

بررسی پایگاه‌های توالی آلرژن‌ها و تعیین پایگاه داده مناسب برای آنالیز بیوانفورماتیکی آلرژنیسیته پروتئین‌های نو ترکیب

نجف الهیاری فرد^{۱*}، زرین مینوچهر^۲، مسعود رهگذر^۳، عبدالمنان اورنگ^۴، محمد فضیلتی^۵

تاریخ دریافت: ۹۳/۱۰/۱۷

تاریخ تصویب: ۹۵/۰۸/۱۲

چکیده

بر اساس اعلام سازمان جهانی آلرژی؛ حدود یک سوم جمعیت جهان متاثر از آلرژی‌اند. مطابق با ضوابط سازمان جهانی سلامت ارزیابی آلرژنیسیته پروتئین‌های نو ترکیب ضروری است. ارزیابی آلرژنیسیته با پیشرفت‌های بدست آمده در آنالیزهای بیوانفورماتیکی و تعیین ساختار سه بعدی آلرژن‌ها تسهیل شده است. آنالیز بیوانفورماتیکی مرحله مقدماتی در ارزیابی آلرژنیسیته پروتئین‌های نو ترکیب است که نتایج آن در سنگینی مستندات و قضاوت آلرژنیسیته اهمیت دارد. کارایی آنالیز بیوانفورماتیکی آلرژنیسیته پروتئین‌های نو ترکیب

۱* استادیار، پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری. (نویسنده مسئول (allahyar@nigeb.ac.ir)

۲ استادیار، پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری

۳ دانشیار، گروه کامپیوتر دانشگاه تهران

۴ کارشناسی ارشد، دانشگاه پیام نور اصفهان

۵ دانشیار، دانشگاه پیام نور اصفهان

وابسته به ماهیت و محتوای پایگاه‌های اطلاعاتی مورد استفاده در تحلیل است. این تحقیق با هدف بررسی و تعیین پایگاه‌های داده توالی مناسب برای آنالیز بیوانفورماتیکی آلرژنیسته پروتئین‌های نو ترکیب انجام شد. لذا پایگاه‌های داده آلرژن پس از مطالعه تطبیقی در سه گروه الف- عمومی، ب- مولکولی و ج- آنالیز موتیف و اپی‌توپ دسته‌بندی شد. نتایج نشان داد این پایگاه‌ها در حجم داده‌ها، سازماندهی و قابلیت‌ها متفاوت هستند. به منظور دستیابی به نتایج بهتر و دقیق‌تر، آنالیز بیوانفورماتیکی آلرژنیسته پروتئین‌های نو ترکیب باید بصورت ترکیبی و بهره‌مندی از گروهی از پایگاه‌های اطلاعاتی آلرژن‌ها انجام شود. بدلیل گستره داده‌ها و الگوریتم‌های مختلف درباره پیش‌بینی آلرژنیسته، پایگاه توالی آلرژن‌ها شامل Allergome, ADFS, SDAP و FARRP و Alegend برای آنالیز بیوانفورماتیکی آلرژنیسته پروتئین‌های نو ترکیب توصیه می‌شود. همچنین در این تحقیق ارزیابی بیوانفورماتیکی آلرژنیسته پروتئین اسموتین انجام شد. نتایج تحقیق حاکی از آن بود که براساس آنالیزهای بیوانفورماتیکی اسموتین حساسیت‌زا است.

واژه‌های کلیدی: آلرژن، پایگاه داده پروتئینی، پروتئین‌های نو ترکیب

مقدمه

صنعتی شدن و افزایش شهرنشینی اپیدمی آن افزایش یافته است (Ono, 2000). آلرژی می تواند هزینه زیادی را به جامعه تحمیل کند و تبعات جدی اقتصادی اجتماعی داشته باشد (Kamble & Bharmal, 2009). آلرژن پروتئین یا گلیکوپروتئینی است که توسط ایمونوگلوبولین E (IgE) سیستم ایمنی افراد دارای حساسیت شناسایی می شود. افزایش آلرژی های مختلف در یک فرد، وضعیت آلرژی را در جهان دشوار تر کرده و از این رو یک راهبرد مناسب برای تشخیص و درمان آلرژی لازم و ضروری است (Bru-sic et al., 2003). با توجه به رشد سریع اطلاعات آلرژی، روش های بیوانفورماتیکی برای مدیریت اطلاعات آلرژن و تجزیه و تحلیل دقیق از توالی ها، ساختارها، خواص عملکردی و آلرژی زایی شان در حال ایجاد و توسعه می باشد (Brusic & Petrovsky, 2003). آلرژن ها به دسته های مختلفی شامل گیاهی، گیاهی خوراکی، جانوری، جانوری خوراکی، قارچی، مایت ها، و سایر موارد تقسیم می شوند (Cui, 2013). هنگامی که یک سیستم ایمنی افراطی در معرض آلرژن قرار می گیرد، ابتدا ایمونوگلوبولین IgE تولید می کند. IgE ها به سطح ماست سل های موجود در دستگاه تنفسی و دستگاه گوارشی (محل اصلی ورود آلرژن ها)، متصل می شوند. در صورتی که فرد بار دیگر در معرض همان

تاریخچه تحقیق و تولید پروتئین های نو ترکیب حاصل از تحقیقات مهندسی ژنتیک به حدود سه دهه قبل می رسد (Catherine H Schein, 1989). تولید اریتروپویتین، هورمون رشد (rHGH)، انسولین انسانی، هورمون FSH، فاکتور هشت، TPA، اینترفرون و سوماتوتروپین از مثال های بارز تولید پروتئین های نو ترکیب هستند. رها سازی پروتئین های نو ترکیب که مصارف مختلفی در پزشکی، صنعت و کشاورزی دارند و ورود آنها به چرخه مصرف و زندگی مردم، منوط به ارزیابی های ایمنی زیستی و اطمینان از سلامتی و بی خطر بودن آنها است (Salehi et al., 2010). یکی از زمینه های مهم علمی در انجام ارزیابی های ایمنی زیستی، بررسی آلرژنیسیته یا امکان بروز واکنش های حساسیت توسط پروتئین های نو ترکیب است که در بررسی مذکور از آنالیزهای بیوانفورماتیک نیز استفاده می شود (Ladics et al., 2011); Towhid-far & Kaviani, (2010). واژه آلرژی برای نخستین بار در اوایل ۱۹۹۰ با تعریف کوتاه شدن تنفس مطرح شد (Cohen & King, 2001). سازمان جهانی آلرژی اعلام کرده است در حال حاضر حدود ۳۰ درصد جمعیت جهان از انواع آلرژی رنج می-برند. شیوع آلرژی افزایشی و با

مصوبات سازمان‌های بین‌المللی سلامت^۱ و غذا (فائو)^۲ بر اساس یک درخت تصمیم‌گیری است (شکل ۱) Randhawa et al., (2011). آنالیزهای بیوانفورماتیکی بخشی از این فرآیند و ابزار موثری برای تعیین آلرژنیسته پروتئین‌های نو ترکیب است.

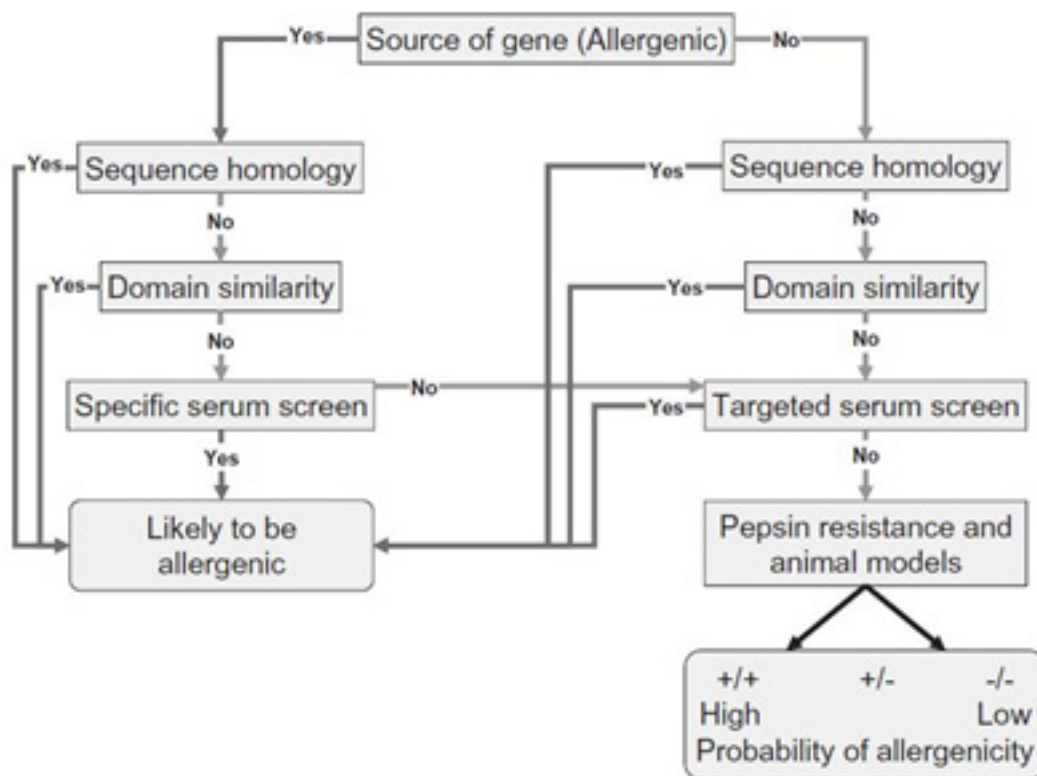
توصیف خواص مولکولی شامل مقایسه توالی و آنالیز ساختار سه بعدی (۳D) و شناخت ساختاری می‌تواند به فراهم کردن درک و بینش کلی نسبت به اساس مولکولی آلرژنی‌زایی و تبیین عملکرد بیولوژیکی از آلرژن کمک کند (Gendel, 2009). بیوانفورماتیک در حال دگرگون کردن زمینه‌های مطالعه آلرژنی از طریق فراهم کردن پایگاه داده برای مدیریت اطلاعات آلرژن، الگوریتم برای توصیف خواص واکنش‌های متقابل آلرژیک، موتیف‌های ساختاری و اپی توپ‌های سلول‌های B و C، ابزارهای پیش بینی آلرژنی‌زایی و تکنیک‌هایی برای آنالیزهای ژنومیک و پروتئومیک آلرژن‌ها می‌باشد (Bru-sic & Petrovsky, 2003). آنالیزهای بیوانفورماتیکی یک کنه کوچک آلرژن به تازگی شناسایی شده به نام Blo t 1 از *Blomia tropicalis* ۳۵% هویت سیستمین پروتئاز کنه را نشان داد (Mora et al., 2003). مدل‌سازی ساختار سه بعدی

آلرژن قرار گیرد، آلرژن‌ها به IgE های موجود بر سطح ماست سل‌ها می‌چسبند. تشکیل کمپلکس آلرژن / IgE در سطح ماست سل‌ها منجر به آزادی هیستامین و اینترلوکین می‌شود که عامل بروز علائم و نشانه‌های آلرژنی است (Alvaro et al., 2013). عوامل مختلفی در ایجاد آلرژنی‌ها دخالت دارند که عبارتند از عوامل مرتبط با میزبان و عوامل محیطی. عوامل مرتبط با میزبان شامل ژنتیک، سن، جنس و نژاد است و عوامل محیطی شامل شرایط محیطی و وضعیت بهداشتی محل زندگی است. همچنین اثرات متقابل عوامل محیطی و ژنتیکی بر یکدیگر نیز در بروز آلرژنی دخالت دارد (Holloway et al., 2010).

بررسی آلرژنیسته پروتئین‌های نو ترکیب و نقش بیوانفورماتیک
یکی از زمینه‌های مهم بررسی ایمنی محصولات تراریخته، ارزیابی آلرژنیسته پروتئین‌های نو ترکیب است. ارزیابی آلرژنیسته پروتئین‌های نو ترکیب دارای مراحل مختلفی است که عبارتند از: الف- بررسی منبع پروتئین، ب- بررسی همولوژی توالی اسیدهای آمینه، ج- بررسی هضم پروتئین‌های مورد مطالعه و مقاومت به پپسین، د- غربالگری سرم، ه- بررسیهای تکمیلی از طریق مدل‌های حیوانی. ارزیابی آلرژنی‌زایی پروتئین‌های نو ترکیب مطابق

1 World Health Organization (WHO)

2 Food and Agriculture Organization (FAO)



شکل ۱: درخت تصمیم‌گیری (نسخه ویرایش‌شده) گام به گام برای ارزیابی آلرژنیسیته پروتئین‌های نو ترکیب (Randhawa et al., 2011)

توضیح: ارزیابی همسانی توالی یک گام کلیدی در فرایند تصمیم‌گیری است. زمانی که شباهت پروتئین موردنظر با آلرژن‌های هدف مطرح است سرم بیماران مبتلا به آلرژی به آلرژن‌های هدف جهت پیدا کردن اینکه آیا پروتئین مورد نظر در بیماران حساس ایجاد واکنش می‌کند، ارزیابی می‌شود (سمت چپ درخت). اگر منفی بود، سرم گرفته شده از بیماران حساس در آلرژی به غذاهای وابسته به آلرژن‌های هدف جهت پیدا کردن اینکه آیا پروتئین مورد نظر افراد مستعد را حساس می‌کند، ارزیابی می‌شود (سمت راست درخت). مراحل آخر درخت تصمیم‌گیری مقاومت به تخریب پروتئولیتیک و آلرژی زایی در مدل‌های حیوانی (+/+): هر دو تست منفی؛ احتمال بالا، (-/+): یکی از دو تست مثبت و دیگری منفی؛ احتمال متوسط و (-/-): هر دو تست منفی؛ احتمال پایینی از آلرژی‌زایی را نمایش می‌دهد.

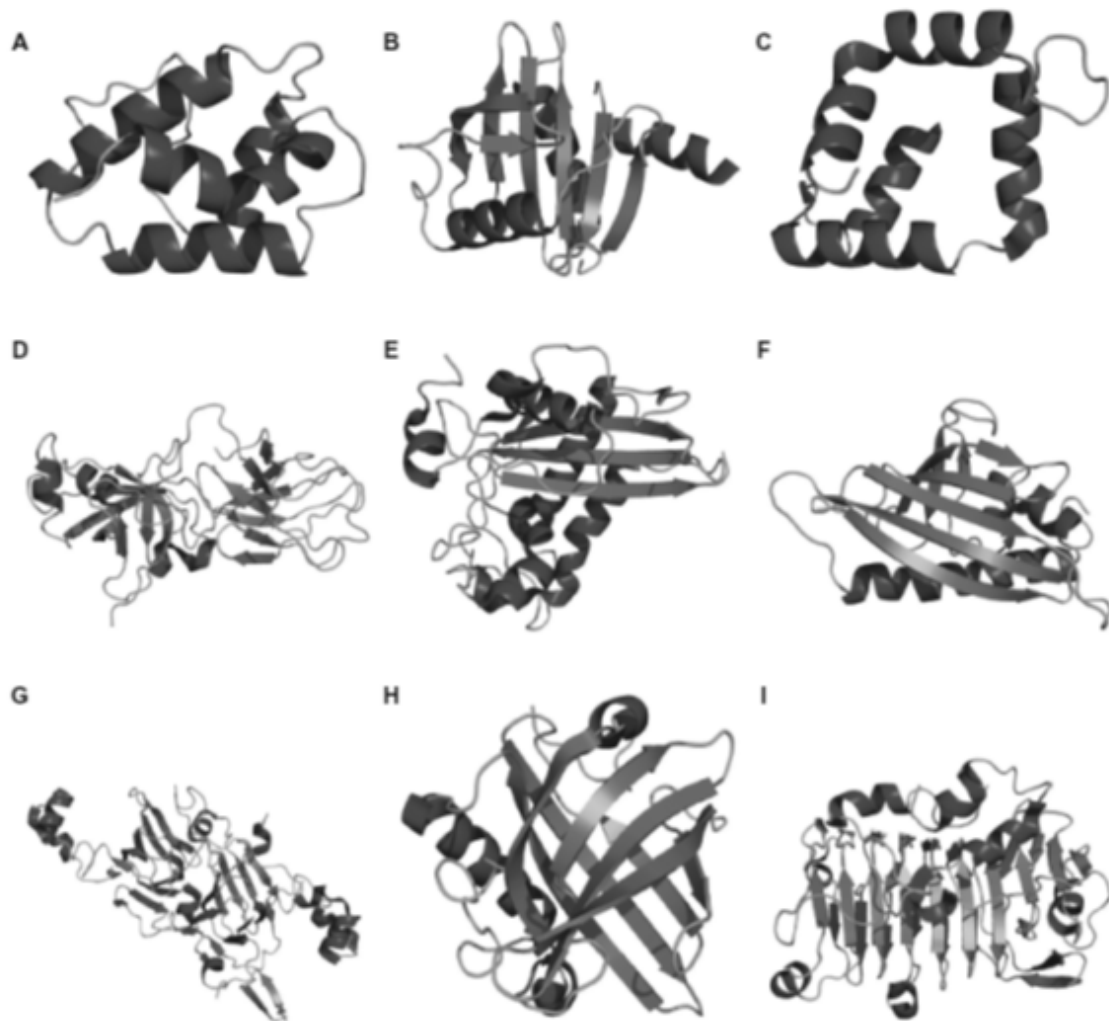
خانواده از پروتئین‌های وابسته به پاتوژن (PR-10) است و تا همین اواخر عملکرد بیولوژیکی Bet v 1 ناشناخته بود. (شکل ۲) Markovic-Housley et al., (2003). Bet v 1 دارای ۵۲ ایزوآلرژن است Ivanciuc et al., (2003)

Blo t 1 ساختار دوم و سوم حفظ شده سیستمین پروتئاز را نمایش داد. این آنالیز ساده بیوانفورماتیکی، همچنین، برای تعیین عملکرد و ساختار Blo t 1 نیز مفید بود (Brusic & Petrovsky, 2003). آلرژن اصلی درخت‌غان Bet v 1 متعلق به یک



شکل ۲: آلرژن اصلی درخت گان ۱ Bet v متعلق به خانواده پروتئین‌های وابسته به پاتوژن
(PR-10) Ivanciuc et al., (2003)

امروزه از تکنیک‌های دیگری مانند NMR و بررسی ساختار با اشعه X در ارزیابی ساختار آلرژن‌ها و کمک به بیوانفورماتیک جهت دستیابی به نتایج دقیق‌تر در پیش‌گویی آلرژنیسته استفاده می‌شود. شکل ۳ نشان دهنده ۹ آلرژن که بوسیله NMR تعیین شده و دارای بیشترین فراوانی در پایگاه داده Pfam (خانواده پروتئینی) هستند، است (C. H. Schein et al., (2010). الگوریتم‌های معمول مقایسه توالی‌ها در پایگاه‌های داده آلرژن‌ها FASTA Pear- (Altschul et al., (2000) و BLAST (1997) هستند. روش‌های محاسباتی برای شناسایی اپی‌توپ‌های سلول‌های T و B به درک بهتر واکنش‌های کامل آلرژیک کمک می‌کند. روش‌های اولیه بیوانفورماتیک برای شناسایی اپی‌توپ‌های بالقوه سلول B از توالی و خواص آمینواسیدی (مانند آب دوستی) و پایین بودن دقت استفاده می‌کنند. روش‌های خیلی جدیدتر از مدل‌سازی ساختار سه بعدی علاوه بر استراتژی‌های اولیه استفاده می‌کنند (Valenta & Kraft, (2001). با رشد انفجاری اطلاعات مربوط به آلرژی از طریق ژنومیکس عملکردی (Saito et al., (2001)، میکرواری (Harwanegg et al., (2003) و پروتئومیکس (Yu et al., (2002)، بیوانفورماتیک به یک ابزار ضروری و کاربردی برای مدیریت داده‌های آلرژی، تفسیر نتایج و طرح‌های تجربی تبدیل شده است. ابزارهای مقایسه توالی در تعیین آلرژی‌زایی و واکنش‌های متقابل آلرژیک معمولاً ضروری‌اند (Gendel, (2002). آنالیز توالی موتیف به عنوان یک روش مهم برای تعیین واکنش‌های متقابل آلرژیک در آلرژن‌های با تشابه کم در حال



شکل ۳: ساختار ۹ NMR آلرژن که دارای بیشترین فراوانی در پایگاه داده پروتئینی هستند. (A) بازدارنده پروتئاز/خانواده LTP؛ آلرژن موجود در ذرت (Zea m ۱۴). (B) خانواده پروفیلین؛ گرده درخت غان واجد پروفیلین (Bet v ۲). (C) آلرژن موجود در تیموتی گراس (Phl p ۷). (D) اکسپانسیو؛ بتا اکسپانسیو موجود در ذرت (Zea m ۱). (E) خانواده پروتئینی CRISP؛ آلرژن موجود در بزاغ مورچه آتشی (Sol i ۳). (F) خانواده پروتئینی PR۱۰؛ آلرژن گیلاس (Pru av ۱). (G) خانواده کوپینز؛ آلرژن بادام زمینی (Ara h ۳). (H) لیپوکالین؛ آلفا-۲-گلوبولین موش (Mus m ۱). (I) خانواده پکتات لیاز؛ آلرژن گرده درخت سدر (C. H. Schein et al., ۲۰۱۰). (Jun a ۱).

ظهور است. بیوانفورماتیک ساختاری نقش مهمی در تبیین اساس ساختاری آلرژی-زایی آلرژن‌های اختصاصی، شناسایی جایگاه‌های شناسایی IgE و مهندسی پروتئین‌های ضد حساسیت دارد (Mar-kovic-Housley et al., 2003). در مجموع، ابزارهای بیوانفورماتیک در حال حاضر برای تحقیقات آلرژی و طراحی واکسن

و درمان بیماری‌های آلرژیک بسیار مهم است (Singh & Bhalla, 2003).
 با توجه به توضیحات فوق پایگاه توالی آلرژن‌ها دارای خصوصیات و کاربردهای مهمی هستند که عبارتند از: مدیریت داده‌های آلرژنی، آنالیز آلرژن‌ها، ادغام منابع داده آلرژنی، ابزارهای مقایسه توالی، تعیین آلرژنی‌زایی و واکنش‌های متقابل آلرژیک، تبیین اساس ساختاری آلرژنی‌زایی آلرژن‌های اختصاصی، شناسایی جایگاه‌های شناسایی IgE (اپی‌توپ) و مهندسی پروتئین‌های ضد حساسیت، تحقیقات آلرژنی و طراحی واکسن و درمان بیماری‌های آلرژیک. با توجه به گستردگی کاربردها این تحقیق با هدف بررسی، دسته‌بندی و تعیین پایگاه‌های داده توالی مناسب برای آنالیز بیوانفورماتیکی آلرژنیسته پروتئین‌های نو ترکیب انجام شده است.

روش کار

الف- مطالعه تطبیقی پایگاه داده آلرژن‌ها
 در این تحقیق که با هدف بررسی، دسته‌بندی و تعیین پایگاه‌های داده توالی مناسب برای آنالیز بیوانفورماتیکی آلرژنیسته پروتئین‌های نو ترکیب انجام شده است، در مرحله اول پایگاه‌های اطلاعاتی آلرژن‌ها با جستجوی پیشرفته در بانک‌های اطلاعاتی اولیه شامل NCBI،

PubMed, PDB, Google Scholar بدست آمد. در مرحله دوم منابع اصلی اطلاعات آلرژن‌ها و نیز پایگاه‌های داده اختصاصی تعیین شدند. مرحله سوم به مطالعه تفصیلی و نیز اجزای پایگاه اطلاعاتی آلرژن‌ها شامل منبع داده‌های پایگاه (Data source)، توالی پروتئینی آلرژن‌ها (-Allergen protein sequence)، ساختار پروتئینی آلرژن‌ها (Allergen protein structure)، دومین‌ها (Domains)، اپی توپ‌ها (-Epi topes)، ابزارهای ارزیابی ایمنی (Safety evaluation tools)، ابزارهای جستجوی شباهت (Similarity search tools)، داده‌های آزمایشگاهی (Experimental data)، تعداد داده (No. Of data)، طبقه بندی داده‌ها (Classification of data)، اطلاعات مولکولی آلرژن (Molecular information of allergens) و الگوریتم‌ها برای جستجو تشابه (Algorithms for similarity search) اختصاص یافت، در مرحله چهارم دسته‌بندی پایگاه داده آلرژن‌ها به سه دسته کلی؛ الف- عمومی، ب- مولکولی و ج- آنالیز موتیف انجام شد و در مرحله انتهایی پایگاه‌های داده توالی مناسب برای آنالیز بیوانفورماتیکی آلرژنیسته پروتئین‌های نو ترکیب با لحاظ خصوصیات مختلفی شامل گستره و تنوع داده‌ها، ارتباط با پایگاه‌های توالی ژنی و پروتئینی اولیه شامل: PDB, NCBI, SWISS-PROT, PIR، ارزیابی بر مبنای استانداردهای FAO



شکل ۴: ساختار سه بعدی پروتئین اسموتین (ژن AP24) با کد ۱PCV از پایگاه داده PDB

شود. اسموتین یک پروتئین چندکاره پاسخ دهنده به استرس است که موجب تحمل گیاهان نسبت به فشارهای اسمزی می شود. اسموتین از طریق تحریک تجمع پرولین و افزایش پایداری باعث مقاومت گیاه در برابر استرس های زیستی و غیر زیستی می شود. این پروتئین عضوی از خانواده پروتئینی PRP بوده و در حفاظت گیاهان در برابر عوامل بیماری زا نقش ایفا می کند. در این بخش بررسی آلرژن زایی پروتئین اسموتین (ژن AP24) بعنوان نمونه جهت استفاده از پایگاه های داده آلرژن انجام شد.

توالی این پروتئین با فرمت Fasta و با کد یونی پرات (UniProt ID: P14170)

WHO، تحلیل جایگاه و پیش بینی اپی توپ در آلرژن، تعیین موتیف های آلرژن با جستجو در بانک های داده ای مرتبط مانند MEME/MAST، انجام شد.

ب- بررسی آلرژن زایی پروتئین اسموتین (ژن AP24)

پس از تعیین پایگاه های داده آلرژن مناسب، تحقیق بوسیله ارزیابی بیوانفورماتیکی آلرژنیسیته یک نمونه پروتئین انجام شد. در این راستا ارزیابی با پروتئین اسموتین انجام گردید. اسموتین اولین بار در سلول های گیاه تنباکو (Nicoti-ana tabacum) یافت شد و قادر است با پتانسیل اسمزی پایین محیط سازگار

استخراج گردید. توالی اسمو تین دارای توالی پروتئین اصلی شامل رزیدوهای ۲۲ تا ۲۴۶ اسید آمینه است که اسید آمینه‌های ۲۴۶ تا ۲۶۶ یعنی ۲۲۵ اسید آمینه است. ساختار ۱ تا ۲۱ سیگنال پپتید را شامل می‌شود. اسمو تین با روش X-ray تعیین ساختار

```
>sp|P14170|OSMO_TOBAC Osmotin OS=Nicotiana tabacum GN=AP24 PE=1 SV=2
MGNLRSSFVFFLLALVITYTYAATIEVRNNCPYTVWAASTPIGGGRRLDRGQTWVINAPRG
TKMARVWGR TNCNFNAAGRGTCTGDCGGVLQCTGWGKPPNTLAEYALDQFSGLDLFDWIS
LVDGFNIPMTFAPTNPSGGKCHAIHCTANINGECPREL RVPGGC NNPCTTFGGQQYCCTQ
GPCGPTFFSKFFKQRCPDAYSYPQDDPTSTFTCPGGSTNYRVIFCPNGQAHPNFLEMPG
SDEVAK
```

شکل ۵: توالی پروتئین اسمو تین (ژن AP۲۴) با کد (UniProt ID: P۱۴۱۷۰)، توالی سیگنال پپتید زیرخطدار است

شده و توالی ۲۲ تا ۲۲۶ (۲۰۵ اسید آمینه) در پایگاه داده PDB با کد 1PCV دارای دو زنجیره A, B است.

یافته‌ها
الف- نتایج مطالعه تطبیقی پایگاه داده آلرژن‌ها
 پایگاه‌های اطلاعاتی معرف آلرژن‌ها به سه دسته کلی؛ الف- عمومی، ب- مولکولی و ج- آنالیز موتیف و اپی توپ دسته بندی شد. جدول ۱ شامل پایگاه‌های اطلاعاتی معرف آلرژن‌ها بصورت عمومی است و شامل- Allergome, AllAllergy, Infor- و mAll, AllergoPharma, Protall می‌باشند. در این گروه آلرژن‌ها بصورت عمومی و البته علمی شامل منبع آلرژن، نامگذاری آلرژن، نوع و گروه آلرژن و موارد مرتبط

آنالیزهای بیوانفورماتیکی مطابق دستورالعمل استاندارد مربوطه جهت بررسی آلرژنیسته توالی اسمو تین در چند پایگاه داده آلرژن‌ها شامل- FAR, RP, SDAP, Allergome, Alegpred, ADFS در سه سطح توالی کامل، توالی ۸۰ اسید آمینه‌ای و توالی ۸ اسید آمینه‌ای انجام شد. پارامترهای مورد استفاده شامل ضریب کلمه برابر با ۲ و ماتریکس بلوسوم ۵۰ بودند و نرم-افزارهای مورد استفاده در این پایگاه‌های داده عبارت بودند از Fasta 3.5 و Blastp. Algorithm: FASTA (3.5 Sept 2006) [op-

جدول ۱: پایگاه‌های داده علمی-عمومی آلرژن‌ها

پایگاه داده آلرژن	URL
Allergome	/http://www.allergome.org
AllAllergy	/http://allallergy.net
InformAll	/http://www.foodallergens.info
AllergoPharma	http://www.allergopharma.com/dokumente/en/allergopharma/ allergen_db/allergen_db.php
Protall	/http://www.ifrn.bbsrc.ac.uk/protall

دیگر تبیین شده‌اند.

ارزیابی ایمنی آلرژنیسیته پروتئین‌های نو ترکیب، ارتباط با پایگاه توالی اولیه ژن‌ها و پروتئین‌ها و سایر موارد مولکولی قرار گرفته است (Radauer et al., 2008). داده‌های این پایگاه داده‌ها بر اطلاعات پایگاه‌داده‌های اولیه ژنی و پروتئینی و نیز منابع مقالات اولیه علمی منتشر شده در مجلات معتبر علمی بین‌المللی توسط

جدول ۲ شامل پایگاه‌های اطلاعاتی معرف آلرژن‌ها بصورت مولکولی است و شامل BFS, FARRP, CSL, SDAP, AllerPredict, AllerMatch, IUIS, Allergome, PSD, AllFam می‌باشند. در این گروه اطلاعات مولکولی و ساختاری آلرژن‌ها،

جدول ۲: پایگاه‌های داده توالی - مولکولی آلرژن‌ها

پایگاه داده آلرژن	URL
Bioinformatics for Food Safety	http://iit.edu/lsgendel/fa.htm
(Allergen Online (FARRP	http://allergenonline.com
Central Science Laboratory	/http://www.csl.gov.uk/allergen
Structural Database of Allergen Proteins	http://fermi.utmb.edu/SDAP/sdap_ver.html
AllerPredict	/http://sdmc.i2r.a-star.edu.sg/Templar/DB/Allergen
AllerMatch	/http://www.allermatch.org
International Immunogenetics Information System	/http://imgt.cines.fr
Allergome	/http://www.allergome.org
(PSD (Protein Structure Discovery	http://www-bionet.sccc.ru/psd/cgi-bin/programs/ Allergen/allergen.cgi
(AllFam (Allergen Families	http://www.meduniwien.ac.at/allergens/allfam
ADFS	/http://allergen.nihs.go.jp/ADFS

مراکز علمی مرتبط مبتنی بوده و به طور مداوم در حال به روز رسانی شدن هستند (Ivanciuc et al., 2002). جدول ۳ شامل پایگاه‌های اطلاعاتی آنالیز موتیف و اپی‌توپ آلرژن‌ها است و شامل Evaluator, WebAllergen, AllerTool, SDAP, AlgPred, IEDB, می‌باشند. در این گروه تحلیل جایگاه و پیش‌بینی اپی‌توپ در آلرژن و تعیین موتیف‌های مرتبط با جستجو در بانک‌های داده‌ای مرتبط مانند MEME/MAST گنجانده شده است (Bailey et al., 2009). آنالیز موتیف و اپی‌توپ آلرژن‌ها، به منظور درک بیشتر ماهیت آلرژنی‌زایی آنها و پیش‌گیری از آلرژنی‌زا بودن پروتئین‌های نو ترکیبی که در مواد غذایی و نیز داروهای

جدول ۳: پایگاه‌های داده آنالیز موتیف و اپی‌توپ آلرژن‌ها

پایگاه داده آلرژن	URL
Evaller	http://bioinformatics.bmc.uu.se/evaller.html
WebAllergen	http://weballergen.bii.a-star.edu.sg/
AllerTool	http://research.i2r.a-star.edu.sg/AllerTool
AlgPred	http://www.imtech.res.in/raghava/algpred
IEDB	http://immuneepitope.org
SDAP	http://fermi.utmb.edu/SDAP/sdap_ver.html

نو ترکیب استفاده می‌شوند، دارای اهمیت روزافزون و قابل توجه است (Ivanciuc et al., 2009); Randhawa et al., (2011). اگرچه الگوریتم‌های معمول مقایسه توالی‌ها در پایگاه‌های داده آلرژن مشابه اکثر پایگاه توالی داده‌های پروتئینی شامل FASTA و BLAST هستند، لیکن بعضاً الگوریتم SSEARCH نیز در برخی پایگاه داده‌ها مانند Allergome به عنوان الگوریتم تکمیلی استفاده می‌شود (Mari et al., 2006). برنامه پژوهش و منابع آلرژنی غذایی (FARRP) Hileman et al., (2002) و پایگاه داده ساختاری پروتئین‌های آلرژن (SDAP) پایگاه‌های داده با قابلیت جستجویی هستند که توالی‌های آلرژن و آنالیز مقایسه‌ای توالی‌های فعال را در برمی‌گیرد (Ivanciuc et al., 2003). پس از بررسی خصوصیات مختلف پایگاه داده آلرژن‌ها شامل گستره و تنوع داده‌ها، ارتباط با پایگاه‌های توالی ژنی و پروتئینی اولیه شامل: PDB, NCBI, SWISS-PROT, PIR، ارزیابی بر مبنای استانداردهای WHO, FAO، تحلیل جایگاه و پیش‌بینی اپی‌توپ در آلرژن، تعیین موتیف‌های آلرژن با جستجو در بانک‌های داده‌ای مرتبط مانند MEME/

جدول ۴: پایگاه‌های داده آلرژن‌ها مناسب برای تحلیل آلرژنیسیته پروتئین‌های نو ترکیب

پایگاه داده آلرژن	URL
Allergome	/http://www.allergome.org
FARRP	http://allergenonline.com
SDAP	http://fermi.utmb.edu/SDAP/sdap_ver.html
ADFS	/http://allergen.nihs.go.jp/ADFS
AlgPred	http://www.imtech.res.in/raghava/algpred

داده PDB) می‌باشد. در این پایگاه داده دسترسی به یک فهرست مروری کامل از آلرژن‌ها و قابل جستجو بر اساس توالی جهت تشخیص پروتئین-هایی که ممکن است دارای یک خطر بالقوه واکنش متقاطع آلرژنی باشند، فراهم شده است. این پایگاه داده برای کمک به ارزیابی امنیت و حساسیت‌زایی پروتئین‌های نو ترکیب موجود در مواد غذایی حاصل از مهندسی ژنتیک طراحی شده است Goodman et al., (2007); Wise et al., (2009).

۲- پایگاه توالی آلرژن و تحلیلی آلرژنیسیته SDAP یکی از کاربردی‌ترین و معروف‌ترین پایگاه‌های اطلاعاتی مولکولی تحلیل آلرژنیسیته است که در آن اطلاعاتی راجع به وضعیت مولکولی و ساختاری، بویژه ساختار سه بعدی آلرژن‌ها در دسترس قرار گرفته است. همچنین امکان مقایسه و تحلیل توالی‌های پروتئین‌های نو ترکیب در آن وجود دارد. تعداد داده‌های

MAST، پایگاه داده‌های Allergome، ALEpred و ADFS، SDAP، FARRP جامع‌تر، دقیق‌تر و مناسب‌تر برای دریافت اطلاعات مربوطه و نیز تحلیل آلرژنیسیته پروتئین‌های نو ترکیب تشخیص داده شد که به اختصار ذیلا تبیین می‌شوند.

۱- پایگاه تحلیلی آلرژنیسیته FARRP این پایگاه داده با قابلیت جستجوی توالی‌های پروتئین، به منظور ارزیابی ایمنی آلرژنیسیته پروتئین‌های نو ترکیب کاربرد فراوانی دارد. نسخه ۱۶ این پایگاه داده در تاریخ ۲۷ ژانویه ۲۰۱۶ به روز شده و شامل یک لیست جامع ۱۹۵۶ توالی آلرژن است. فرمت مورد استفاده پروتئینی -FAS-TA است. فرآیند به‌روز رسانی بصورت سالانه و شامل جمع‌آوری توالی جدید تعیین شده به عنوان آلرژنی در فایل‌های مرجع از بانک اطلاعاتی NCBI بخش پروتئین (وارد شده از RefSeq، GeneBank و پایگاه‌های داده TPA و همچنین توالی پروتئین از SwissProt، PIR، PRF و پایگاه

IgE اپی‌توپ، و ارتباط با پایگاه داده پروتئینی (UniProt, PDB, InterPro, Pfam) ، طبقه بندی تفصیلی و اطلاعات کربوهیدرات آنها است. گردآوری داده‌ها در ADFS هم شامل استفاده از داده‌های آلرژن‌ها و اپی‌توپ‌ها از پایگاه داده‌های SDAP و FARRP بوده و هم مبتنی بر جستجوی علمی در پایگاه توالی‌های اولیه بوده است Nakamura et al. (2005). تحلیل الرژنیسته شامل ارزیابی بر اساس معیارهای WHO و FAO و نیز جستجو بر اساس موتیف (Stadler, 2003) & می‌باشد.

۴- پایگاه توالی آلرژن و تحلیلی Aller-gome

این پایگاه آلرژن توسط Mari و همکاران، دارای توالی‌های مربوط برای عرضه اطلاعات در مولکول‌های آلرژن طراحی شده است و ضمن در بردارندگی اطلاعات علمی و عمومی، دارای ابزارهای تحلیل توالی درخواستی و بررسی همولوژی آن است. مولکول‌های شناسایی شده که منجر بیماری (آلرژیک، اتوپیک) با واسطه IgE (آنافیلاکسی، آسم، درماتیت اتوپیک، ورم ملتحمه، رینیت، کهیر) از مجلات علمی بین المللی و از منابع مبتنی بر وب انتخاب شده است. مولکول‌ها و منابع آلرژنی را در پایگاه داده Allergome بر اساس منابع مستقل (جانوران و گیاهان) و بافت آنها

موجود در این بانک داده که در گروه‌های مختلفی و نیز بصورت الفبایی دسته بندی شده اند بیش از ۱۵۰۰ مورد است. SDAP ، اطلاعات مربوط به پایگاه داده و ابزارهای مختلف محاسباتی را برای مطالعه آلرژن‌ها فراهم می‌کند. اجزاء پایگاه داده SDAP حاوی اطلاعات نام آلرژن، منبع، توالی، ساختار، اپی‌توپ‌های IgE ، و مقالات منبع و ارجاع به سرویس‌های اصلی پروتئین (PDB , PIR , SWISS-PROT , NCBI , MED-) و مقالات (PubMed , LINE) می‌باشند. SDAP اطلاعات اپی‌توپ‌های شناخته شده IgE و سلولهای B و T را خلاصه و ثبت کرده است که مبتنی بر اطلاعات مقالات است Ivanciuc et al., (2003); Catherine H Schein et al., (2006).

۳- پایگاه توالی آلرژن و تحلیلی آلرژنیسته ADFS

این پایگاه اطلاعاتی توسط محققان غذایی و شیمی‌ایمنی موسسه ملی علوم سلامت کشور ژاپن تهیه شده است. در این پایگاه اطلاعاتی نیز اطلاعات و داده‌های تخصصی راجع به وضعیت مولکولی آلرژن‌ها در دسترس قرار گرفته است. همچنین امکان مقایسه و تحلیل توالی‌های پروتئینی در آن وجود دارد. ADFS شامل اطلاعات متنوعی از آلرژن‌ها شامل نام، منبع، توالی، دومین‌ها،

موقعیت اپی-توپ در پروتئین مورد نظر، تعیین موتیف‌های آلرژن با جستجو در MEME/MAST، پیش‌بینی آلرژن‌ها براساس الگوی ماشین بردار پشتیبان؛ SVM، تسهیل و تسریع در BLAST با جستجو در پپتیدهای آلرژن ARPs معرفی شده قبلی پایگاه داده داخلی سرور، استفاده از رویکرد تلفیقی (SVMc + IgE) (MAST) + BLAST + ARPs epitope در پیش‌بینی آلرژنیسیته.

(ریشه، میوه، دانه گرده، دانه، هاگ، زهر، کل بدن و غیره)، و یا از نظر مسیرهای مختلف در معرض قرار گرفتن (تماس، بلع، استنشاق، تزریق و غیره) تقسیم‌بندی شده‌اند. بسیاری از مولکول‌های آلرژن‌زا به طور رسمی توسط اتحادیه بین‌المللی انجمن‌های ایمونولوژیک (IUIS) زیر کمیته نام-گذاری آلرژن، نام‌گذاری شده و در پایگاه داده Allergome مشخص شده‌اند (Mari et al., 2006).

ب- نتایج آلرژن‌زایی پروتئین اسموتین (ژن AP24)

بر اساس نتایج آنالیزهای بیوانفورماتیکی (جدول ۵) و بررسی توالی اسموتین در پایگاه‌ها داده آلرژن‌های مناسب‌تر برای تحلیل آلرژنیسیته شامل FARRP, SDAP, Allergome, Alegpred, ADFS چنین استنباط می‌شود که اسموتین پروتئینی حساسیت‌زا است و باعث بروز آلرژنیسیته در افراد می‌شود.

۵- پایگاه توالی آلرژن و تحلیلی Alegpred در این پایگاه داده آلرژن ارزیابی بر مبنای استانداردهای WHO, FAO، تحلیل جایگاه اپی-توپ در آلرژن، تعیین و آنالیز موتیف‌های آلرژن با جستجو در MEME/MAST ارائه شده است (Saha & Raha-va, 2006). قابلیت‌های مختلف ارائه شده در Alegpred عبارتند از: پیش‌بینی آلرژنیسیته بر اساس مشابهت اپی-توپ‌های شناخته شده در هر ناحیه پروتئینی، مکان‌یابی

جدول ۵: نتایج آنالیزهای بیوانفورماتیکی بررسی آلرژنیسیته اسموتین در چند پایگاه داده آلرژن. * موارد آلرژن متقاطع ثبت شده در پایگاه داده آلرژن بصورت نمونه است و به یک مورد از موارد متعدد اکتفا شده است.

پایگاه داده آلرژن	سطح بررسی			آلرژن متقاطع ثبت شده در پایگاه داده *	
	توالی کامل	توالی ۸۰ مر	توالی ۸ مر	لینک توالی	نام آلرژن
FARRP	Yes	Yes	Yes	AAR21072,1	3,2 Jun r
SDAP	Yes	Yes	Yes	P12670	Lyc e NP24
Allergome	Yes	-	-	AAG34078	1 Cap a
ADFS	Yes	Yes	Yes	P81370	2 Act c
AlgPred	Yes	-	-	Allergen	Allergen

بحث

ارزیابی آلرژنیسته پروتئین‌های نو ترکیب یکی از زمینه‌های مهم بررسی ایمنی محصولات تراریخته است. این اقدام با استناد به درخت تصمیم‌گیری و با استفاده از ابزارهای بیوانفورماتیکی و پایگاه‌های اطلاعاتی آلرژن‌ها انجام می‌شود (Aalberse, 2005). امروزه آنالیزهای بیوانفورماتیکی به عنوان ابزار موثری جهت ارزیابی ایمنی پروتئین‌های نو ترکیب حاصل از تغییرات ژنتیکی، از جمله بررسی آلرژنی‌زا بودن آنها، کاربرد وسیعی یافته‌اند (Gibson, 2006). این کاربرد با توسعه قابلیت‌های بیوانفورماتیک بویژه در زمینه شناسایی، پیش‌بینی ساختار سوم پروتئین‌ها و ایجاد و توسعه دانش مهندسی آنتی‌بادی‌ها رو به گسترش است (Vincent & Zurini, 2012).

Tong و همکاران به منظور طراحی پایگاه داده Allergen Atlas در سال ۲۰۰۹ میلادی با مطالعه داده‌های آلرژن‌ها در پایگاه‌های داده‌های مختلف آلرژن‌ها دریافتند سری‌های داده‌ای در این پایگاه‌های داده متفاوت هستند. همچنین ابزارهای تحلیل آلرژنیسته در برخی از آنها ناکارآمد می‌باشند (Tong et al., 2009). این دیدگاه در نتایج و یافته‌های این تحقیق نیز مورد تاکید قرار گرفته است. Gendel در سال ۲۰۰۸ میلادی با بررسی پایگاه

داده آلرژن‌ها و مفاهیم آلرژنی‌زایی، ضمن شناسایی ۱۵ پایگاه داده اصلی، آنها را به سه گروه عمومی، مولکولی و تحلیل موتیف تقسیم‌بندی کرد (Gendel, 2009). باتوجه به رشد تعداد داده‌ها و نیز اهمیت روزافزون بهره مندی از بیوانفورماتیک در تحلیل آلرژنی‌زایی پروتئین‌های نو ترکیب، یافته‌های این تحقیق مکمل نتایج قبلی است. Grover و همکاران در سال ۲۰۱۱ میلادی برای طراحی یک پایگاه داده دومین آلرژن‌ها از داده‌های FARRP استفاده کردند. Brusic و همکاران در سال ۲۰۰۳ میلادی با بررسی ۹ پایگاه داده اصلی آلرژن‌ها مشخص کردند داده‌ها طی سال‌های اخیر روندی افزایشی دارد همچنین در پایگاه‌های داده آلرژنی تکرار و افزونگی داده‌ای وجود دارد (Catherine Brusic et al., 2003). و همکاران در سال ۲۰۱۰ با بررسی ۱۰ پایگاه توالی اصلی آلرژن‌ها، نظام طبقه‌بندی و خانواده‌های پروتئینی و انجام واکنش متقاطع آلرژن‌ها در پایگاه داده‌های مزبور را بررسی کردند (C. H. Schein et al., 2010). برخی نتایج این تحقیق موید بررسی قبلی اخیرالذکر نیز می‌باشد. Gendel و همکاران در سال ۲۰۰۶ میلادی با بررسی ۱۰ پایگاه داده آلرژن شامل عمومی و مولکولی، تعداد توالی‌های آلرژن‌های اختصاصی را در آنها بررسی کردند و دریافتند این

مشترکات، دارای تنوع هم در تعداد و هم در نوع آلرژن‌ها می‌باشند (Allahyari Fard & Minucheher, 2013). لذا استفاده تکمیلی از آنها در تحلیل آلرژنیسیته پروتئین‌های نو ترکیب لازم است. ضمن آنکه لازم است محققان با توجه به فرضیات خود از گروه خاصی از آنها استفاده نمایند که در جداول فوق ذکر شد.

نتیجه گیری

طراحی، ساختار و نگهداری پایگاه‌های داده بیوانفورماتیک فرآیندی دشوار است (Gendel & Jenkins, 2006). از سوی دیگر دستیابی به نتایج دقیق‌تر، مستلزم بهره‌برداری تخصصی از پایگاه‌های داده بیوانفورماتیک و از جمله پایگاه توالی آلرژن‌ها می‌باشد. کارآیی، درجه تاثیر و سودمندی آنالیزهای بیوانفورماتیکی آلرژنیسیته پروتئین‌های نو ترکیب ارتباط مستقیمی با ماهیت و محتوای پایگاه‌های توالی آلرژن‌ها دارد که در تحلیل از آن استفاده می‌شود. اخیرا تعداد پایگاه‌های داده توالی آلرژن‌ها توسعه یافته‌اند و هر کدام با توجه به فرضیات و اهداف تحقیق، پاسخگویی ویژه‌ای دارند. پایگاه‌های توالی آلرژن‌ها در مقصود از ایجاد ساختار کلی، حجم داده‌ها، سازماندهی و قابلیت‌ها تفاوت دارند. تفاوت‌های مزبور، بویژه اختلافات محتوایی این پایگاه‌های اطلاعاتی نیاز به تحلیل‌های تطبیقی به منظور دستیابی به گستره و حوزه اختلاف

توالی‌های اختصاصی در این ۱۰ پایگاه داده آلرژن دارای تفاوت معنی‌دار هستند (Mari Gendel & Jenkins, 2006). همکاران که خود مبدع پایگاه توالی AI- Iergome هستند در سال ۲۰۰۷ میلادی پس از بررسی ۱۴ پایگاه توالی آلرژن را شناسایی و خصوصیات و مشخصات این پایگاه داده‌ها را بررسی کردند (Mari et al., 2006). و همکاران در سال ۲۰۰۹ میلادی با بررسی تفصیلی ۶ پایگاه داده توالی آلرژن‌ها شامل Allergome, FARRP, SDAP, ADFS, BIFS, IUIS آنها را بر اساس ۱۵ شاخص مورد تحلیل تطبیقی قرار دادند و تفاوت‌های آنها را مشخص کردند (Mari et al., 2009). این تفاوت‌ها در بررسی حاضر نیز مورد تایید قرار گرفت.

پایگاه‌های داده مختلف آلرژن‌ها با محتوی ارائه اطلاعات مختلف (عمومی، مولکولی و موتیف‌ها) و نیز تحلیل ساختارهای پروتئینی آلرژن‌ها با الگوریتم‌های مختلف در دانشگاه‌ها و مراکز علمی مختلف جهان طراحی و راه اندازی شده‌اند (Mari et al., 2006). در مجموع بانک‌های داده آلرژن‌ها از حیث تعداد و تنوع داده متفاوت‌اند و هیچکدام به تنهایی کامل نیستند و نیازهای محققان را بطور کامل برآورده نمی‌سازند، ضمنا داده‌های موجود همسانی ندارند. این پایگاه‌های اطلاعاتی ضمن دارا بودن

Allergome, ADFS, SDAP, FARRP و نیز ضرورت یک پایگاه اطلاعاتی و جامع با اشتراک‌گذاری همه داده‌ها را ضروری می‌نماید. اما با وضعیت موجود پایگاه‌های توالی آلرژن‌ها، لازم است محققان جهت بررسی آلرژنیسته پروتئین‌های نوترکیب و برای یافتن پاسخ آزمون فرضیات خود از ترکیبی از پایگاه‌های توالی آلرژن‌ها به نحوی که پوشش داده‌ای و تنوع کارآیی آنها کامل باشد، استفاده نمایند. علاوه بر این بدلیل گستره داده‌ها، ارزیابی بر مبنای استانداردهای FAO, WHO، تحلیل جایگاه اپی‌توپ در آلرژن، تعیین موتیف‌های آلرژن با جستجو در MEME/MAST؛ پنج پایگاه توالی آلرژن‌ها شامل

آنالیزهای بیوانفورماتیکی آلرژنیسته پروتئین‌های نوترکیب توصیه می‌شود. از اینرو در این تحقیق، بصورت نمونه ارزیابی بیوانفورماتیکی آلرژنیسته پروتئین‌های نوترکیب با استفاده از این پنج پایگاه‌های داده آلرژن انجام شد. نتایج تحقیق حاکی از آن بود که بر اساس آنالیزهای بیوانفورماتیکی اسمو تین حساسیت‌زا است.

سپاسگزاری از پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری در جهت فراهم شدن زمینه و

فرصت مناسب برای تهیه و ارائه این مقاله سپاسگزاری می‌شود.

منابع

- Aalberse, R. C. (2005). Assessment of sequence homology and cross-reactivity. [Review]. *Toxicology Applied Pharmacology*, 207(2 Suppl): 149-151.
- Allahyari Fard, N., & Minuchehr, Z. (2013). In Silico Analysis for Allergenicity Assessment of Novel Proteins of GMOs. *Genetics in the 3rd millennium*, 10(4): 0-15.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25(17): 3389-3402.
- Alvaro, M., Sancha, J., Larramona, H., Lucas, J. M., Mesa, M., Tabar, A. I., & Martinez-Canavate, A. (2013). Allergen-specific immunotherapy: Update on immunological mechanisms. *Allergol Immunopathol (Madr)*, 41(4): 265-272.
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Noble, W.

- S. (2009). MEME SUITE: tools for motif discovery and searching. [Research Support, N.I.H., Nucleic Acids Research, 37(Web Server issue): W202-208.
- Brusic, V., Millot, M., Petrovsky, N., Gendel, S. M., Gigonzac, O., & Stelman, S. J. (2003). Allergen databases. *Allergy*, 58(11): 1093-1100.
- Brusic, V., & Petrovsky, N. (2003). Bioinformatics for characterisation of allergens, allergenicity and allergic crossreactivity. [Review]. *Trends in Immunology*, 24(5): 225-228.
- Cohen, S. G., & King, J. R. (2001). Skin tests: a historic trail. *Immunol Allergy Clin North Clinics of North America*, 21(2): 191-249.
- Cui, Y. (2013). Structural biology of mite allergens. *Molecular Biology Reports*, 40(1): 681-686. Gendel, S. M.(2002). Sequence analysis for assessing potential allergenicity. 964: 87-98.
- Gendel, S. M. (2009). Allergen databases and allergen semantics. *Regulatory Toxicology and Pharmacology Pharmacol*, 54(3 Suppl), S7-10.
- Gendel, S. M., & Jenkins, J. A. (2006). Allergen sequence databases. *Molecular Nutrition & Food Research*, 633-637
- Gibson, J. (2006). Bioinformatics of protein allergenicity. *Molecular Nutrition & Food Research*, 50(7): 591.
- Goodman, R., Ebisawa, M., Sampson, H., van Ree, R., Vieths, S., Wise, J., & Taylor, S (2007) AllergenOnline, a peer-reviewed protein sequence database for assessing the potential allergenicity of genetically modified organisms and novel food proteins. *World Allergy Organization Journal*, S285.
- Harwanegg, C., Laffer, S., Hiller, R., Mueller, M. W., Kraft, D., Spitzauer, S., & Valenta, R. (2003). Microarrayed recombinant allergens for diagnosis of allergy. *Clinical & Experimental Allergy* , 33(1): 7-13.
- Hileman, R. E., Silvanovich, A., Goodman, R. E., Rice, E. A., Holleschak, G., Astwood, J. D., & Hefle, S. L. (2002). Bioinformatic methods for allergenicity assessment using a comprehensive allergen database. *International Archives of Allergy and Immunology*, 128(4): 280-291.
- Holloway, J. W., Yang, I. A., & Holgate, S. T. (2010). Genetics of allergic dis-

- ease. *Journal of Allergy and Clinical Immunology*, 125(2): 81-94.
- Ivanciuc, O., Schein, C. H., & Braun, W. (2002). Data mining of sequences and 3D structures of allergenic proteins. *Bioinformatics*, 18(10): 1358-1364.
- Ivanciuc, O., Schein, C. H., & Braun, W. (2003). SDAP: database and computational tools for allergenic proteins. *Nucleic Acids Research*, 31(1): 359-362.
- Ivanciuc, O., Schein, C. H., Garcia, T., Oezguen, N., Negi, S. S., & Braun, W. (2009). Structural analysis of linear and conformational epitopes of allergens. *Regulatory Toxicology Pharmacology*, 54(3 Suppl), S11-19.
- Kamble, S., & Bharmal, M. (2009). Incremental direct expenditure of treating asthma in the United States. *Journal of Asthma*, 46(1): 73-80.
- Ladics, G. S., Cressman, R. F., Herouet-Guicheney, C., Herman, R. A., Privalle, L., Song, P., McClain, S. (2011). Bioinformatics and the allergy assessment of agricultural biotechnology products: industry practices and recommendations. *Regulatory Toxicology and Pharmacology*, 60(1): 46-53.
- Mari, A., Rasi, C., Palazzo, P., & Scala, E. (2009). Allergen databases: current status and perspectives. *Current Allergy and Asthma Reports*, 9(5): 376-383.
- Mari, A., Scala, E., Palazzo, P., Ridolfi, S., Zennaro, D., & Carabella, G. (2006). Bioinformatics applied to allergy: allergen databases, from collecting sequence information to data integration. The Allergome platform as a model. *Cellular Immunology*, 244(2): 97-100.
- Markovic-Housley, Z., Degano, M., Lamba, D., von Roepenack-Lahaye, E., Clemens, S., Susani, M., Breiteneder, H. (2003). Crystal structure of a hypoallergenic isoform of the major birch pollen allergen Bet v 1 and its likely biological function as a plant steroid carrier. *Journal of Molecular Biology*, 325(1): 123-133.
- Martinez Barrio, A., Soeria-Atmadja, D., Nister, A., Gustafsson, M. G., Hammerling, U., & Bongcam-Rudloff, E. (2007). EVALLER: a web server for in silico assessment of potential protein allergenicity. *Nucleic Acids Research*, 35(Web Server issue), W694-700.
- Mora, C., Flores, I., Montealegre, F., & Diaz, A. (2003). Cloning and expression of Blo t 1, a novel allergen from the dust mite *Blomia tropicalis*, homolo-

- gous to cysteine proteases. *Clinical & Experimental Allergy*, 33(1): 28-34.
- Nakamura, R., Teshima, R., Takagi, K., & Sawada, J. (2005). [Development of Allergen Database for Food Safety (ADFS): an integrated database to search allergens and predict allergenicity. *Kokuritsu Iyakuhin Shokuhin Eisei Kenkyusho Hokoku*(123): 32-36.
- Ono, S. J. (2000). Molecular genetics of allergic diseases. *Annual Review of Immunology*, 18, 347-366.
- Pearson, W. R. (2000). Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol Biol*, 132, 185-219.
- Radauer, C., Bublin, M., Wagner, S., Mari, A., & Breiteneder, H. (2008). Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *Journal of Allergy and Clinical Immunology*, 121(4): 847-852 e847.
- Randhawa, G. J., Singh, M., & Grover, M. (2011). Bioinformatic analysis for allergenicity assessment of *Bacillus thuringiensis* Cry proteins expressed in insect-resistant food crops. *Food and Chemical Toxicology*, 49(2): 356-362.
- Saha, S., & Raghava, G. P. (2006). AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Research*, 34(Web Server issue), W202-209.
- Saito, H., Nakajima, T., & Matsumoto, K. (2001). Human mast cell transcriptome project. [Comparative Study Review]. *Int Arch Allergy Immunol*, 125(1): 1-8. doi: 53790
- Salehi, J., Tohidfar, M., & Sadeghi, A. (2010). Biosafety of genetically modified products: ABRIL.
- Schein, C. H. (1989). Production of soluble recombinant proteins in bacteria. *Nature Biotechnology*, 7(11): 1141-1149.
- Schein, C. H., Ivanciuc, O., Braun, W., Maleki, S., Burks, A., & Helm, R. (2006). Structural database of allergenic proteins (SDAP). *Food allergy*, 257-283.
- Schein, C. H., Ivanciuc, O., Midoro-Horiuti, T., Goldblum, R. M., & Braun, W. (2010).

- An Allergen Portrait Gallery: Representative Structures and an Overview of IgE Binding Surfaces. *Bioinformatics and Biology Insights*, 4: 113-125.
- Singh, M. B., & Bhalla, P. L. (2003). Hypoallergenic derivatives of major grass pollen allergens for allergy vaccination. *Immunology & Cell Biology*, 81(1): 86-91.
- Stadler, M. B., & Stadler, B. M. (2003). Allergenicity prediction by protein sequence. [Comparative Study]. *FASEB J*, 17(9), 1141-1143.
- Tong, J. C., Lim, S. J., Muh, H. C., Chew, F. T., & Tammi, M. T. (2009). Allergen Atlas: a comprehensive knowledge center and analysis resource for allergen information. *Bioinformatics*, 25(7), 979-980.
- Towhidfar, M., & Kaviani, M. (2010). Biosafety aspects of cotton biotechnology: ABRII. Valenta, R., & Kraft, D. (2001). Recombinant allergen molecules: tools to study effector cell activation. 179: 119-127.
- Vincent, K. J., & Zurini, M. (2012). Current strategies in antibody engineering: Fc engineering and pH-dependent antigen binding, bispecific antibodies and antibody drug conjugates. [Review]. *Biotechnology Journal*, 7(12):1444-1450.
- Wise, J., Taylor, S. L., & Goodman, R. E. (2009). Development and use of the AllergenOnline. org curated protein database for assessing the potential allergenicity of genetically modified organisms and novel food ingredients. *BIOT-2009*, 116.
- Yu, K., Petrovsky, N., Schonbach, C., Koh, J. Y., & Brusica, V. (2002). Methods for prediction of peptide binding to MHC molecules: a comparative study. *Molecular Medicine*, 8(3): 137-148.