

مروری بر روش‌های پیش‌بینی سایت‌های تعامل کمپلکس‌های آنتی‌بادی-پروتئین مبتنی بر هوش مصنوعی

مرضیه عبدی^۱، مهدی سعادت‌مند طرزجان^{۲*}، محمد طاهرزاده ثانی^۳، علیرضا حق‌پرست^۴

• پذیرش مقاله: ۹۶/۱۲/۱۰

• دریافت مقاله: ۹۶/۶/۳۰

مقدمه: سرطان یکی از مهمترین چالش‌های بهداشتی قرن اخیر و آینده می‌باشد. طراحی داروهای ضدسرطان هدفمند، مبتنی بر آنتی‌بادی‌های مونوکلونال، نیازمند درک مکانیسم تعامل آنتی‌بادی-پروتئین در سطح باقی‌مانده‌ها است. اولین گام برای تولید آنتی‌بادی‌های مونوکلونال، پیش‌بینی ساختار آن‌ها می‌باشد.

روش: در این مقاله، مهم‌ترین تحقیقات منتشر شده در پایگاه‌های اطلاعاتی PubMed، ScienceDirect، Springer و IEEE، برای پیش‌بینی سایت‌های تعامل کمپلکس‌های آنتی‌بادی-پروتئین و تعیین ساختار مؤثر آنتی‌بادی‌ها، به صورت ساختاریافته مورد بررسی قرار گرفت. معمولاً برای این منظور، از شبکه‌های عصبی مصنوعی یا وب‌سرورها استفاده می‌شود. به علاوه، برخی محققین نیز از الگوریتم‌های تکاملی برای پیش‌بینی ساختار مؤثر آنتی‌بادی‌ها استفاده نموده‌اند. بر این اساس، تعداد ۱۴ روش مبتنی بر ساختار فضایی پروتئین‌ها، ۲۸ روش مبتنی بر توالی اسیدهای آمینه (مستقل از ساختار فضایی) و ۱۸ روش پیش‌بینی ساختار آنتی‌ژن/آنتی‌بادی مورد بررسی قرار گرفت.

نتایج: مطالعه حاضر نشان داد که دقت روش‌های مبتنی بر ساختار فضایی تا ۸۰٪ قابل افزایش می‌باشد؛ در حالی که دقت روش‌های پیش‌بینی مبتنی بر توالی اسیدهای آمینه به ندرت بهتر از ۷۵٪ بود. از آنجا که ساختار فضایی بسیاری از آنتی‌بادی‌ها در دسترس نمی‌باشد؛ برخی محققین برای بهبود دقت (حتی تا ۹۶٪)، تنها از توالی آنتی‌بادی‌های مؤثر بر چند آنتی‌ژن مشابه در آموزش شبکه عصبی استفاده نموده‌اند؛ لذا با توجه به دقت بالای به دست آمده، پیشنهاد می‌شود که از روش اخیر برای پیش‌بینی ساختار آنتی‌بادی‌های مونوکلونال استفاده گردد.

نتیجه‌گیری: در این مقاله، پس از مرور روش‌های موجود برای پیش‌بینی سایت‌های تعامل آنتی‌بادی-پروتئین، پیشنهادهایی برای پیش‌بینی ساختار آنتی‌بادی‌های مونوکلونال پیشنهاد گردید.

کلید واژه‌ها: ایمونولوژی، آنتی‌بادی‌های مونوکلونال، کمپلکس آنتی‌بادی-پروتئین، هوش مصنوعی، شبکه‌های عصبی

• ارجاع: عبدی مرضیه، سعادت‌مند طرزجان مهدی، طاهرزاده ثانی محمد، حق‌پرست علیرضا. مروری بر روش‌های پیش‌بینی سایت‌های تعامل کمپلکس‌های آنتی‌بادی-پروتئین مبتنی بر هوش مصنوعی. مجله انفورماتیک سلامت و زیست پزشکی ۱۳۹۷؛ ۵(۱): ۶۹-۵۶.

۱. کارشناسی ارشد مهندسی پزشکی (بیوالکتریک)، آزمایشگاه تصویربرداری پزشکی، گروه مهندسی برق، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران

۲. دکتری مهندسی پزشکی (بیوالکتریک)، استادیار آزمایشگاه تصویربرداری پزشکی، گروه مهندسی برق، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران

۳. دکتری مهندسی برق (الکترونیک)، استادیار، گروه مهندسی برق، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران

۴. دکتری پاتوبیولوژی، دانشیار، گروه پاتوبیولوژی، دانشکده دامپزشکی، دانشگاه فردوسی مشهد، مشهد، ایران

* نویسنده مسئول: مشهد، میدان آزادی، دانشگاه فردوسی مشهد، دانشکده مهندسی، گروه برق

• Email: saadatmand@um.ac.ir

• شماره تماس: ۰۵۱۳۷ ۵۱۳۸۰ ۹۸۵۱۳

مقدمه

سرطان یکی از نگران‌کننده‌ترین بیماری‌ها در سراسر جهان بوده و هزینه‌های بسیار زیادی به خانواده‌ها و دولت‌ها تحمیل می‌کند. روش‌های متداول برای درمان سرطان همچون جراحی، پرتو درمانی و شیمی درمانی عوارض جانبی زیادی به دنبال دارند؛ زیرا علاوه بر نابودی سلول‌های سرطانی موجب مختل شدن عملکرد سلول‌های طبیعی نیز می‌شوند؛ لذا محققان به سمت گسترش درمان‌های هدفمند برای سرطان گام برداشته‌اند [۱-۳].

درمان هدفمند مبتنی بر آنتی‌بادی در سال ۱۹۸۳ توسط [۱] Oldham و Dillman [۱] معرفی گردید. آن‌ها با آزمایش آنتی‌بادی‌های مونوکلونال موشی در حیوانات و تعداد محدودی از بیماران نشان دادند که این نوع آنتی‌بادی‌ها می‌توانند پیشرفت عمده‌ای در درمان سرطان ایجاد نمایند. این روش به عنوان یکی از مؤثرترین روش‌های درمان سرطان شناخته شده است؛ زیرا به صورت اختصاصی، تنها بر نابودی سلول‌های سرطانی (با یک درصد خطا) تمرکز دارد. تاکنون، بیش از ۱۰۰ آنتی‌بادی مونوکلونال مختلف در پژوهش‌های تجربی در حیوانات و کارآزمایی‌های انسانی برای درمان سرطان مورد استفاده قرار گرفته‌اند که از میان آن‌ها تنها شمار اندکی برای مصارف بالینی توصیه شده‌اند [۲]. این نوع درمان نیازمند درک عمیقی از سرم‌شناسی سرطان، تکنیک‌های مهندسی پروتئین، مکانیسم عمل سلول‌های سرطانی و نحوه تعامل بین سیستم ایمنی و سلول‌های سرطانی است. یک چالش کلیدی در این نوع درمان، شناسایی آنتی‌ژن‌های مناسب برای درمان‌های مبتنی بر آنتی‌بادی است [۳]. علاوه بر این، برای درمان مؤثرتر، می‌توان آنتی‌بادی‌های مونوکلونال را با مولکول‌های سمی، رادیوایزوتوپ‌ها و یا داروهای ضد توموری ترکیب نمود. به این ترتیب، مواد سمی به طور اختصاصی در مجاورت سلول‌های توموری قرار گرفته و فرایند تخریب سلول‌ها تسریع می‌گردد [۲].

معمولاً برای پیش‌بینی میزان برهم‌کنش آنتی‌بادی - پروتئین از روش‌های هوشمند استفاده می‌شود. در این روش‌ها معمولاً فرایند پیش‌بینی بر مبنای تعامل بین باقی‌مانده‌های موجود در سایت تعامل (باقی‌مانده‌های واسطه) بین دو پروتئین انجام می‌شود. بدیهی است که اگر ساختار سه‌بعدی دو پروتئین با قدرت تفکیک بالا در دسترس باشد، (به دلیل استفاده از خصوصیات فضایی کمپلکس آنتی‌بادی - پروتئین) پیش‌بینی ساختار مولکولی آنتی‌بادی با دقت بهتری انجام می‌شود [۴-۱۹].

اما از آنجا که معمولاً ساختارهای سه‌بعدی در دسترس نیستند، برای پیش‌بینی ساختار مولکولی آنتی‌بادی باید از توالی اسید آمینه‌ها استفاده نمود که با کاهش دقت (نسبت تعداد باقیمانده‌های واسطه به درستی پیش‌بینی شده به تعداد کل پیش‌بینی‌ها) و پوشش (نسبت تعداد سایت‌های پروتئین-پروتئین پیش‌بینی شده به تعداد سایت‌های پروتئین-پروتئین موجود در دنباله پروتئینی موردنظر) همراه است [۲۶-۲۰]. برخی از روش‌های مذکور در قالب وب سرور (web-server) در دسترس می‌باشند [۲۷-۴۶]. با توجه به اهمیت آنتی‌ژن‌ها و آنتی‌بادی‌ها در سیستم ایمنی بدن انسان، برخی محققین تحقیقات مستقلی در زمینه پیش‌بینی ساختار آنتی‌ژن‌ها و ساختار آنتی‌بادی‌ها انجام داده‌اند [۴۴-۴۷].

با توجه به اینکه روش‌های آزمایشگاهی برای مطالعه ساختار آنتی‌بادی‌های مونوکلونال بسیار زمان‌بر و گران می‌باشد، پیش‌بینی ساختار مؤثر آنتی‌بادی‌های مونوکلونال می‌تواند منجر به صرفه‌جویی قابل توجه در زمان و هزینه گردد. در این مقاله، ابتدا، الگوریتم‌های هوشمند ارائه شده توسط محققین، در زمینه پیش‌بینی ساختارهای مؤثر پروتئین‌ها، آنتی‌ژن‌ها و آنتی‌بادی‌ها بررسی شده است. سپس، با مقایسه و بررسی روش‌های موجود مختلف، بهترین راه کار برای تعیین ساختار مؤثر آنتی‌بادی‌های مونوکلونال پیشنهاد گردید.

روش

در این تحقیق، ابتدا تعداد ۶ واژه کلیدی مناسب شامل کمپلکس پروتئین-پروتئین، کمپلکس آنتی‌بادی-پروتئین، آنتی‌بادی مونوکلونال، ایمونولوژی، هوش مصنوعی و شبکه عصبی برای انجام یک جستجوی جامع و کامل در زمینه روش‌های پیش‌بینی سایت‌های تعامل پروتئین‌ها انتخاب شد. سپس، تمامی مقالات مرتبط با واژه‌های کلیدی فوق از پایگاه‌های اطلاعاتی PubMed، ScienceDirect، Springer و IEEE استخراج شدند. با توجه به هدف این تحقیق مبنی بر بررسی و مرور ساختاریافته روش‌های مبتنی بر هوش مصنوعی، در همه مراحل جستجو، از ترکیب کلید واژه هوش مصنوعی یا شبکه عصبی با کلید واژه‌های دیگر استفاده شد. همچنین، به منظور محدود نمودن تعداد نتایج جستجو و دستیابی به مقاله‌های مرتبط‌تر جستجو به بخش‌های عنوان، کلمات کلیدی و چکیده مقاله‌های موجود در پایگاه‌های مذکور محدود گردید. در نهایت، با مطالعه چکیده مقاله‌های استخراج شده در فرایند جستجو و تبادل نظر، نویسندگان تعداد ۶۰ مقاله مرتبط با

پیش بینی سایت های تعامل پروتئین ها را انتخاب نمودند. همچنین سعی شد که فرآیند جستجو تا حد امکان جامع بوده و همه تحقیقات مرتبط را شامل گردد. به علاوه، در فرآیند انتخاب مقاله ها هیچ محدودیتی لحاظ نشده و همه مقاله های مرتبط با موضوع این تحقیق، مورد بررسی قرار گرفته است. در مرحله بعد، به دسته بندی مقاله های جمع آوری شده پرداختیم. به طور کلی، تحقیقات انجام شده در خصوص پیش بینی سایت های تعامل پروتئین ها را به دو دسته روش های پیش بینی ساختار مؤثر آنتی بادی و آنتی ژن تقسیم شد. اگرچه آنتی بادی ها و آنتی ژن ها نیز در حقیقت نوعی پروتئین می باشند؛ اما به دلیل اهمیت آن ها در مطالعه و شناخت عملکرد سیستم ایمنی بدن و خصوصاً پیش بینی ساختار آنتی بادی های مونوکلونال، تحقیقات مربوط به آنتی بادی ها و آنتی ژن ها در دسته جداگانه ای قرار داده شد. به عبارت دیگر، در دسته اول تحقیقات انجام شده در زمینه پیش بینی سایت های تعامل پروتئین ها به طور عام و در دسته دوم، تحقیقات اختصاصی در رابطه با آنتی بادی ها و آنتی ژن ها قرار داده شده اند. به علاوه، دسته اول براساس استفاده یا عدم استفاده از اطلاعات ساختار فضایی پروتئین ها در فرآیند پیش بینی به دو زیربخش تقسیم شد. این در حالی است که در دسته دوم نیز الگوریتم های مربوط به ساختار آنتی بادی ها و ساختار آنتی ژن ها در دو زیربخش جداگانه دسته بندی شده اند.

در مجموع، می توان گفت که از میان ۶۰ مقاله استخراجی، ۱۴ روش مبتنی بر ساختار فضایی پروتئین ها، ۲۸ روش مبتنی بر توالی اسیدهای آمینه (مستقل از ساختار فضایی) پروتئین ها، ۱۱ روش پیش بینی ساختار آنتی ژن ها و ۷ روش پیش بینی سایت های تعامل آنتی بادی ها مورد بررسی قرار گرفته است. با وجود این، تحقیق مستقلاً در خصوص پیش بینی ساختار مؤثر آنتی بادی های مونوکلونال یافت نشد.

نتایج

۳-۱- پیش بینی سایت های تعامل پروتئین-پروتئین:

درک مکانیسم تعامل میان پروتئین ها به منظور طراحی داروها، نیازمند شناسایی باقی مانده های در حال تعامل است. بدیهی است که اگر ساختار فضایی پروتئین در دسترس باشد، شناسایی قابل اطمینان تر خواهد بود؛ اما برای گستره وسیعی از پروتئین ها، ساختار فضایی با قدرت تفکیک بالا در دسترس نیست. در نتیجه، تنها باید با استفاده از توالی اسیدهای آمینه به پیش بینی باقی مانده ها (اسید آمینه هایی که با تشکیل پیوند پپتیدی، مولکول آب خود را از دست داده اند) پرداخت. بر این اساس می توان روش های پیش بینی ساختار مولکولی را به دو دسته با دسترسی به ساختار فضایی و بدون دسترسی به ساختار فضایی، تقسیم نمود [۵].

• پیش بینی با استفاده از اطلاعات ساختار فضایی

زمانی که ساختار فضایی پروتئین (جهت گیری فضایی اسید آمینه ها) در دسترس باشد، معمولاً برای پیش بینی بهتر از

جدول ۱: روش های پیش بینی سایت های تعامل پروتئین-پروتئین با استفاده از اطلاعات ساختار فضایی

روش پیش بینی	سال	نوع طبقه بند	مشخصات طبقه بند	نتایج ارزیابی
Shan و Zhou [۶]	۲۰۰۱	دو شبکه عصبی پرسپترون متوالی	شبهه اول: ۴۲۰ نرون ورودی، ۷۵ نرون در لایه پنهان و ۲ نرون خروجی شبهه دوم: ۶۰ نرون ورودی، ۳۰ نرون در لایه پنهان و ۲ نرون خروجی	دقت ۷۰٪، پوشش ۴۷٪
Fariselli و همکاران [۷]	۲۰۰۲	شبکه عصبی پیشرو	۲۲۰ نرون ورودی، ۴ نرون در لایه پنهان و ۱ نرون خروجی	دقت ۷۲٪، پوشش ۵۶٪
Takagi و Koike [۸]	۲۰۰۴	ماشین بردار پشتیبان	کرنل گاوسی	دقت ۵۶٪، پوشش ۸۷٪
Abagyan و Bordner [۹]	۲۰۰۵	ماشین بردار پشتیبان	کرنل گاوسی	دقت ۸۰٪، پیش بینی مثبت ۳۹٪، حساسیت ۶۴٪
Wang و همکاران [۱۰]	۲۰۰۶	ماشین بردار پشتیبان	کرنل گاوسی	دقت ۶۵/۴٪، حساسیت ۶۶/۳٪، ویژگی ۴۹/۷٪
Chung و همکاران [۱۱]	۲۰۰۶	ماشین بردار پشتیبان	کرنل پایه شعاعی	دقت ۵۱٪، حساسیت ۶۳/۷٪
Nguyen و Rajapakse [۱۲]	۲۰۰۶	ماشین بردار پشتیبان	کرنل گاوسی	دقت ۷۳/۸٪، حساسیت ۶۸٪، ویژگی ۷۶٪
Dong و همکاران [۱۳]	۲۰۰۷	ماشین بردار پشتیبان	کرنل پایه شعاعی	دقت ۷۳/۵٪، حساسیت ۷۰٪
Li و همکاران [۱۴]	۲۰۰۸	ماشین بردار پشتیبان	کرنل پایه شعاعی	حساسیت ۶۰/۶٪، ویژگی ۵۳/۴٪
Liu و همکاران [۱۵]	۲۰۰۹	ماشین بردار پشتیبان	مدل مارکوف پنهان	دقت ۶۹/۳٪، حساسیت ۵۳/۶٪، ویژگی ۵۸/۶٪
Sikic و همکاران [۱۶]	۲۰۰۹	الگوریتم جنگل تصادفی	-	پیش بینی مثبت ۷۶٪، حساسیت ۳۸٪
Chen و همکاران [۱۷]	۲۰۱۲	شش شبکه عصبی تابع پایه شعاعی	آموزش با استفاده از روش بهینه سازی ازدحام ذرات	دقت ۸۰٪
Li و همکاران [۱۸]	۲۰۱۲	الگوریتم جنگل تصادفی	-	دقت ۶۷٪، حساسیت ۷۹٪، ویژگی ۵۵٪
Hwang و همکاران [۱۹]	۲۰۱۴	ماشین بردار پشتیبان	کرنل خطی	؟

آنالیز تکه‌ای (patch analysis)، مساحت سطح در دسترس (accessible surface area)، همسایگان فضایی باقی‌مانده‌ها و ... استفاده می‌شود [۵]. در جدول ۱، ۱۴ روش مختلف پیش‌بینی سایت‌های تعامل با استفاده از ساختار فضایی پروتئین‌ها معرفی گردید. همچنین، برای هر روش، طبقه‌بند مورد استفاده و عملکرد آن برحسب معیارهای مختلف نیز گزارش شد. به عنوان مثال، Shan و Zhou [۶]، روش PPISP را ارائه نموده‌اند. در این روش، با استفاده از دو شبکه عصبی پرسپترون متوالی، واسط (سایت تعامل) بودن یا نبودن یک باقیمانده سطحی پروتئین مورد بررسی قرار می‌گیرد (یک باقیمانده واسط یا سایت تعامل است اگر بتواند با باقیمانده‌ای از یک پروتئین دیگر در یک کمپلکس شرکت نماید). ورودی‌های شبکه عصبی ابتدایی شامل پروفایل توالی یک باقی‌مانده سطحی و ۱۹ باقی‌مانده از نزدیک‌ترین همسایگان فضایی آن و همچنین مساحت قابل دسترس آن‌ها برای حلال بوده است در حالی که ورودی‌های شبکه عصبی دوم شامل مقادیر خروجی شبکه اول و مساحت سطح قابل دسترس ۲۰ باقی‌مانده بوده است.

در مقاله دیگری که توسط Fariselli و همکاران [۷] ارائه شد، پیش‌بینی سایت‌های تعامل تنها به کمپلکس‌های پروتئینی از نوع هترودایمر (دو پروتئین غیر یکسان) محدود شده است. در این روش، برای پیش‌بینی از یک شبکه عصبی پیش‌رو با ورودی پروفایل دنباله یک باقی‌مانده سطحی و ده باقی‌مانده از نزدیک‌ترین همسایه‌های فضایی آن استفاده شده است.

Koike و Takagi [۸] با استفاده از ماشین بردار پشتیبان SVM (Support Vector Machine) به پیش‌بینی سایت‌های تعامل پرداختند. آن‌ها برای به دست آوردن بهترین پیش‌بینی، از ویژگی‌های مختلفی شامل پروفایل توالی، مساحت قابل دسترس باقی‌مانده هدف و ده همسایه فضایی آن و نیز سهم سایت‌های تعامل از کل توالی برای آموزش طبقه‌بند استفاده نمودند که بهترین نتایج با ترکیبی از آن‌ها و با بهره‌گیری از کرنل گاوسی حاصل شد. لازم به ذکر است که پروفایل توالی میزان شباهت توالی موردنظر با توالی‌های هم‌خانواده‌اش را در یک ماتریس بیان می‌کند که این اطلاعات با هم‌ترازسازی توالی‌های پروتئین با استفاده از نرم‌افزارهای خاص (مانند PSI-BLAST) به دست می‌آید.

Bordner و Abagyan [۹] از پروفایل توالی و نرخ تکاملی (محاسبه شده با استفاده از الگوریتم بیزی) برای کد کردن پنجره ورودی بهره بردند. این دو ویژگی برای پنجره‌ای شامل باقی‌مانده تحت بررسی و ۱۴ باقی‌مانده از نزدیک‌ترین همسایه‌های

فضایی آن محاسبه شده است. آن‌ها نیز از طبقه‌بند SVM با کرنل گاوسی برای نگاشت ورودی استفاده نمودند.

همچنین Wang و همکاران [۱۰] برای پیش‌بینی سایت تعامل از پروفایل توالی و نرخ تکاملی به عنوان ورودی طبقه‌بند SVM استفاده نمودند. این دو ویژگی برای باقی‌مانده هدف و ۱۰ باقی‌مانده از نزدیک‌ترین همسایه‌های فضایی آن محاسبه و در بردار ورودی قرار گرفته است.

درمقاله دیگری، Chung و همکاران [۱۱] از SVM با تابع کرنل پایه شعاعی برای طبقه‌بندی داده‌ها استفاده کردند. برای آموزش SVM، دو بردار ورودی مختلف به کار رفته و نتایج آن‌ها مقایسه شده است. اولین بردار ورودی شامل پروفایل توالی و مساحت در دسترس پنجره تحت بررسی (شامل باقی‌مانده مورد نظر و ۱۱ باقی‌مانده از نزدیک‌ترین همسایه‌های فضایی آن) است. بردار ورودی دوم، علاوه بر ورودی‌های قبلی شامل نمرات حفاظت وابسته به پنجره ورودی نیز می‌باشد. همچنین، Nguyen و Rajapakse [۱۲] از پروفایل توالی و مساحت در دسترس برای کدگذاری پنجره ورودی استفاده نمودند. هر پنجره شامل باقی‌مانده هدف و ۱۴ باقی‌مانده از نزدیک‌ترین همسایگان فضایی آن بوده است. برای طبقه‌بندی داده‌ها نیز از SVM با کرنل گاوسی استفاده گردیده است.

Dong و همکاران [۱۳] اساس پیش‌بینی خود را بر مبنای گرایش اسیدآمین‌های موجود در سایت‌های تعامل (گرایش هر اسیدآمین متناسب است با نسبت فراوانی آن در باقیمانده واسط به فراوانی آن در کل باقیمانده‌های سطحی)، مساحت سطح در دسترس و پروفایل باینری قرار دادند. ویژگی‌های فوق برای پنجره‌ای شامل باقی‌مانده هدف و ۱۱ همسایه فضایی آن محاسبه و به عنوان ورودی به طبقه‌بند SVM با کرنل تابع پایه شعاعی داده شده است.

Li و همکاران [۱۴]، به منظور پیش‌بینی باقی‌مانده‌های واسط، مفهوم جدیدی برای آن‌ها ارائه دادند. آن‌ها برای تمایز بیشتر کلاس‌ها، مجموعه مثبت را تنها باقی‌مانده‌های واسطی در نظر گرفتند که نسبت همسایگان واسط آن‌ها به کل همسایگان بیشتر از آستانه بهینه ۰/۲ باشد. در مجموعه داده نیز دو باقی‌مانده در حال تعامل در نظر گرفته می‌شوند، اگر فاصله بین اتم‌های آن‌ها کمتر از ۵ آنگستروم باشد. پنجره ورودی شامل ۴ همسایه فضایی و ۸ همسایه از توالی باقی‌مانده هدف بوده است. برای کدگذاری همسایگان فضایی از ۱۰ ویژگی شامل تعداد اتم‌ها، تعداد بار الکترواستاتیکی، تعداد باند هیدروژنی (ویژگی‌های فیزیکی - شیمیایی)، آب‌گریزی، مساحت نسبی در دسترس

حفاظت، سطح در دسترس و تنوع توالی در نظر گرفته شده است. به علاوه، هر شبکه عصبی یک لایه پنهان با نرون‌هایی از نوع RBF (Radial Basis Function) و یک نرون خروجی نیز داشته است. برای تنظیم مرکز و عرض توابع RBF و وزن‌های بین نرون‌های لایه‌های پنهان و خروجی، روش بهینه‌سازی ازدحام ذرات (Particle Swarm Optimization) PSO به خدمت گرفته شده است. در نهایت، پاسخ نهایی با ادغام نتایج حاصل از شش شبکه عصبی به روش میانگین وزنی به دست آمده است. نتایج تجربی در این تحقیق، بیانگر آن است که روش‌های تصمیم‌گیری جمعی می‌توانند دقت و پوشش بیشتری فراهم آورند.

Li و همکاران [۱۸] با بهره‌گیری از یک پیش‌بینی‌کننده جدید مبتنی بر معیار حداکثر ارتباط-حداقل افزونگی، mRMR (minimum Redundancy Maximal Relevance) و الگوریتم جنگل تصادفی به پیش‌بینی سایت‌های تعامل پروتئین پرداختند. در این روش، برای پیش‌بینی علاوه بر اطلاعات توالی، از ویژگی‌های ساختار دوم و سوم پروتئین نیز بهره‌برده شده است. برای کدگذاری پنجره ورودی (شامل ۲۱ باقی‌مانده) از ۳۴ ویژگی شامل ماتریس امتیازدهی موقعیت، خصوصیات فیزیکی-شیمیایی اسیدآمینه‌ها، نمره بی‌نظمی، شکل اسیدهای آمینه در ساختار دوم و خصوصیات ساختار سوم شامل شاخص عمق، شاخص بیرون‌زدگی، مساحت سطح در دسترس حلال، مساحت سطح مولکولی و انحنا سطح برای هر باقی‌مانده استفاده شده است. در ادامه، با استفاده از روش حداکثر ارتباط - حداقل افزونگی ویژگی‌های کم اهمیت تر حذف و تعداد آن به ۵۱ عنصر کاهش یافته است. در نهایت برای طبقه‌بندی داده‌ها از الگوریتم جنگل تصادفی استفاده شده است.

Hwang و همکاران [۱۹] علاوه بر ساختار پروتئین مورد نظر از ساختار پروتئین شریک نیز برای پیش‌بینی سایت‌های تعامل کمک گرفتند (اصطلاحاً به این نوع الگوریتم‌ها، روش‌های مبتنی بر اتصال (داکینگ) پروتئین‌ها گفته می‌شود). برای این منظور، ابتدا برای هر اتم باقی‌مانده، تعداد تماس‌ها با اتم‌های پروتئین شریک محاسبه گردیده و پس از نرمال‌سازی، ضریبی به هر اتم اختصاص داده می‌شود. سپس، هر باقی‌مانده براساس مقایسه مجموع ضرایب اتم‌هایش با آستانه‌ای مشخص به عنوان واسط یا غیر واسط در نظر گرفته می‌شود. این روش در مقایسه با دیگر روش‌های متداول پیش‌بینی سایت‌های تعامل پروتئین (که توجهی به ساختار پروتئین شریک ندارند) عملکرد بهتری داشته است. به علاوه، به منظور بهبود نتایج، الگوریتم فوق با استفاده از

حلال، ساختار دوم باقی‌مانده (شامل ۳ نوع: helix, coil, sheet)، نمره حفاظت، محیط زنجیره جانبی (براساس درجه غیرسطحی و مساحت زنجیره جانبی که توسط اتم‌های قطبی پوشیده شده است به ۶ ناحیه تقسیم می‌شود)، فاصله توالی و فاصله فضایی (حداقل فاصله بین باقیمانده موردنظر با باقیمانده هدف) بهره‌برده شده است. برای باقی‌مانده هدف نیز علاوه بر ۱۰ ویژگی مذکور، تعداد اتم‌ها و تعداد باقی‌مانده‌ها به تفکیک نوع برای کدگذاری محاسبه گردیده است (در مجموع، بردار ورودی شامل $10 \times 8 + 10 \times 4 + 18 + 20 = 158$ عنصر است). برای طبقه بندی داده‌ها نیز از SVM با کرنل تابع پایه شعاعی استفاده شده است.

Liu و همکاران [۱۵] برای پیش‌بینی سایت‌های تعامل پروتئین (برای مجموعه‌ای شامل هتروکمپلکس‌ها و هوموکمپلکس‌ها)، استفاده از SVM مدل مارکوف پنهان، را پیشنهاد دادند. آن‌ها از پروفایل توالی و مساحت نسبی در دسترس حلال پنجره‌ای شامل باقی‌مانده هدف و ۱۳ باقی‌مانده از نزدیک‌ترین همسایه‌های فضایی آن برای کدگذاری استفاده نمودند. برچسب باقی‌مانده‌های پنجره ورودی بر اساس تعریف باقی‌مانده واسط و غیر واسط تعیین شده و ارتباط بین برچسب‌ها نیز به عنوان یک ویژگی در کدگذاری پنجره به کار رفته است (به طوری که اگر دو برچسب مجاور برابر باشند، مقدار این ویژگی صفر و در غیر این صورت یک خواهد بود).

Siki و همکاران [۱۶] پیش‌بینی سایت‌های تعامل را با استفاده از توالی و خصوصیات فضایی باقی‌مانده‌ها انجام دادند. خصوصیات فضایی به کار رفته شامل مساحت در دسترس و مساحت نسبی باقی‌مانده‌های غیر قطبی، ماکزیمم شاخص عمق، میانگین شاخص عمق و مینیمم شاخص برآمدگی می‌باشد. پنجره ورودی شامل N ($1 < N < 9$) باقی‌مانده بوده و از الگوریتم جنگل تصادفی برای پیش‌بینی واسط بودن آن استفاده گردیده است. برای هر پنجره به ازای طول‌های مختلف، ۹ پیش‌بینی انجام شده و در نهایت، اگر تعداد پیش‌بینی‌های مثبت (پیش‌بینی مثبت یعنی باقیمانده مرکزی به عنوان واسط شناخته شود) بیشتر از منفی بود، پنجره به عنوان سایت تعامل در نظر گرفته می‌شد.

Chen و همکاران [۱۷] از شش شبکه عصبی تابع پایه شعاعی (RBF) برای پیش‌بینی سایت‌های تعامل پروتئین استفاده نمودند. ورودی‌های هر کدام از شبکه‌ها به ترتیب پنجره‌هایی به طول ۱، ۳، ۵، ۷، ۹ و ۱۱ باقی‌مانده بوده و برای هر باقی‌مانده ۶ ویژگی شامل پروفایل توالی، آنتروپی، آنتروپی نسبی، نمره

آنتروپی هر ستون به هر باقی‌مانده، یک ضریب ارزش اختصاص می‌دهد.

ورودی طبقه‌بند شامل پروفایل توالی و ارزش تکاملی باقی‌مانده هدف و ۸ باقی‌مانده از نزدیکترین همسایه‌های آن خواهد بود.

Ofran و Rost [۲۳] با ترکیب اطلاعات ساختاری با اطلاعات تکاملی به تکمیل کار پیشین خود پرداختند. در این کار، علاوه بر ویژگی‌های فوق از ساختار دوم و مساحت در دسترس حلال برای پیش‌بینی واسط بودن باقی‌مانده موردنظر استفاده شده است.

Chen و Jeong [۲۴] پیش‌بینی سایت‌های تعامل را با استفاده از روش تلفیقی انجام دادند. در این روش، از ۳ گروه ویژگی برای کد کردن پنجره‌ای شامل باقی‌مانده هدف و ۲۰ باقی‌مانده از نزدیک‌ترین همسایه‌های آن بهره برده شد: گروه اول شامل ویژگی‌های فیزیکی-شیمیایی (از جمله آب‌گریزی، آب-دوستی، تمایل به حلال، نقطه ایزوالکتریک، جرم و ...) و نمره حفاظت، گروه دوم شامل فاصله باقی‌مانده هدف از ۲۰ باقی‌مانده همسایه آن و گروه سوم شامل ماتریس امتیازدهی موقعیت می‌باشد. در جدول ۲، ساختار طبقه‌بندهای مورد استفاده در روش‌های فوق و نتایج ارزیابی آن‌ها گزارش شده است. علاوه بر این، وب‌سرورهای متعددی (که به صورت برخط و از طریق اینترنت در دسترس می‌باشند) نیز به صورت کلی، به پیش‌بینی تعاملات پروتئین‌ها می‌پردازند. از جمله وب‌سرورهای پرکاربرد و مشهور می‌توان به موارد زیر اشاره کرد.

SVM (با کرنل خطی) با روش meta-PPISP [۳۵] نیز ترکیب شده است.

• پیش‌بینی با استفاده از ساختار مولکولی (بدون ساختار فضایی)

در روش‌های پیش‌بینی ساختار مولکولی بدون دسترسی به ساختار فضایی معمولاً از اطلاعات مربوط به مولکول پروتئین شامل ترکیب اسیدآمینه‌ها در توالی، نظم خاص آن‌ها و اطلاعات تکاملی به عنوان ورودی شبکه عصبی استفاده می‌شود [۵].

به عنوان مثال، Bock و Gough [۲۰] با همین رویکرد به پیش‌بینی تعاملات پروتئین‌ها پرداختند. آن‌ها از خواص فیزیکی-شیمیایی باقی‌مانده‌ها شامل بار الکتریکی، آب‌گریزی و فشار سطح برای کدگذاری توالی بهره بردند.

در کار دیگری، Ofran و Rost [۴] برای پیش‌بینی ساختار مولکولی از یک شبکه عصبی پیش‌رو استفاده نمودند. در این روش، ورودی شبکه عصبی شامل پروفایل توالی ۹ باقی‌مانده متوالی به مرکز باقی‌مانده تحت بررسی بوده است. به علاوه، به هر بردار یک متغیر با مقدار یک برای باقی‌مانده تحت بررسی و مقدار صفر برای سایر باقی‌مانده‌ها اضافه گردیده است. همچنین، Yan و همکاران [۲۱] از پروفایل توالی پنجره‌ای شامل ۱۱ باقی‌مانده متوالی به مرکز باقی‌مانده هدف برای تشخیص کمپلکس‌های آنتی‌بادی - آنتی‌ژن استفاده نمودند.

در مقاله دیگری که توسط Res و همکاران [۲۲] ارائه شد از پروفایل توالی و اطلاعات تکاملی برای پیش‌بینی سایت تعامل استفاده کردند. اطلاعات تکاملی براساس درخت فیلوژنتیک و

جدول ۲: روش‌های پیش‌بینی سایت‌های تعامل پروتئین-پروتئین با استفاده از اطلاعات ساختار مولکولی

روش پیش‌بینی	سال	نوع طبقه‌بند	مشخصات طبقه‌بند	نتایج ارزیابی
Bock و Gough [۲۰]	۲۰۰۱	ماشین بردار پشتیبان	-	دقت ۸۰٪
Ofran و Rost [۴]	۲۰۰۳	شبکه عصبی پیش‌رو	۱۰ نرون ورودی، ۳۰۰ نرون در لایه پنهان و ۲ نرون خروجی	دقت ۷۰٪، پوشش ۲۰٪
Yan و همکاران [۲۱]	۲۰۰۴	ماشین بردار پشتیبان	کرنل چندجمله‌ای	حساسیت ۸۲/۳٪، ویژگی ۸۱٪
Res و همکاران [۲۲]	۲۰۰۵	ماشین بردار پشتیبان	کرنل چندجمله‌ای	دقت ۵۸/۷٪، ارزش پیش‌بینی مثبت ۲۶٪
Ofran و Rost [۲۳]	۲۰۰۷	شبکه عصبی پیش‌رو	-	دقت ۶۸٪
Jeong و Chen [۲۴]	۲۰۰۹	الگوریتم جنگل تصادفی	-	حساسیت ۷۳٪، ویژگی ۷۰٪

سطحی و همسایگان فضایی آن به عنوان ورودی شبکه عصبی استفاده شده است [۲۷].

PPI-PRED: در این وب‌سرور، ابتدا پیچ‌های سطح تولید شده و سپس برای راس هر پیچ، ۷ ویژگی مختلف شامل شاخص سطح (عددی در بازه (۱، -۱) که تعیین کننده میزان تقعر یا تحدب پیچ می‌باشد)، میزان خمیدگی، نمره حفاظت، پتانسیل الکترواستاتیک، آب‌گریزی، گرایش به واسط بودن باقی‌مانده و

ProMate: براساس بعضی از خصوصیات سطح از جمله توزیع اسیدهای آمینه به تفکیک نوع، توزیع اتم‌ها، توزیع جفت اسیدهای آمینه، نمره حفاظت، ساختار دوم، اندازه پیچ (patch size) و ... به پیش‌بینی می‌پردازد [۲۶].

cons-PPISP: با اجماعی از شبکه‌های عصبی و بر اساس ساختار فضایی به پیش‌بینی باقی‌مانده‌های واسط می‌پردازد. در این وب‌سرور از پروفایل توالی و مساحت در دسترس باقی‌مانده

مساحت در دسترس حلال محاسبه می شود. سپس، برای پیش بینی سایت های تعامل، بردار ویژگی فوق به عنوان ورودی به SVM داده می شود [۲۸].

WHISCY: در این وب سرور، به هر باقی مانده براساس نمره حفاظت و گرایش به واسط بودن، ضریبی اختصاص می یابد. سپس، براساس ضرایب باقی مانده موردنظر و همسایه های آن، واسط بودن آن مشخص می شود [۲۹].

SPPIDER: در این وب سرور برای کدینگ پنجره ورودی شبکه عصبی از میانگین وزن دار ویژگی ها شامل مساحت نسبی در دسترس حلال (RSA)، اختلاف بین مقادیر پیش بینی شده و مشاهده شده RSA، آب گریزی، میزان بقای بار، میزان بقای آب گریزی، میزان بقای اندازه، میزان بقای نوع اسید آمینه در پنجره ای با ۱۱ باقی مانده استفاده شده است [۳۰].

PIER: در این وب سرور از پارامترهای آماری (مساحت در دسترس و نمره حفاظت) به دست آمده از گروه های اتمی سطح پروتئین برای پیش بینی استفاده می شود. [۳۱].

meta-PSIVER: سایت های تعامل را تنها براساس ساختار اول پروتئین و با استفاده از طبقه بند بیز ساده پیش بینی می کند [۳۲].

meta-PIS: در این وب سرور، نتایج پنج الگوریتم مختلف شامل SPPIDER [۳۰]، meta-PSIVER [۳۲]، الگوریتم Sikic و همکاران [۱۶] و دو نسخه بهبود یافته از الگوریتم Liu و همکاران [۱۵] به شکلی مؤثر با یکدیگر ترکیب شده اند [۳۳]. meta-PPISP: که با ترکیب خطی نتایج وب سرورهای ProMate [۲۶]، cons-PPISP [۲۷] و PINUP [۳۴] به پیش بینی باقی مانده های واسط می پردازد [۳۵].

عملکرد وب سرورهای فوق برحسب معیارهای ارزیابی مختلف در جدول ۳ با یکدیگر مقایسه شده است. همچنین، علاوه بر نمونه های فوق، می توان به وب سرورهای IPPRED [۲۵]، PatchDock [۳۷]، HADDOCK [۳۶]، InterPreTS [۳۸]، ClusPro [۳۹]، GRAMM-X [۴۰]، FireDock [۴۱]، RosettaDock [۴۲]، Struct2Net [۴۳]، Hex [۴۴]، CPORT [۴۵]، pyDockWEB [۴۶] نیز اشاره نمود.

جدول ۳: مقایسه عملکرد وب سرورهای پیش بینی سایت های تعامل پروتئین-پروتئین (با استفاده از تنها اطلاعات ساختار مولکولی) براساس معیارهای پوشش، ضریب همبستگی، دقت، پیش بینی مثبت، ویژگی و حساسیت

وب سرورها	سال	پوشش	ضریب همبستگی	دقت	پیش بینی مثبت	ویژگی	حساسیت
ProMate [۲۶]	۲۰۰۴	-	-	٪۷۰	-	-	-
cons-PPISP [۲۷]	۲۰۰۵	٪۵۰	-	٪۷۱	-	-	-
PPI-PRED [۲۸]	۲۰۰۵	-	-	٪۷۶	-	٪۵۰	٪۲۰
PINUP [۳۴]	۲۰۰۶	٪۳۰/۵	-	٪۲۹/۴	-	-	-
WHISCY [۲۹]	۲۰۰۶	-	-	٪۶۰	-	-	٪۱۰
SPPIDER [۳۰]	۲۰۰۷	-	۰/۴۳	٪۷۴/۲	-	٪۶۳/۷	٪۶۰/۳
meta-PPISP [۳۵]	۲۰۰۷	٪۵۰/۵	-	٪۴۹/۵	-	-	-
PIER [۳۱]	۲۰۰۷	-	-	-	-	٪۶۰	٪۵۰
meta-PSIVER [۳۲]	۲۰۱۰	-	۰/۱۳۵	-	-	٪۲۵	٪۴۶/۵
meta-PIS [۳۳]	۲۰۱۳	-	۰/۱۸۱	٪۶۶/۳	٪۳۳	٪۷۰/۶	٪۴۹/۵

پیش بینی ساختار مؤثر آنتی ژن ها و آنتی بادی ها:

از آنجا که همه آنتی بادی ها و آنتی ژن ها زیرمجموعه پروتئین ها هستند؛ لذا برای پیش بینی سایت های تعامل آن ها می توان از هر یک از روش های عمومی پیش بینی سایت های تعامل پروتئین ها (که در بخش های قبلی مورد بررسی قرار گرفتند) استفاده نمود؛ اما به دلیل اهمیت عملکرد آنتی بادی ها و نحوه تعامل آن ها با آنتی ژن ها در سیستم ایمنی بدن انسان، برخی محققین برای

دستیابی به کارایی بیشتر، اقدام به توسعه روش های اختصاصی برای آنتی بادی ها و آنتی ژن ها نموده اند. بدیهی است که تعیین ساختارهای مؤثر آنتی ژن ها و آنتی بادی ها در طراحی داروهای مؤثرتر از اهمیت ویژه ای برخوردار می باشد. لازم به ذکر است که در همه روش های ارائه شده برای پیش بینی ساختار آنتی ژن ها و آنتی بادی ها، تنها از اطلاعات توالی اسیدهای آمینه استفاده شده

است؛ زیرا ساختار فضایی بسیاری از آنتی‌ژن‌ها و آنتی‌بادی‌های شناخته شده، هنوز مشخص نشده است.

• ساختار مؤثر آنتی‌ژن‌ها

Resch و همکاران [۴۷] به پیش‌بینی فنوتیپ‌های مختلف ویروس نقص ایمنی انسانی نوع ۱ (آنتی‌ژن) پرداختند. ایزوله‌های این ویروس دارای دو نوع فنوتیپ می‌باشند. فنوتیپ اول براساس توانایی ویروس برای القا و تکثیر چند هسته‌ای شدن سلول‌های MT-2 به دو گونه SI (Syncytium Inducing) و NSI (Non-SI) می‌باشد. این در حالی است که فنوتیپ دوم بر اساس اینکه ویروس برای داخل شدن به سلول‌های بدن از چه گیرنده کموکائینی استفاده می‌کند به دو گونه CXCR4-using (X4) و CCR5-using تقسیم‌بندی می‌شود. آن‌ها برای پیش‌بینی فنوتیپ‌ها از یک شبکه عصبی پیشرو با ۱۶ نرون در ورودی و ۳ نرون در لایه پنهان و یک نرون در لایه خروجی استفاده نمودند. آموزش شبکه با الگوریتم پس‌انتشار خطا انجام شده است (با نرخ یادگیری ۰/۵). اسیدآمینه‌های موجود در چند موقعیت خاص در حلقه سوم متغیر ویروس ایدز (V3) و بار کلی آن به عنوان ورودی‌های شبکه عصبی در نظر گرفته شده است. نتایج بیانگر توانایی شبکه برای پیش‌بینی فنوتیپ اول است؛ اما از دقت مناسبی در پیش‌بینی فنوتیپ دوم برخوردار نیست.

Chen و همکاران [۴۸] با استفاده از طبقه‌بند SVM و تجزیه

توالی آنتی‌ژن به ترکیبات دوتایی به پیش‌بینی اپی‌توپ‌ها پرداختند. برای محاسبه بردار ورودی، ابتدا توالی به ترکیبات دوتایی تجزیه شده و سپس، یک بردار ۴۰۰ عنصری براساس فراوانی ترکیب‌های مذکور ساخته می‌شود. در نهایت، این روش به ویژگی ۵۸/۹۴٪، حساسیت ۷۴/۳۱٪ و دقت ۶۹/۶۵٪ رسیده است.

علاوه بر مقالات ذکر شده، وب‌سرورهای متعددی نیز در ارتباط با تعیین ساختار آنتی‌ژن (به صورت خاص تعیین اپی‌توپ‌های آن) براساس توالی آن ارائه شده است. به عنوان مثال، می‌توان از وب‌سرورهای BcePred [۴۹] (پیش‌بینی براساس خصوصیات فیزیکی-شیمیایی اسیدآمینه مانند آبدوستی، انعطاف‌پذیری، میزان قطبی بودن، سطح در معرض و میزان دسترسی به حلال)، CEP [۵۰] (پیش‌بینی براساس میزان دسترس‌پذیری حلال و توزیع فضایی اسیدهای آمینه)، DiscoTope [۵۱] (پیش‌بینی براساس ساختار و نمرات گرایش اسیدهای آمینه)، ELLIPRO [۵۲] (پیش‌بینی براساس ساختار فضایی)، EPITOPIA [۵۳] (پیش‌بینی براساس خصوصیات فیزیکی-شیمیایی و ویژگی‌های هندسی ساختار فضایی)، EPCES [۵۴] و EPSVR [۵۵]، SEPPA [۵۶]، EPMeta [۵۵] و Bpredictor [۵۷] نام برد.

برای مقایسه روش‌های فوق، جدول ۴ را ببینید.

جدول ۴: مقایسه عملکرد وب‌سرورهای پیش‌بینی اپی‌توپ آنتی‌ژن‌ها براساس معیارهای پوشش، ضریب همبستگی، دقت، پیش‌بینی مثبت، ویژگی و حساسیت

الگوریتم پیش‌بینی	سال	دقت	پیش‌بینی مثبت	ویژگی	حساسیت
Bcepred [۴۹]	۲۰۰۴	۵۸/۷٪	-	۶۱٪	۵۶٪
CEP [۵۰]	۲۰۰۵	۷۵٪	-	-	-
DiscoTope [۵۱]	۲۰۰۸	-	-	۷۵٪	۴۷/۳٪
ElliPro [۵۲]	۲۰۰۸	۸۴٪	۲۹/۱٪	۸۶/۲٪	۶۰/۱٪
Epitopia [۵۳]	۲۰۰۹	۸۹/۴٪	-	-	-
EPCES [۵۴]	۲۰۰۹	-	-	۶۹/۵٪	۴۷/۸٪
SEPPA [۵۶]	۲۰۰۹	-	-	۷۰/۷٪	۵۸٪
EPSVR [۵۵]	۲۰۱۰	۵۹/۱٪	-	-	-
EPMeta [۵۵]	۲۰۱۰	-	-	-	-
Bpredictor [۵۷]	۲۰۱۱	۶۷/۲٪	-	۶۹/۲٪	۴۹٪

• ساختار مؤثر آنتی‌بادی‌ها

محمودی و همکاران [۵۸] و طاهرزاده و همکاران [۵۹] در دو تحقیق متفاوت، با استفاده از شبکه عصبی پرسپترون چندلایه و بهره‌گیری از دو ویژگی آبدوستی و PH ایزوالکتریک، به دسته بندی اتوانتی‌بادی‌ها (در بیماری لوپوس) پرداختند. در این کار، برای آموزش و تست شبکه عصبی از نواحی متغیر آنتی‌بادی‌های ضد DNA (مجموعه مثبت) و آنتی‌بادی‌های غیر اتوانتی‌بادی (مجموعه منفی) استفاده گردیده است. همچنین، ساختار شبکه عصبی، شامل ۲۶۶ نرون در لایه ورودی، ۹ نرون در هر یک از دو لایه پنهان و ۱ نرون در لایه خروجی می‌باشد که به روش آزمون و خطا تنظیم گردیده است. برای آموزش شبکه نیز از الگوریتم پس انتشار خطا استفاده شده است. این روش توانست به حساسیت و ویژگی ۷۱٪ دست یابد.

در پژوهش دیگری، عبدی و همکاران [۶۰] روشی جدید برای ارزیابی و تعیین آنتی‌بادی‌های مؤثر بر HIV با استفاده از SVM ارائه شد. در این روش، از ترکیب ۳ ویژگی مختلف شامل احتمال وقوع اسید آمینه در هم‌ترازی، همسایگی اسیدآمینه‌ها در توالی و آبدوستی برای کدگذاری زنجیره پروتئینی استفاده شد. در واقع، روش فوق با معرفی بردار جدید همسایگی، در ارزیابی آنتی‌بادی‌های مؤثر بر HIV به حساسیت ۸۱/۸۰٪ و ویژگی ۹۹/۶۶٪، ارزش پیش‌بینی مثبت ۹۸/۲۹٪ و دقت ۹۶/۲۰٪ دست یافته است.

در حقیقت، در دو تحقیق اخیر به جای آموزش طبقه‌بند با انواع مختلفی از ترکیب‌های آنتی‌بادی-آنتی‌ژن، تنها ساختار آنتی‌بادی‌های قابل تعامل با آنتی‌ژن مورد نظر (مثلاً HIV) و آنتی‌ژن‌هایی با ساختار مشابه، به طبقه‌بند آموزش داده شده است. به این ترتیب، با توجه به ساده شدن فضای مدل‌سازی، طبقه‌بند می‌تواند به دقت بسیار بهتری از خود نشان دهد.

در ارتباط با تعیین پاراتوپ‌ها نیز هم‌اکنون چندین وب‌سرور به صورت برخط در دسترس می‌باشد. به عنوان مثال، وب‌سرور

Paratome [۶۱] از هم‌ترازسازی ساختاری پروتئین‌ها و شباهت‌های ساختاری آن‌ها استفاده می‌نماید. به بیانی دیگر، این وب‌سرور با مقایسه توالی اسیدآمینه‌ای یا ساختار سوم آنتی‌بادی موردنظر با ساختار آنتی‌بادی‌هایی شناخته شده، به پیش‌بینی نواحی اتصال با آنتی‌ژن می‌پردازد. همچنین، وب‌سرور proABC [۶۲] برای تعیین نواحی اتصال، الگوریتم جنگل تصادفی را به خدمت گرفته است. این وب‌سرور قادر به پیش‌بینی احتمال اتصال هر اسیدآمینه بر اساس نوع تعامل (تعاملات با تشکیل پیوند هیدروژنی، تعاملات آب‌گریز، تعاملات بدون اتصال) و نوع اتم‌های آن (زنجیره اصلی، زنجیره جانبی و یا هردو) نیز می‌باشد. دو وب‌سرور مذکور از هم‌ترازسازی ساختاری یا هم‌ترازسازی توالی برای پیش‌بینی بهره می‌برند. این در حالی است که به دلیل جهش‌های غیرهمبسته در کمپلکس‌های آنتی‌بادی-آنتی‌ژن (و در نتیجه، عدم دسترسی به اطلاعات تکاملی) استفاده از هم‌ترازی چندان قابل قبول نیست. لذا، Krawczyk و همکاران وب‌سرور Antibody i-Patch را معرفی نمودند. در این وب‌سرور به جای هم‌ترازی، از ساختار همولوژی آنتی‌بادی‌ها برای تعیین احتمال اتصال اسیدآمینه‌ها به آنتی‌ژن استفاده شده است. به بیانی دقیق‌تر، این وب‌سرور با مدل‌سازی ساختار آنتی‌بادی-موردنظر به روش همولوژی و براساس نمره گرایش اسیدآمینه (براساس فراوانی اسیدآمینه در سایت تعامل نسبت به نواحی دیگر تعیین می‌شود)، به هر اسیدآمینه نواحی CDR، امتیاز واسط بودن اختصاص می‌دهد. Asti و همکاران [۶۴] نیز روشی مبتنی بر آنتروپی برای پیش‌بین میزان کشش بین آنتی‌بادی و آنتی‌ژن پیشنهاد کردند. پیش‌بینی موردنظر با استفاده از مدل‌سازی حداکثر آنتروپی بر روی مجموعه توالی به دست آمده از بیمار آلوده به ویروس HIV حاصل شد. نتایج به دست آمده از این روش بر روی ۳۰ آنتی‌بادی، بیانگر ضریب همبستگی ۰/۷۷ می‌باشد. برای مقایسه روش‌های فوق، جدول ۵ را ببینید.

جدول ۵: مقایسه عملکرد روش‌های پیش‌بینی ساختار آنتی‌بادی‌ها براساس معیارهای پوشش، ضریب همبستگی، دقت، پیش‌بینی مثبت، ویژگی و حساسیت

الگوریتم پیش‌بینی	سال	دقت	پیش‌بینی مثبت	ویژگی	حساسیت
اتوانتی‌بادی‌ها [۵۹]	۲۰۰۴	-	-	۷۱٪	۷۱٪
Paratome [۶۱]	۲۰۱۲	۴۱٪	۳۶٪	۱۱٪	۱۰۰٪
ProABC [۶۲]	۲۰۱۳	۸۱٪	۶۹٪	۸۲٪	۸۰٪
Antibody i-Patch [۶۳]	۲۰۱۳	-	۴۰٪	-	۹۴٪
عبدی و همکاران [۶۰]	۲۰۱۵	۹۶/۲٪	۹۸/۳٪	۹۹/۷٪	۸۱/۸٪

بحث و نتیجه‌گیری

همان‌طور که بیان گردید، درک مکانیسم تعامل میان پروتئین‌ها به منظور طراحی داروها، نیازمند شناسایی باقی‌مانده‌های در حال تعامل است. در این مقاله، با بررسی آخرین و پرکاربردترین منابع در این زمینه سعی گردید که بررسی جامعی از انواع روش‌های پیش‌بینی سایت‌های تعامل کمپلکس‌های آنتی‌بادی-پروتئین ارائه گردد.

به عنوان مثال، بسیاری از محققین برای دستیابی به دقت قابل قبول در پیش‌بینی سایت‌های تعامل، از اطلاعات ساختار فضایی پروتئین برای آموزش شبکه عصبی سود جسته‌اند. در مطالعه حاضر به بررسی ۱۴ منبع مختلف در این زمینه نشان داده شد که با استفاده از اطلاعات ساختار فضایی امکان بهبود دقت پیش‌بینی تا ۸۰٪ (روش Bordner و Abagyan [۹]) نیز میسر بوده است. همچنین، ماشین بردار پشتیبان پرکاربردترین طبقه‌بند برای پیش‌بینی سایت‌های تعامل با استفاده از اطلاعات ساختار فضایی بود. با وجود این، اطلاعات فضایی برای بسیاری از پروتئین‌ها در دسترس نمی‌باشد؛ لذا برای پیش‌بینی سایت‌های تعامل پروتئین-پروتئین تنها با استفاده از اطلاعات توالی اسیدهای آمینه، الگوریتم‌های هوشمند و وب‌سرورهای متعددی تاکنون معرفی شده‌اند. در این تحقیق، ۲۸ منبع مختلف در این رابطه مورد بررسی قرار گرفت.

به عنوان مثال، در جدول ۲، شش روش مختلف پیش‌بینی سایت‌های تعامل با استفاده از ساختار مولکولی مورد بررسی قرار گرفته که بهترین دقت توسط Bock و Gough [۲۰] ارائه شده است. همچنین، در جدول ۳، عملکرد ۱۰ وب‌سرور پرکاربرد در پیش‌بینی سایت‌های تعامل پروتئین-پروتئین (تنها با استفاده از توالی اسیدهای آمینه)، براساس معیارهای ارزیابی مختلف شامل پوشش، ضریب همبستگی، پیش‌بینی مثبت و منفی، ویژگی و حساسیت با یکدیگر مقایسه شده است. همان‌طور که مشاهده شد، پوشش، دقت، ویژگی و حساسیت روش‌های مذکور بدون استفاده از اطلاعات ساختار فضایی به ترتیب ۵۰/۵٪ (وب سرور meta-PPISP [۳۵])، ۷۶٪ (وب سرور PPI-PRED [۲۸])، ۶۳/۷٪ (وب سرور SPPIDER [۳۰]) و ۶۰/۳٪ (وب سرور SPPIDER [۳۰]) تجاوز نمی‌کند.

با توجه به دقت پایین روش‌های پیش‌بینی سایت‌های تعامل، برخی محققین اقدام به توسعه الگوریتم‌هایی هوشمند برای پیش‌بینی ساختارهای مؤثر آنتی‌ژن‌ها و آنتی‌بادی‌ها نموده‌اند. در این گونه روش‌ها، به جای آموزش کمپلکس‌های پروتئین-پروتئین، تنها آنتی‌ژن‌ها یا آنتی‌بادی‌های مرتبط به شبکه عصبی آموزش داده می‌شوند. در همین رابطه، مطالعه حاضر به بررسی

۱۱ منبع مختلف در خصوص پیش‌بینی ساختار آنتی‌ژن‌ها و ۷ منبع مختلف در زمینه پیش‌بینی ساختار آنتی‌بادی‌ها پرداخت. به عنوان مثال، در جدول ۴ عملکرد ۱۰ روش پیش‌بینی ساختار آنتی‌ژن‌ها و همچنین، در جدول ۵، کارایی ۵ روش پیش‌بینی ساختار آنتی‌بادی‌ها، بر حسب معیارهای ارزیابی مختلف شامل دقت، پیش‌بینی مثبت، ویژگی، و حساسیت با یکدیگر مقایسه شد. همان‌طور که مشاهده می‌شود، بهترین دقت، ویژگی و حساسیت به دست آمده برای پیش‌بینی اپی‌توپ آنتی‌ژن‌ها به ترتیب عبارت است ۸۹/۴٪ (Epitopia [۵۳])، ۸۶/۲٪ (ElliPro [۵۲])، و ۶۰/۱٪ (ElliPro [۵۲]). این در حالی است که روش عبدی و همکاران [۶۰] با دقت ۹۶/۲٪، پیش‌بینی مثبت ۹۸/۳٪، ویژگی ۹۹/۷٪ و حساسیت ۸۱/۸٪ بهترین کارایی را در میان همه روش‌های پیش‌بینی ساختار آنتی‌بادی‌ها داشته است.

از آنجا که تاکنون تحقیق مستقلی در رابطه با پیش‌بینی ساختار آنتی‌بادی‌های مونوکلونال انجام نشده است، نتایج اخیر می‌تواند در این زمینه بسیار امیدبخش باشد. به عبارت دقیق‌تر، عبدی و همکاران [۶۰] نشان دادند که به جای انواع مختلفی از کمپلکس‌های آنتی‌بادی-آنتی‌ژن، می‌توان تنها از ساختار آنتی‌بادی‌های قابل تعامل با آنتی‌ژن مورد نظر (یا چند آنتی‌ژن با ساختار مشابه) برای آموزش شبکه عصبی (یا طبقه‌بند) استفاده نمود. به این ترتیب، با افزایش شباهت الگوهای آموزشی و ساده شدن فضای ویژگی، دقت شبکه عصبی تا ۹۶/۲٪ افزایش یافته است. تنها دشواری در این زمینه، تعداد محدود آنتی‌بادی‌های شناخته‌شده مؤثر بر گیرنده‌های سلول سرطانی است؛ لذا در این حالت، طبقه‌بند SVM به دلیل عدم حساسیت به تفاوت تعداد نمونه‌های آموزشی در کلاس‌های مثبت منفی و قابلیت آموزش با تعداد نمونه‌های کم، عملکرد بهتری در مقایسه با شبکه‌های عصبی مصنوعی داشته است.

با توجه به رشد روزافزون سرطان و عوارض درمان‌های رایج، توجه‌ها به سمت درمان‌های اختصاصی و هدفمند جلب شده است. این نوع درمان بر پایه استفاده از آنتی‌بادی‌های مونوکلونال می‌باشد. اولین گام در تولید این نوع آنتی‌بادی‌ها، پیش‌بینی ساختار آن‌ها است. پیش‌بینی ساختار باید با توجه به تعامل آنتی‌ژن مربوطه با آنتی‌بادی‌های تأثیرگذار بر آن انجام شود؛ لذا در این مقاله روش‌های موجود برای پیش‌بینی سایت‌های تعامل پروتئین-پروتئین در حالت عام و آنتی‌بادی-آنتی‌ژن در حالت خاص بررسی شد. در این مطالعه نشاد داده شد که تنها روش‌های توسعه داده شده برای پیش‌بینی ساختار آنتی‌بادی‌های مشابه

با یک آنتی ژن خاص زیاد نمی باشد، استفاده از طبقه بند SVM برای این منظور پیشنهاد گردید.

(مثلاً مؤثر بر یک یا چند آنتی ژن مشابه) از دقت کافی برای پیش بینی ساختار آنتی بادی های مونوکلونال برخوردار هستند. از آنجا که معمولاً تعداد آنتی بادی های شناخته شده با قابلیت تعامل

References

1. Oldham RK, Dillman RO. Monoclonal antibodies in cancer therapy: 25 years of progress. *J Clin Oncol* 2008;26(11):1774-7.
2. Abbas AK, Lichtman AH. Cellular and Molecular Immunology. 9th ed. Canada: Elsevier/Saunders; 2017.
3. Scott AM, Wolchok JD, Old LJ. Antibody therapy of cancer. *Nat Rev Cancer* 2012;12(4):278-87.
4. Ofra Y, Rost B. Predicted protein-protein interaction sites from local sequence information. *FEBS Lett* 2003;544(1-3):236-9.
5. Keskin O, Tuncbag N, Gursoy A. Characterization and prediction of protein interfaces to infer protein-protein interaction networks. *Curr Pharm Biotechnol* 2008;9(2):67-76.
6. Zhou HX, Shan Y. Prediction of protein interaction sites from sequence profile and residue neighbor list. *Proteins* 2001;44(3):336-43.
7. Fariselli P, Pazos F, Valencia A, Casadio R. Prediction of protein-protein interaction sites in heterocomplexes with neural networks. *Eur J Biochem* 2002;269(5):1356-61.
8. Koike A, Takagi T. Prediction of protein-protein interaction sites using support vector machines. *Protein Eng Des Sel* 2004;17(2):165-73.
9. Bordner AJ, Abagyan R. Statistical analysis and prediction of protein-protein interfaces. *Proteins* 2005;60(3):353-66.
10. Wang B, Chen P, Huang DS, Li JJ, Lok TM, Lyu MR. Predicting protein interaction sites from residue spatial sequence profile and evolution rate. *FEBS Lett* 2006;580(2):380-4.
11. Chung JL, Wang W, Bourne PE. Exploiting sequence and structure homologs to identify protein-protein binding sites. *Proteins* 2006;62(3):630-40.
12. Nguyen MN, Rajapakse JC. Protein-Protein Interface Residue Prediction with SVM Using Evolutionary Profiles and Accessible Surface Areas. Symposium on Computational Intelligence and Bioinformatics and Computational Biology; 2006 Sep 28-29; Toronto, Ont., Canada: IEEE; 2007.
13. Dong Q, Wang X, Lin L, Guan Y. Exploiting residue-level and profile-level interface propensities for usage in binding sites prediction of proteins. *BMC Bioinformatics* 2007;8:147.
14. Li N, Sun Z, Jiang F. Prediction of protein-protein binding site by using core interface residue and support vector machine. *BMC Bioinformatics* 2008;9:553.
15. Liu B, Wang X, Lin L, Tang B, Dong Q, Wang X. Prediction of protein binding sites in protein structures using hidden Markov support vector machine. *BMC Bioinformatics* 2009; 10:381.
16. Sikic M, Tomic S, Vlahovicek K. Prediction of protein-protein interaction sites in sequences and 3D structures by random forests. *PLoS Comput Biol* 2009;5(1):e1000278.
17. Chen Y, Xu J, Yang B, Zhao Y, He W. A novel method for prediction of protein interaction sites based on integrated RBF neural networks. *Comput Biol Med* 2012;42(4):402-7.
18. Li BQ, Feng KY, Chen L, Huang T, Cai YD. Prediction of protein-protein interaction sites by random forest algorithm with mRMR and IFS. *PLoS One* 2012;7(8):e43927.
19. Hwang H, Vreven T, Weng Z. Binding interface prediction by combining protein-protein docking results. *Proteins* 2014; 82(1): 57-66.
20. Bock JR, Gough DA. Predicting protein-protein interactions from primary structure. *Bioinformatics* 2001;17(5):455-60.
21. Yan C, Honavar V, Dobbs D. Identification of interface residues in protease-inhibitor and antigen-antibody complexes: a support vector machine approach. *Neural Comput Appl* 2004;13(2):123-9.
22. Res I, Mihalek I, Lichtarge O. An evolution based classifier for prediction of protein interfaces without using protein structures. *Bioinformatics* 2005;21(10):2496-501.
23. Ofra Y, Rost B. ISIS: interaction sites identified from sequence. *Bioinformatics* 2007;23(2):e13-6.
24. Chen XW, Jeong JC. Sequence-based prediction of protein interaction sites with an integrative method. *Bioinformatics* 2009;25(5):585-91.
25. Goffard N, Garcia V, Iragne F, Groppi A, de Daruvar A. IPPRED: server for proteins interactions inference. *Bioinformatics* 2003;19(7):903-4.
26. Neuvirth H, Raz R, Schreiber G. ProMate: a structure based prediction program to identify the location of protein-protein binding sites. *J Mol Biol* 2004;338(1):181-99.
27. Chen H, Zhou HX. Prediction of interface residues in protein-protein complexes by a consensus neural network method: test against NMR data. *Proteins* 2005;61(1):21-35.
28. Bradford JR, Westhead DR. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics* 2005;21(8):1487-94.
29. de Vries SJ, van Dijk AD, Bonvin AM. WHISCY: what information does surface conservation yield? Application to data-driven docking. *Proteins* 2006;63(3):479-89.

30. Porollo A, Meller J. Prediction-based fingerprints of protein-protein interactions. *Proteins* 2007;66(3):630-45.
31. Kufareva I, Budagyan L, Raush E, Totrov M, Abagyan R. PIER: protein interface recognition for structural proteomics. *Proteins* 2007;67(2):400-17.
32. Murakami Y, Mizuguchi K. Applying the Naive Bayes classifier with kernel density estimation to the prediction of protein-protein interaction sites. *Bioinformatics* 2010;26(15):1841-8.
33. Huang J, Deng R, Wang J, Wu H, Xiong Y, Wang X. metaPIS: a sequence-based meta-server for protein interaction site prediction. *Protein Pept Lett* 2013;20(2):218-30.
34. Liang S, Zhang C, Liu S, Zhou Y. Protein binding site prediction using an empirical scoring function. *Nucleic Acids Res* 2006;34(13):3698-707.
35. Qin S, Zhou HX. meta-PPISP: a meta web server for protein-protein interaction site prediction. *Bioinformatics* 2007;23(24):3386-7.
36. Aloy P, Russell RB. InterPreTS: protein interaction prediction through tertiary structure. *Bioinformatics* 2003;19(1):161-2.
37. Dominguez C, Boelens R, Bonvin AM. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 2003;125(7):1731-7.
38. Schneidman-Duhovny D, Inbar Y, Polak V, Shatsky M, Halperin I, Benyamini H, et al. Taking geometry to its edge: fast unbound rigid (and hinge-bent) docking. *Proteins* 2003;52(1):107-12.
39. Comeau SR, Gatchell DW, Vajda S, Camacho CJ. ClusPro: an automated docking and discrimination method for the prediction of protein complexes. *Bioinformatics* 2004;20(1):45-50.
40. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res* 2006;34(Web Server issue):W310-4.
41. Mashiach E, Schneidman-Duhovny D, Andrusier N, Nussinov R, Wolfson HJ. FireDock: a web server for fast interaction refinement in molecular docking. *Nucleic Acids Res* 2008;36(Web Server issue):W229-32.
42. Lyskov S, Gray JJ. The RosettaDock server for local protein-protein docking. *Nucleic Acids Res* 2008;36(Web Server issue):W233-8.
43. Singh R, Park D, Xu J, Hosur R, Berger B. Struct2Net: a web service to predict protein-protein interactions using a structure-based approach. *Nucleic Acids Res* 2010;38(Web Server issue):W508-15.
44. Macindoe G, Mavridis L, Venkatraman V, Devignes MD, Ritchie DW. HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res* 2010;38(Web Server issue):W445-9.
45. de Vries SJ, Bonvin AM. CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PloS one* 2011;6(3):e17695.
46. Jimenez-Garcia B, Pons C, Fernandez-Recio J. pyDockWEB: a web server for rigid-body protein-protein docking using electrostatics and desolvation scoring. *Bioinformatics* 2013;29(13):1698-9.
47. Resch W, Hoffman N, Swanstrom R. Improved success of phenotype prediction of the human immunodeficiency virus type 1 from envelope variable loop 3 sequence using neural networks. *Virology* 2001;288(1):51-62.
48. Chen J, Liu H, Yang J, Chou KC. Prediction of linear B-cell epitopes using amino acid pair antigenicity scale. *Amino Acids* 2007;33(3):423-8.
49. Saha S, Raghava GP. Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. *Proteins* 2006;65(1):40-8.
50. Kulkarni-Kale U, Bhosle S, Kolaskar AS. CEP: a conformational epitope prediction server. *Nucleic Acids Res* 2005;33(Web Server issue):W168-71.
51. Haste Andersen P, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci* 2006;15(11):2558-67.
52. Ponomarenko J, Bui HH, Li W, Fusseder N, Bourne PE, Sette A, et al. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinformatics* 2008;9:514.
53. Rubinstein ND, Mayrose I, Martz E, Pupko T. Epitopia: a web-server for predicting B-cell epitopes. *BMC Bioinformatics* 2009; 10: 287.
54. Liang S, Zheng D, Zhang C, Zacharias M. Prediction of antigenic epitopes on protein surfaces by consensus scoring. *BMC Bioinformatics* 2009;10:302.
55. Liang S, Zheng D, Standley DM, Yao B, Zacharias M, Zhang C. EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results. *BMC Bioinformatics* 2010;11:381.
56. Sun J, Wu D, Xu T, Wang X, Xu X, Tao L, et al. SEPPA: a computational server for spatial epitope prediction of protein antigens. *Nucleic Acids Res* 2009;37(Web Server issue):W612-6.
57. Zhang W, Xiong Y, Zhao M, Zou H, Ye X, Liu J. Prediction of conformational B-cell epitopes from 3D structures by random forests with a distance-based feature. *BMC Bioinformatics* 2011;12:341.
58. Mahmoudi M, Khademi M, Mirsalehi M, Taherzadeh-Sani M, Azemi A, Naghibi MB. Determination of effective structures of antibodies in binding to DNA using artificial neural networks. *Iranian Journal of Basic Medical Sciences* 2004; 6(4):315-22. Persian
59. Taherzadeh M, Mirsalehi M, Khademi M, Mahmoudi M, Naghibi MB, Azemi A. Usage of the genetic algorithm and multi-layer perceptron to predict effective structures for antibody binding, *Amirkabir* 2004; 33(1). Persian
60. Abdi M, Saadatmand-Tarzan M, Taherzadeh M, Haghparsat A. The new algorithm for for determining antibodies effective to HIV using support vector machine. 1st National E-Conference of Technology Development on Electrical, Electronics and Computer Engineering; 2015 Feb 15; Mashhad: Ferdowsi University of Mashhad; 2015

61. Kunik V, Ashkenazi S, Ofran Y. Paratome: an online tool for systematic identification of antigen-binding regions in antibodies based on sequence or structure. *Nucleic Acids Res* 2012;40(Web Server issue):W521-4.
62. Olimpieri P, Chailan A, Tramontano A, Marcatili P. Prediction of site-specific interactions in antibody-antigen complexes: the proABC method and server. *Bioinformatics* 2013; 29(18): 2285–91.
63. Krawczyk K, Baker T, Shi J, Deane CM. Antibody i-Patch prediction of the antibody binding site improves rigid local antibody-antigen docking. *Protein Eng Des Sel* 2013;26(10):621-9.
64. Asti L, Uguzzoni G, Marcatili P, Pagnani A. Maximum-Entropy Models of Sequenced Immune Repertoires Predict Antigen-Antibody Affinity. *PLoS Comput Biol* 2016;12(4):e1004870.

A Review of Prediction Methods of Interaction Sites of Antibody-Protein Complexes Based on Artificial Intelligence

Abdi Marzieh¹, Saadatmand-Tarzjan Mahdi^{2*}, Taherzadeh-Sani Mohammad³,
Haghparast Alireza⁴

• Received: 21 Sep, 2017

• Accepted: 1 Mar, 2018

Introduction: Cancer is one of the most important health issues in the current and next centuries. Understanding the mechanism of interaction between antibody-protein residues is essential for designing targeted anticancer drugs based on monoclonal antibodies. Prediction of the effective structure is the first step for production of monoclonal antibodies.

Methods: This paper is a systematic review of the state-of-the-art researches on prediction of interaction sites and specification of antibody structures. Artificial neural networks or web servers are frequently used for evaluation of interaction sites while some researchers have employed evolutionary algorithms for prediction of the effective structure of antibodies. Accordingly, 14 methods based on the protein spatial structure, 28 researches based on the molecular amino-acide sequence (without usage of the spatial structure), and 18 antigen/antibody structure prediction techniques were reviewed.

Results: We demonstrated that the accuracy of structure-based methods can be increased up to 80% while the accuracy of sequence-based methods was rarely better than 75%. Since the spatial structure of many antibodies is unknown, some researchers raised the accuracy (even to 96%) by only antibody sequences able to interact with some similar antigens in training neural networks. Therefore, we suggest this approach for structure prediction of monoclonal antibodies because of its adequate high accuracy.

Conclusion: In this paper, after reviewing available methods for prediction of antibody-protein interaction sites, some suggestions were made for effective prediction of structure of monoclonal antibodies.

Keywords: Immunology, Monoclonal Antibodies, Antibody-Protein Complexes, Artificial Intelligence, Neural Networks

• **Citation:** Abdi M, Saadatmand Tarzjan M, Taherzadeh Sani M, Haghparast AR. A Review of Prediction Methods of Interaction Sites of Antibody-Protein Complexes Based on Artificial Intelligence. *Journal of Health and Biomedical Informatics* 2018; 5(1): 56-69.

1. M.Sc in Bioelectronc, Medical Imaging Lab, Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.

2. Ph.D in Bioelectronc, Assistant Professor, Medical Imaging Lab, Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

3. Ph.D in Electronic Engineering, Assistant Professor, Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

4. Ph.D in Pathobiology, Associate Professor, Department of Veterinary Medicine, Ferdowsi University of Mashhad, Mashhad, Iran

***Correspondence:** Department of Electrical Engineering, Faculty of Engineering, Ferdowsi University of Mashhad, Vakil-Abad Blv., 9177948974 Mashhad, Iran

• **Tel:** +98 51 3880 5137

• **Email:** saadatmand@um.ac.ir