر عدد تماس، آن هم برای اولین بار، استفاده از ساختارهای متفاوت از شبکه های عصبی می باشد.

قدردانی

با حمد و ثنای پروردگار و با درود بر پیامبر اعظم (ص) و با آرزوی تعجیل در فرج امام عصر (عج) ،

از استادان گرانقدرم سرکار خانم دکتر هایده اهرابیان و آقای دکتر عباس نوذری که در تمام دوران تحصیل مشوق من بوده‌اند و هم چنین

از آقای دکتر مهدی صادقی، نهایت سپاس و قدردانی را دارم و از دوستان عزیزم سید امیر مرعشی و سید حسین علوی سلطانی نیز

کمال تشکر و قدردانی را دارم .

فهرست مطالب

ث	پیشگفتار
ج	قدردانی
ح	فهرست مطالب
د	فهرست جداول
ذ	فهرست اشکال
۱	۱ آشنایی با مقدمات زیست شناسی مولکولی
۱.....	۱-۱ سلول
۲.....	۱-۱-۱ ساختار سلول

۴	تقسیم سلول	۲-۱-۱	
۴	چرخه سلولی	۳-۱-۱	
۵	ساختار شیمیایی سلول	۲-۱	
۶	هیدرات های کربن	۱-۲-۱	
۶	لیپیدها	۲-۲-۱	
۶	پروتئین ها	۳-۲-۱	
۷	اسیدهای نوکلئیک	۴-۲-۱	
۹	ساختمان رشته ای اسیدهای نوکلئیک	۵-۲-۱	
۱۲	کنش ماکرومولکول ها	۳-۱	
۱۲	همانندسازی DNA	۱-۳-۱	
۱۳	نسخه برداری	۲-۳-۱	
۱۳	ترجمه	۳-۳-۱	
۱۴	جهش ها، جهش زایی و ترمیم DNA	۴-۳-۱	
۱۴	ژن ها	۴-۱	
۱۵	ساخت پروتئین	۱-۴-۱	
۲۰	شبکه های عصبی	۲	
۲۲	شبکه های عصبی بیولوژیکی (طبیعی)	۱-۲	
۲۵	شبکه های عصبی مصنوعی	۲-۲	
۲۸	تشابهات و انتظارات شبکه های عصبی مصنوعی با شبکه های عصبی بیولوژیکی	۳-۲	
۲۸	تشابهات	۱-۳-۲	
۲۸	انتظارات	۲-۳-۲	
۳۰	تاریخچه شبکه های عصبی	۴-۲	
۳۲	کاربرد شبکه های عصبی	۵-۲	
۳۳	شبکه های عصبی چگونه به کار برده می شوند	۶-۲	
۳۳	ساختارهای شبکه های عصبی مصنوعی	۱-۶-۲	

۳۵	یادگیری شبکه های عصبی مصنوعی	۲-۶-۲
۴۷	روش های پیش بینی ساختار دوم پروتئین	۳
۴۸	مهندسی پروتئین	۱-۳
۴۹	ساختار های پروتئین	۲-۳
۵۰	ساختار اول	۱-۲-۳
۵۰	ساختار ثانویه	۲-۲-۳
۵۰	ساختار فرا ثانویه	۳-۲-۳
۵۱	ساختار سوم	۴-۲-۳
۵۱	ساختار چهارم	۵-۲-۳
۵۱	اهمیت پیش بینی ساختار دوم پروتئین	۳-۳
۵۳	پارامترهای بررسی دقت پیش بینی	۴-۳
۵۷	روش های پیش بینی ساختار دوم پروتئین	۵-۳
۵۸	روش PHD	۱-۵-۳
۵۹	مجموعه های یادگیری	۱-۱-۵-۳
۵۹	ساختن الگوهای ورودی - خروجی جهت آموزش شبکه های عصبی	۲-۱-۵-۳
۶۰	شبکه عصبی و ساختار آن	۳-۱-۵-۳
۶۳	یادگیری شبکه	۴-۱-۵-۳
۶۳	نتایج	۵-۱-۵-۳
۶۴	روش SABLE	۲-۵-۳
۶۴	مجموعه یادگیری	۱-۲-۵-۳
۶۵	ساختن الگوهای ورودی جهت آموزش شبکه های عصبی	۲-۲-۵-۳
۶۶	شبکه عصبی و ساختار آن	۳-۲-۵-۳
۶۷	یادگیری شبکه	۴-۲-۵-۳
۶۹	نتایج	۵-۲-۵-۳
۷۱	روش SPINE	۳-۵-۳

۷۲	مجموعه یادگیری شبکه عصبی	۱-۳-۵-۳
۷۳	ساختن الگوهای ورودی جهت آموزش شبکه عصبی	۲-۳-۵-۳
۷۴	یادگیری شبکه عصبی	۳-۳-۵-۳
۷۶	نتایج	۴-۳-۵-۳

۸۶

۸۷	برنامه DSSP	۷-۳
--------------	-------------	-----

۸۸	پیش بینی ساختار دوم پروتئین به کمک شبکه‌های عصبی	۴
----	--	---

۸۹	معرفی	۱-۴
--------------	-------	-----

۹۱	مجموعه یادگیری مدل	۲-۴
--------------	--------------------	-----

۹۱	ساختن الگوهای ورودی جهت آموزش شبکه های عصبی	۳-۴
--------------	---	-----

۹۲	یادگیری شبکه عصبی	۴-۴
--------------	-------------------	-----

۹۳	نتایج	۵-۴
--------------	-------	-----

۱۰۲	نتیجه گیری	۶-۴
---------------	------------	-----

۱۰۳	مراجع	
-----	-------	--

۱۰۸	واژه‌نامه فارسی به انگلیسی	
-----	----------------------------	--

فهرست جداول

مقایسه دقت پیش بینی با استفاده و بدون استفاده از مقادیر سطح دسترسی بدست آمده از ساختارهای پروتئینی معلوم	۷۰	۱-۳
مقایسه دقت پیش بینی با استفاده و بدون استفاده از مقادیر سطح دسترسی پیش بینی شده	۷۰	۲-۳
پارامترهای مختلف روش SPINE	۷۵	۳-۳
شش آزمایش مختلف برای بررسی اثر پارامترهای مختلف . در این جدول PSSM به نحالی اشاره می کند از جداول پروفایل استفاده شده است و PROP بیانگر حالتی است که از هفت ویژگی اسید آمینه ها استفاده شده است	۷۹	۴-۳
پیاده سازی مدل با دو مجموعه یادگیری معروف	۸۰	۵-۳
دقت پیش بینی ساختار دوم برای هر کدام از انواع اسید آمینه	۸۲	۶-۳
مقدار هر یک از خطاهای دسته بندی	۸۴	۷-۳

۸-۳	بررسی اثر بکار بردن هفت ویزگی اسیدآمینه ها و تعداد نرون های لایه پنهان در پیش بینی سطح دسترسی
	. در این جدول PSSM به حالتی اشاره می کند از جداول پروفایل و PROP بیانگر حالتی است که از
۸۵	هفت ویزگی اسیدآمینه ها استفاده شده است .
۹-۳	دقت پیش بینی سطح دسترسی در دو حالت و در چهار آزمایش مختلف .
۸۵	
۱۰-۳	دقت پیش بینی سطح دسترسی به کمک مجموعه یادگیری شامل ۲۶۴۰ زنجیره پروتئینی .
۸۶	
۱-۴	نتایج حاصله از مدل پیشینه در حالتی که از مقادیر عدد تماس استفاده نمی شود .
۹۶	
۲-۴	نتایج حاصله از مدل پیشینه در حالت استفاده از مقادیر عدد تماس .
۹۷	
۳-۴	نتایج حاصله از مدل میانگین در حالتی که از مقادیر عدد تماس استفاده نمی شود .
۹۸	
۴-۴	نتایج حاصله از مدل میانگین در حالت استفاده از مقادیر عدد تماس .
۹۹	
۵-۴	نتایج حاصله از مدل ترکیب در حالتی که از مقادیر عدد تماس استفاده نمی شود .
۱۰۰	
۶-۴	نتایج حاصله از مدل ترکیب در حالت استفاده از مقادیر عدد تماس .
۱۰۱	

فهرست اشکال

۱-۱	تصویر سلول موجود چند سلولی (سمت راست) و سلول موجود تک سلولی (سمت چپ).	۲
۲-۱	اجزاء سلول.	۳
۳-۱	نمای کلی از یک کروموزوم.	۳
۴-۱	مراحل انجام تقسیم میوز.	۵
۵-۱	مولکول ریبوز سمت راست و داکسی ریبوز سمت چپ.	۷
۶-۱	تصویر نمادین ساختار نوکلئوتید.	۸
۷-۱	ساختار مولکولی DNA.	۹
۸-۱	ساختار مولکولی بازهای نیتروژنی RNA و DNA و ساختار نمادین سه بعدی آنها.	۱۰
۹-۱	چگونگی پیوند فسفودی استر.	۱۰
۱۰-۱	پیوند هیدروژنی بین دو رشته DNA و پیوند فسفودی استر نوکلئوتیدهای رشته.	۱۱
۱۱-۱	مراحل همانند سازی رشته DNA.	۱۳
۱۲-۱	قسمت های DNA بیانگر ژن.	۱۵

۱۶	ایجاد mRNA	۱۳-۱
۱۷	مراحل ایجاد یک رشته پروتئین	۱۴-۱
۱۸	فرآیند ایجاد پروتئین	۱۵-۱
۱۸	معادل بودن هر سه حرف نوکلئوتید با یک کدن و معادل بودن هر کدن با یک اسید آمینه	۱۶-۱
۱۹	جدول نمایش معادل اسید آمینه هر کدن	۱۷-۱
۱۹	تصویر نمادین پروتئین هموگلوبین	۱۸-۱
۲۳	یک سلول عصبی (نرون)	۱-۲
۲۶	شباهت یک شبکه عصبی مصنوعی با یک شبکه عصبی طبیعی	۲-۲
۲۷	یک نرون مصنوعی با n ورودی	۳-۲
۲۷	یک تابع فعالیت دو مقداری	۴-۲
۲۷	تابع فعالیت لگاریتم سیگموئید $f(x) = \frac{1}{1 + e^{-ax}}$	۵-۲
۳۴	شبکه عصبی یک لایه به سمت جلو با n نرون در ورودی و m نرون در لایه خروجی	۶-۲
۳۴	یک شبکه عصبی دو لایه به سمت جلو	۷-۲
۳۴	ساختار کلی شبکه های عصبی بازگشتی	۸-۲
۳۵	یک شبکه عصبی المان با یک لایه پنهان و نحوه بروز شدن وزن ها در آن	۹-۲
۳۷	شبکه عصبی پرسپترون یک لایه	۱۰-۲
۳۹	یک شبکه پس انتشار خطا با یک لایه پنهان	۱۱-۲
۵۲	ساختار های مختلف پروتئین	۱-۳
۵۴	مقایسه میزان خوبی پارامتر های SOV و Q_2	۲-۳
۵۶	نشان دادن محاسبه SOV(E)	۳-۳
۵۸	دیاگرام جریان پیش بینی ساختار یک پروتئین	۴-۳
۶۰	پیمایش زنجیره پروتئینی بطوری که هر اسید آمینه در وسط یک پنجره به طول پنج قرار گیرد	۵-۳
۶۱	یک پنجره به طول ۱۳ به مرکز پنج آمین اسید آمینه	۶-۳
۶۱	ساختار شبکه دنباله به ساختار در روش PHD	۷-۳
۶۶	ساختار شبکه عصبی المان با دو لایه پنهان استفاده شده در روش SABLE	۸-۳

۷۵	SPINE مدل	۹-۳
۷۷	بررسی اندازه پنجره روی دقت پیش بینی	۱۰-۳
۸۰	دقت پیش بینی ساختار دوم به عنوان تابعی از دقت شفافیت ray-X ساختار های پروتئین	۱۱-۳
۸۱	دقت پیش بینی ساختار دوم به عنوان تابعی از سطح دسترسی	۱۲-۳
۸۳	دقت پیش بینی ساختار دوم برای دو اسید آمینه به عنوان تابعی از سطح دسترسی آن ها	۱۳-۳
۸۷	بخشی از یک جدول پروفایل	۱۴-۳
۹۰	نقشه تماس برای زنجیره A از پروتئین 1IO1	۱-۴
۹۲	بخشی از یک فایل DSSP برای یک زنجیره پروتئینی	۲-۴
۹۴	ساختار شبکه های عصبی موجود در مدل بیشینه	۳-۴
۹۴	ساختار شبکه های عصبی موجود در مدل میانگین	۴-۴
۹۵	ساختار شبکه های عصبی موجود در مدل ترکیب	۵-۴

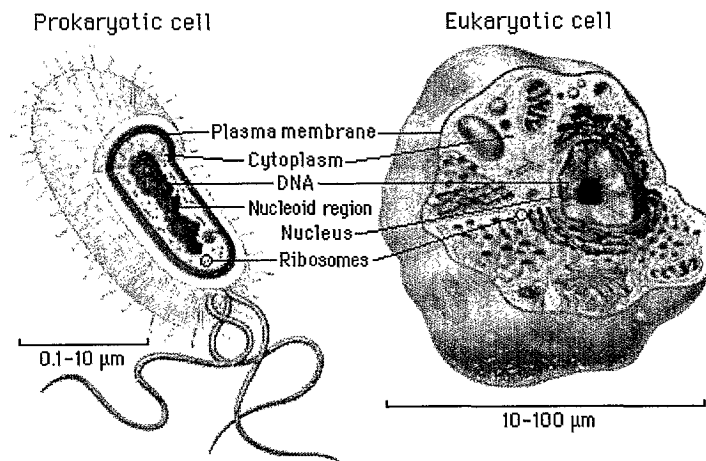
فصل ۱

آشنایی با مقدمات زیست شناسی مولکولی

در این فصل مفاهیم مقدماتی زیست شناسی که برای پایان نامه لازم است بیان می شود. در این قسمت ابتدا سلول را بررسی می کنیم سپس با ساختار شیمیایی سلولها آشنا می شویم. در ادامه ژنها و مراحل ایجاد پروتئین و در انتها ساختارهای پروتئین توضیح داده می شود. مطالب این فصل از مرجع [۳۲] آورده شده است.

۱-۱ سلول

یاخته یا سلول، واحد بنیادین ساختاری و کارکردی همه ارگانیسم های زنده است (شکل ۱-۱). تمام گیاهان و جانوران از سلول تشکیل شده اند. حتی باکتری هم با همه کوچکی خود از یک سلول (فقط یک سلول) تشکیل شده است. بدن ما هم از سلول های مغزی، سلولهای پوستی، سلول های کبدی و غیره تشکیل شده که هر سلول وظایف و ویژگی های خاص خود را داراست. بر



شکل ۱-۱: تصویر سلول موجود چند سلولی (سمت راست) و سلول موجود تک سلولی (سمت چپ).

این اساس تمامی موجودات زنده به دو دسته موجودات زنده چند سلولی و موجودات زنده تک سلولی تقسیم می شوند.

۱-۱-۱ ساختار سلول

هر سلول از سه بخش اصلی: غشاء، سیتوپلاسم و هسته ساخته شده است. غشاء پرده ای نرم است که پیرامون سلول را می پوشاند و آن را از محیط خارجی محافظت می کند. همانطور که در شکل ۱-۲ نشان داده شده است، فضای درون سلول را ماده ای بنام سیتوپلاسم پر می کند که ترکیب اصلی آن آب و پروتئین های محلول در آب اند. یکی از مهمترین اجزای سیتوپلاسمی ریبوزوم ها هستند. ریبوزوم ها، محل های ساخته شدن پروتئین در سلول هستند.

هسته در میان سیتوپلاسم قرار گرفته است و نقش اساسی و اصلی را در زندگی سلول بر عهده دارد. هسته را پوشش دو لایه ای از سیتوپلاسم جدا می کند. در این پوشش حفره ها یا روزنه هایی موسوم به منفذ پیچیده هسته ای وجود دارند که از طریق آنها عمل تبادل مواد بین هسته و سیتوپلاسم انجام می گیرد. درون هسته چندین رشته بلند و بسیار نازک وجود دارد که آنها را کروماتین می نامند. ماده اصلی کروماتین به اختصار DNA نامیده می شود (شکل ۱-۳). در ساختار کروماتین علاوه بر DNA، پروتئین نیز شرکت دارد. DNA فعالیت های سلول را در کنترل دارد. جسمی بنام هستک که ریبوزوم ها در آنجا ساخته می شوند و مایعی بنام شیره هسته نیز درون هسته وجود دارد.

هنگامی که سلول ها در حال آماده شدن برای تقسیم هستند، هر رشته کروماتین مضاعف گردیده سپس فشرده و ضخیم می شود و به جسم مشخصی بنام کروموزوم تبدیل می گردد. تعداد کروموزوم های موجود در هسته از ویژگی های هر نوع موجود زنده است. مثلاً در هر یک از هسته های سلول سازنده بدن انسان ۴۶ کروموزوم وجود دارد.