مجله دانشکده پزشکی اصفهان
سال سی و پنجم/شماره‌ی ٤٤٠/هفته‌ی چهارم شهریور ماه ۱۳۹۶

تاریخ دریافت: ۱۳۹۶/۳/۱۷
تاریخ پذیرش: ۱۳۹۶/۳/۲۸

ارزیابی جامع بیوانفورماتیکی بیان ترشحی هورمون رشد انسانی با توالی‌های راهنمای مناسب در Escherichia coli: مطالعه‌ی In Silico

عارف دوزنده جویباری[۱]، حمیدرضا وزیری[۲]، شاهرخ قوتی[۳]، محمدمهدی سوهانی[۴]

مقاله پژوهشی

چکیده

مقدمه: امروزه، مهندسی ژنتیک و بیوانفورماتیک، تولید انبوه پروتئین‌های نوترکیب دارویی در باکتری Escherichia coli را که دارای ویژگی‌های منحصر به فرد بیانی است، به امری معمول و اقتصادی تبدیل کرده است. در این ارزیابی بیوانفورماتیکی، تولید پری‌پلاسمیک هورمون رشد انسانی بررسی گردید. هدف از انجام این تحقیق، ارزیابی جامع بیوانفورماتیکی ۴۸ توالی راهنمای انسانی توسط پایگاه‌های داده‌ی معتبر به منظور بررسی بیان پروتئین نوترکیب هورمون رشد انسانی در میزبان Escherichia coli بود.

روش‌ها: با استفاده از پایگاه داده‌ی قدرتمند SignalP نسخه‌ی ۴/۱، صحت و دقت جایگاه برش تعداد ۴۸ توالی راهنما ارزیابی و گزینش شدند. خصوصیات فیزیکوشیمیایی توالی راهنمای باقی‌مانده، توسط پایگاه‌های داده‌ی Genescript و Protparam مورد بررسی قرار گرفتند. قابلیت حل شدن پروتئین، فعالیت ترشحی پس از بیان و ساز و کار انتقال توالی‌های راهنما توسط پایگاه‌های داده‌ی Solpro، ProtCompB و PRED-TAT، بررسی شدند.

یافته‌ها: بر اساس نتایج به دست آمده به صورت تئوری، توالی‌های راهنمای مناسب برای اتصال به پروتئین هورمون رشد انسانی به ترتیب فسفوپروتئین بزاقی غنی شده با پرولین (Proline rich protein HaeIII subfamily 1 یا PRH1)، پروتئین ترشح شده‌ی C10orf99 و پپتید رها کننده‌ی پرولاکتین (Prolactin-releasing hormone یا PRLH) پیش‌بینی شد.

نتیجه‌گیری: بیان ترشحی، مزایایی را در مقابل بیان سیتوپلاسمیک ایجاد می‌کند. نتایج این تحقیق، نشان داد که با بررسی توالی‌های راهنمای مختلف در اتصال با پروتئین هورمون رشد انسانی، دستیابی به توالی‌های دارای پتانسیل و قابلیت بیان ترشحی بهتر امکان پذیر شده است. لازم است در مطالعات آتی صحت این نتایج بررسی گردد.

واژگان کلیدی: Escherichia coli، هورمون رشد انسانی، پری‌پلاسم، بیوانفورماتیک

ارجاع: دوزنده جویباری عارف، وزیری حمیدرضا، قوتی شاهرخ، سوهانی محمدمهدی. **ارزیابی جامع بیوانفورماتیکی بیان ترشحی هورمون رشد انسانی با توالی‌های راهنمای مناسب در Escherichia coli: مطالعه‌ی In Silico.** مجله دانشکده پزشکی اصفهان ۱۳۹۶؛ ۳۵ (۴۴۰): ۸۹۹-۸۹۰

مقدمه

هورمون رشد انسانی (Human growth hormone یا hGH) یک پلی‌پپتید تک زنجیره‌ای شامل ۱۹۱ آمینواسید است که در غده‌ی هیپوفیز ساخته می‌شود. این هورمون که با نام سوماتوتروپین نیز شناخته می‌شود، به دلیل نقش محوری در اعمال زیستی مختلف شامل سوخت و ساز و تکثیر سلول، یکی از مهم‌ترین هورمون‌ها در بدن انسان می‌باشد (۱) و در درمان بیماری‌هایی نظیر نقص هورمون رشد کودکان، نقص رشد ناشی از نقص کلیوی مزمن تا زمان پیوند، نقص هورمون رشد بالغین، سندرم Turner، سندرم Prader-Willi، نقص ژن SHOX و سندرم Noonan نقش دارد (۲).

hGH طبیعی، پروتئینی غیر گلیکوزیله است و فرم های نوترکیب آن، به طور گسترده در میزبان‌های پروکاریوتی، بیان شده‌اند. در میان

۱- دانشجوی کارشناسی ارشد، گروه زیست‌شناسی سلولی و مولکولی، دانشکده‌ی علوم پایه، دانشگاه گیلان، رشت، ایران
۲- دانشیار، گروه زیست‌شناسی سلولی و مولکولی، دانشکده‌ی علوم پایه، دانشگاه گیلان، رشت، ایران
۳- استادیار، گروه بیوتکنولوژی، دانشکده‌ی علوم کشاورزی، دانشگاه گیلان، رشت، ایران
۴- دانشیار، گروه بیوتکنولوژی، دانشکده‌ی علوم کشاورزی، دانشگاه گیلان، رشت، ایران
نویسنده‌ی مسؤول: شاهرخ قوتی
Email: ghovvati@guilan.ac.ir

میزبان‌های پروکاریوتی، باکتری Escherichia coli (E. coli) به دلیل داشتن ویژگی‌هایی مانند کشت به نسبت آسان و ارزان، دستیابی سریع به تراکم بالا، ژنتیک شناخته شده و تنوع ابزارهای ملکولی در دسترس، دارای برتری است. البته، فرابیان hGH در E. coli منجر به تجمع پروتئین نامحلول به صورت اجسام انکلوزیونی می‌شود. یک روش برای جلوگیری از تشکیل اجسام انکلوزیونی، ترشح hGH به فضای پری‌پلاسمیک E. coli با استفاده از توالی‌های راهنمای مناسب است (۳). در تولید ترشحی، به دلیل کاهش آلودگی ناشی از اجزای مختلف سلولی و تخریب پروتئولیتیک توسط پروتئازهای درون سلولی، جداسازی و خالص‌سازی محصولات فرابیان شده، تسهیل می‌شود (٤).

انتقال از سیتوزول به فضای پری‌پلاسم E. coli توسط مسیر وابسته به Sec برای پری‌پروتئین‌های تانخورده و مسیر جابه‌جایی دوگانه‌ی آرژنین (Tat) برای انتقال پری‌پروتئین‌های تاخورده، انجام می‌شود. هنگام عبور پری‌پروتئین از طریق کانال Tat A یا Sec YEG از غشای داخلی، توالی راهنما به وسیله‌ی سیگنال پپتیداز، جدا می‌شود (٥).

مطالعات مختلف اثبات کرده‌اند که انتخاب توالی راهنما، گزینه‌ی اصلی برای ترشح کارآمد پروتئین نوترکیب در E. coli است. برخی از مطالعات تجربی و آزمایشگاهی از ساز و کار ترشحی برای بیان هورمون رشد انسانی نوترکیب در E. coli استفاده کرده‌اند (٦). مطالعات In silico نشان داده‌اند که hGH نوترکیب هدایت شده به پری‌پلاسم، می‌تواند از توالی راهنما جدا شود (۷).

در این مطالعه، توالی‌های راهنمای انسانی و پتانسیل آن‌ها با استفاده از پایگاه‌های داده‌ی بیوانفورماتیکی برای ارزیابی ترشح کارآمد هورمون رشد انسانی نوترکیب در E. coli بررسی شد. به طور خلاصه، هدف از انجام این مطالعه، بررسی ٤٨ عدد از توالی‌های راهنمای انسانی به منظور دستیابی به توالی‌هایی بود که از لحاظ تئوری، بیشترین قابلیت بیان ترشحی به صورت فیوژن با هورمون رشد انسانی را در E. coli دارند.

روش‌ها

استخراج توالی‌های راهنما: توالی اسید آمینه‌ی ٤٨ عدد از توالی‌های راهنمای ترشحی با منشأ انسانی که پتانسیل ترشحی آن‌ها تأیید شده بود، با استفاده از پایگاه داده‌ی UniPortKB، از پایگاه داده‌ی Expasy جمع‌آوری شد. توالی‌های به دست آمده، با هدف معرفی توالی راهنمای مناسب برای افزایش ترشح hGH در میزبان E. coli، مورد بررسی قرار گرفتند.

پیش‌بینی جایگاه برش توالی‌های راهنما: در این پژوهش، از پایگاه داده‌ی SignalP استفاده شد که به آدرس http://www.cbs.dtu.dk/services/SignalP/ قابل دسترسی می‌باشد. SignalP از دقیق‌ترین و قابل اطمینان‌ترین ابزارهای موجود برای پیش‌بینی توالی‌های راهنما است و بهترین عملکرد را در تشخیص توالی‌های راهنما از توالی‌های پروتئینی با دقت ۸۷ درصد دارد (۹-۸). با استفاده از پایگاه داده‌ی SignalP نسخه‌ی ٤/۱ و بر اساس هوش مصنوعی پیش‌بینی توالی‌های راهنمای احتمالی و جایگاه برش آن‌ها صورت گرفت. برای مطالعه‌ی نواحی سه گانه‌ی n، h و c توالی‌های راهنما از نسخه‌ی ۳ پایگاه داده‌ی SignalP استفاده شد (۱۰).

تجزیه و تحلیل In Silico ویژگی‌های فیزیکوشیمیایی توالی‌های راهنما: با استفاده از پایگاه داده‌ی Protparam که به آدرس http://web.expasy.org/protparam/ قابل دسترسی می‌باشد، ویژگی‌های فیزیکوشیمیایی توالی‌های راهنمای مختلف شامل تعداد اسیدهای آمینه‌ی تشکیل دهنده، وزن ملکولی، شاخص ناپایداری، شاخص آلیفاتیک و GRAVY (میانگین کلی هیدروپاتیسیته) بررسی شد. ویژگی‌هایی نظیر بار خالص، pH ایزوالکتریک (Isoelectric point یا PI) توالی‌های راهنما نیز توسط قسمت Protein property calculator از سایت Genescript به آدرس https://www.genscript.com/ssl-bin/site2/peptide_calculation.cgi/ بررسی شد.

پیش بینی قابلیت حل شدن پروتئین: با استفاده از پایگاه داده‌ی SOLpro که به آدرس http://scratch.proteomics.ics.uci.edu/ در دسترس می‌باشد، قابلیت حل شدن پروتئین پس از بیش بیانی در E. coli پیش‌بینی شد. SOLpro از دقیق‌ترین نرم‌افزارهای موجود است و دقت آن بیش از ٧٤ درصد برآورد شده است (۱۲-۱۱).

پیش‌بینی فعالیت ترشحی پروتئین: با استفاده از نرم‌افزار SecretomeP نسخه‌ی ۲ که به آدرس http://www.cbs.dtu.dk/services/SecretomeP/ در دسترس است، مقدار SecP score که معیاری برای سنجش قابلیت ترشح پروتئین است، به منظور ارزیابی میزان ترشح هورمون رشد انسانی به صورت فیوژن با استفاده از توالی‌های راهنمای مختلف بررسی شد (۱۳).

پیش‌بینی نوع ساز و کار انتقال پروتئین ترشحی: با استفاده از پایگاه داده‌ی PRED-TAT به آدرس http://www.compgen.org/tools/PRED/TAT/، انواع ساز و کار ترشح Sec، Tat و TM هورمون رشد انسانی در اتصال با توالی‌های راهنمای مختلف پیش‌بینی شد (١٤).

یافته‌ها

برای دستیابی به توالی‌های راهنمای مطلوب، توالی آمینواسیدی ٤٨ عدد از توالی‌های راهنمای ترشحی با منشأ انسانی دارای پتانسیل ترشحی تأیید شده، از پایگاه داده‌ی UniPortKB برای بررسی دقیق در اتصال با توالی هورمون رشد انسانی، استخراج و ارزیابی شدند (جدول ١).

در این مطالعه، از نسخه‌ی ٤/١ پایگاه داده‌ی SignalP برای پیش‌بینی توالی‌های راهنمای مناسب در اتصال با hGH در میزبان E. coli استفاده شد. توالی‌های راهنما به طور معمول از ٣ ناحیه‌ی متفاوت تشکیل شده‌اند که شامل ناحیه‌ی انتهای N اغلب حاوی اسیدهای آمینه‌ی بازی، ناحیه‌ی آب گریز (ناحیه‌ی h) حاوی حداقل ٦ اسید آمینه و ناحیه‌ی انتهای C دارای اسیدهای آمینه‌ی قطبی بدون بار می‌باشند. برای تعیین نواحی سه‌گانه‌ی n، h و c در توالی‌های راهنما، از نسخه‌ی ٣ پایگاه داده‌ی SignalP استفاده شد (١٠) (جدول ٢).

بررسی ویژگی‌های فیزیکوشیمیایی تعداد ٣٢ توالی راهنمای مختلف باقی مانده، با استفاده از پایگاه داده‌ی Protparam صورت گرفت. بار خالص، GRAVY، شاخص‌های آلیفاتیک و شاخص ناپایداری توالی‌های راهنما مورد بررسی قرار گرفتند (جدول ٣).

با بررسی‌های صورت گرفته و نتایج حاصل از آن‌ها، به منظور انتخاب و رتبه‌بندی مناسب‌ترین توالی‌های راهنما در اتصال با hGH، توالی‌های تجزیه و تحلیل شده بر اساس ٥ شاخص آلیفاتیک، GRAVY Grand average of hydropathy، D-Score، طول ناحیه‌ی h و محل استقرار با یکدیگر مقایسه شدند. سپس مرتب‌سازی توالی‌های راهنما بر اساس شاخص آلیفاتیک انجام شد و تعداد ١٦ توالی راهنما با شاخص بالاتر از ١٥٠، برای ارزیابی‌های بیشتر مورد استفاده قرار گرفتند (جدول ٤).

از ویژگی‌های مهم هر پروتئین، قابلیت حل شدن است که رابطه‌ی مستقیم با چیدمان توالی اسیدآمینه‌ی آن دارد و از عوامل مهم در پروتئین‌های نوترکیب دارویی محسوب می‌شود. پایگاه‌های داده‌ی مختلفی برای پیش‌بینی قابلیت حل شدن پروتئین وجود دارد که در بین همه‌ی آن‌ها، پایگاه داده‌ی SOLpro دارای بیشترین دقت و اطمینان است (١٢). قابلیت حل شدن پروتئین hGH در اتصال با توالی‌های راهنمای مختلف و نتایج درصد حلالیت مربوط به هر ساختار در جدول ٣ آمده است.

با استفاده از نسخه‌ی ٢ پایگاه داده‌ی SecretomeP، می‌توان میزان ترشح پروتئین مورد نظر را با استفاده از توالی‌های راهنمای مختلف بررسی کرد و مقادیر SecP score را برای هر توالی راهنما به دست آورد. بررسی ساز و کار انتقال توالی‌های راهنمای منتخب توسط پایگاه داده‌ی PRED-TAT، بر اساس ساز و کار ترشحی Sec، Tat و یا TM انجام شد (جدول ٤).

بحث

در مطالعات مختلف، روش‌های بیوانفورماتیکی در بسیاری از زمینه‌های زیست‌شناختی به منظور کاهش هزینه‌ها و افزایش صحت مطالعات تجربی به کار رفته است، اما ماهیت پیش‌بینی کننده‌ی مطالعات بیوانفورماتیکی علاوه بر کاهش زمان و هزینه‌ی آزمایش، می‌تواند به عنوان یک محدودیت نیز برشمرده شود. در این مطالعه نیز با توجه به موارد پیش‌گفته، سعی گردید از پایگاه‌های داده‌ی مختلفی استفاده شود تا نتایج دارای صحت بالاتری حاصل شود. در پروتئین‌ها، فرایبان پروتئین هترولوگ منجر به تاخوردگی نامناسب می‌شود (١٦-١٥). با افزودن توالی راهنمای مناسب به انتهای N، می‌توان پروتئین‌ها را به فضای پری‌پلاسمیک هدایت کرد و مشکل را برطرف نمود (١٧). برای این هدف، ابزارهای محاسباتی مختلفی برای پیش‌بینی و شناسایی ویژگی‌های فیزیکوشیمیایی یک توالی راهنما مثل وزن مولکولی، نقطه‌ی ایزوالکتریک، GRAVY، شاخص آلیفاتیک و ... وجود دارد (١٦، ٧).

توالی‌های راهنمای انتخاب شده جهت کسب مقادیر بیشتر ترشح hGH در میزبان E. coli، مقایسه شدند. پایگاه UniProtKB از صحت مناسبی برخوردار است؛ به نحوی که در بیشتر مطالعات آزمایشگاهی و بیوانفورماتیکی، برای استخراج داده‌های خام جهت بررسی قابلیت ترشحی آن‌ها، از این پایگاه استفاده می‌شود (١٨).

مهم‌ترین خروجی حاصل از نسخه‌ی ٤/١ پایگاه داده‌ی SignalP شامل نمره‌ی D می‌باشد، که برای تشخیص توالی‌های راهنمای مناسب به کار می‌رود. در این مطالعه، این مقدار به طور پیش‌فرض روی ٠/٥ تنظیم شد. معیار دیگر برای بررسی قدرت برش توالی‌های راهنما که از پایگاه داده‌ی SignalP قابل استخراج می‌باشد، درصد احتمال برش است که در توالی‌های راهنمای بررسی شده به طور معمول بالاتر از ٠/٨ بود و به همراه D-score نشان دهنده‌ی توانایی لازم توالی‌های مورد بررسی برای برش در جایگاه صحیح است.

نتایج ارزیابی با استفاده از پایگاه داده‌ی SignalP نشان داد که در بین توالی‌های مورد بررسی، ٩ توالی راهنما دارای D-score پایین‌تر از ٠/٥ بودند و در نتیجه، به دلیل عدم توانایی کافی به عنوان توالی راهنمای مناسب و همچنین، عدم وجود پتانسیل لازم برای برش خوردن در جایگاه صحیح، از ادامه‌ی فرایند بررسی حذف شدند. همچنین، نتایج ارزیابی با این پایگاه داده، نشان داد اگر چه ٧ عدد از توالی‌ها D-score بالاتر از ٠/٥ دارند، اما فاقد پتانسیل برش پذیری در جایگاه مناسب هستند؛ در نتیجه، این توالی‌ها نیز حذف شدند؛ چرا که فاقد قابلیت لازم جهت ادامه‌ی مراحل بررسی بودند و به سبب باقی گذاردن بقایای اسید آمینه‌ای نادرست، احتمال ایجاد

جدول ۱. لیست توالی‌های راهنما

ردیف	نام توالی راهنما	نام ژن	شماره‌ی دسترسی	توالی اسیدآمینه‌ای
۱	Progonadoliberin-1	GNRH1	P01148	**MKPIQK**LLAGLILLTWCVEGCSS*
۲	Calcitonin gene-related peptide 1	CALCA	P06881	**MGFQKFSP**FLALSILVLLQAGSLHA
۳	ADM	ADM	P35318	**MKL**VSVALMYLGSLAFLGADT
۴	Relaxin-3	RLN3	Q8WXF3	**MARYML**LLLLLAVWVLTGELWPGAEA
۵	Urocortin-2	UCN2	Q96RP3	**MTR**CALLLLMVLML GRVLVVPV
۶	Guanylate cyclase activator 2B	GUCA2B	Q16661	**MGCRAASGLLP**GVAVVLLLLLQSTQS
۷	Calcitonin gene-related peptide 2	CALCB	P10092	**MGFRKFSP**FLALSILVLYQAGSLQA
۸	Neuroendocrine protein 7B2	SCG5	P05408	**MVSRMVST**MLSGLLFWLASGWTPAFA
۹	Prolactin-releasing peptide	PRLH	P81277	**MKVLRA**WLLCLLMLGLALRGAA
۱۰	Transforming growth factor beta-2	TGFB2	P61812	**MHY**CVLSAFLILHLVTVALS
۱۱	Insulin-like 3	INSL3	P51460	**MDPRLP**AWALVLLGPALVFA
۱۲	Klotho	KL	Q9UEF7	**MPASAPPRRPRPPPPS**LSLLLVLLGLGGRRLRA
۱۳	Meprin A subunit beta	MEP1B	Q16820	**MDLWNL**SWFLFLDALLVISGLA
۱۴	Salivary acidic proline-rich phosphoprotein	PRH1	P02810	**ML**LILLSVALLAFSSA
۱۵	Interleukin-8	CXCL8	P10145	**MTSK**LAVALLAAFLISAALC
۱۶	Transforming growth factor beta-1	TGFB1	P01137	**MPPSGLRL**LLLLLPLLWLLVLTPGRPAAG
۱۷	Basement membrane-specific heparan sulfate proteoglycan core protein	HSPG2	P98160	**MGWRA**AGALLLALLLHGRLLA
۱۸	Alpha-1-antitrypsin	SERPINA1	P01009	**MPSSVSWG**ILLLAGLCCLVPVSLA
۱۹	Plasminogen	PLG	P00747	**MEHKE**VVLLLLLFLKSGQG
۲۰	Vitamin K-dependent protein C	PROC	P04070	**MWQ**LTSLLLFVATWGISG
۲۱	Regenerating islet-derived protein 3-alpha	REG3A	Q06141	**MLPPMALPSV**SWMLLSCLMLLSQVQG
۲۲	Pituitary adenylate cyclase-activating polypeptide	ADCYAP1	P18509	**MTMCSGAR**LALLVYGIIMHSSVYS
۲۳	Pro-neuropeptide Y	NPY	P01303	**MLGNKRLGLSG**LTLALSLLVCLGALAEA
۲۴	Xylosyltransferase 1	XYLT1	Q86Y38	**MVAAPCARRLARRSHS**ALLAALTVLLLQTLVVWN
۲۵	Neuroendocrine secretory protein 55	GNAS	O95467	**MDRRSRAQQWRRARHNYNDLCPPIGRRAA**TA LLWLSCSIALLRALA
۲۶	Transforming growth factor beta-3	TGFB3	P10600	**MKMHLQR**ALVVLALLNFATVSLS
۲۷	Glutaminyl-peptide cyclotransferase	QPCT	Q16769	**MAGGRHRRV**VGTLHLLLLVAALPWASRG
۲۸	Protachykinin-1	TAC1	P20366	**MK**ILVALAVFFLVSTQLFA
۲۹	Beta-defensin 4A	DEFB4A	O15263	**MRV**LYLLFSFLFIFLMPLPGVFG
۳۰	C-X-C motif chemokine 5	CXCL5	P42830	**MSLLSSRAARVPGPSSS**LCALLVLLLLLTQPGPIAS
۳۱	Galanin peptides	GAL	P22466	**MARGS**ALLLASLLLAAALS
۳۲	CD5 antigen-like	CD5L	O43866	**MA**LLFSLILAICTRPGFLA
۳۳	Progonadoliberin-2	GNRH2	O43555	**MASSRRG**LLLLLLLTAHLGPSEA
۳۴	Tuberoinfundibular peptide of 39 residues	PTH2	Q96A98	**METRQVSRSPRVR**LLLLLLLLLVVPWGVRT
۳۵	Neurotensin/neuromedin N	NTS	P30990	**MMAGMKI**QLVCMLLLAFSSWSLC
۳۶	Gastrin-releasing peptide	GRP	P07492	**MRGRELP**LVLLALVLCLAPRGRA
۳۷	Promotilin	MLN	P12872	**MVSRKA**VAALLVVHVAAMLASQTEA
۳۸	C-type natriuretic peptide	NPPC	P23582	**MHLSQ**LLACALLLTLLSLRPSEA
۳۹	Orexigenic neuropeptide QRFP	QRFP	P83859	**MVRPYP**LIYFLFLPLGAC
۴۰	V-set and transmembrane domain-containing protein 4	VSTM4	Q8IW00	**MR**LLALAAAALLARAPAPEVCAA
۴۱	Galanin-like peptide	GALP	Q9UBC7	**MAPPSVP**LVLLLVLLLSLAETPAS
۴۲	Putative peptide YY-3	PYY3	Q5JQD4	**MVSVCRPWPA**VAIALLALLVCLG
۴۳	Follicular dendritic cell secreted peptide	FDCSP	Q8NFU4	**MKK**VLLLITAILAVAVG
۴۴	Secreted protein C10orf99	C10orf99	Q6UWK7	**MRL**LVLSSLLCILLLCFSIFSTEG
۴۵	Liver-expressed antimicrobial peptide 2	LEAP2	Q969E1	**MWHLK**LCAVLMIFLLLLGQIDG
۴۶	Serine protease inhibitor Kazal-type 4	SPINK4	O60575	**MAVRQW**VIALALAALLVVDREVPVAA
۴۷	Urotensin-2B	UTS2B	Q765I0	**MNKILSSTV**CFGLLTLLSVLSFLQSVHG
۴۸	Somatotropin	GH1	P01241	**MATGSRTS**LLLAFGLLCLPWLQEGSA

*آمینواسیدهای برجسته، آن‌هایی که زیرشان خط کشیده شده و توالی میانی، به ترتیب نشان دهنده نواحی n، c و h می‌باشند.

جدول ۲. بررسی **In silico** توالی‌های راهنما توسط پایگاه داده‌ی **SignalP** نسخه‌ی ۴/۱

توالی‌های راهنما	ناحیه‌ی n	ناحیه‌ی h	ناحیه‌ی c	جایگاه برش	احتمال برش (درصد)	نمره‌ی D
GNRH1	۶ (۱–۶)	۱۱ (۷–۱۷)	۶ (۱۸–۲۳)	CSS-MF	۰/۷۹۲	۰/۷۷۶
CALCA	۸ (۱–۸)	۱۰ (۹–۱۸)	۷ (۱۹–۲۵)	LHA-MF	۰/۸۹۷	۰/۷۹۵
ADM	۳ (۱–۳)	۱۱ (۴–۱۴)	۷ (۱۵–۲۱)	LGA-DT	۰/۴۳۷	۰/۶۱۵
RLN3	۵ (۱–۵)	۱۱ (۶–۱۶)	۹ (۱۷–۲۵)	AEA-MF	۰/۹۶۰	۰/۸۷۰
UCN2	۳ (۱–۳)	۱۱ (۴–۱۴)	۸ (۱۵–۲۲)	VLV-VP	۰/۷۲۲	۰/۵۵۷
GUCA2B	۱۱ (۱–۱۱)	۱۰ (۱۲–۲۱)	۵ (۲۲–۲۶)	TQS-MF	۰/۹۳۵	۰/۸۵۹
CALCB	۸ (۱–۸)	۹ (۹–۱۷)	۸ (۱۸–۲۵)	LQA-MF	۰/۷۶۳	۰/۷۱۴
SCG5	۸ (۱–۸)	۱۰ (۹–۱۸)	۸ (۱۹–۲۶)	AFA-MF	۰/۹۲۵	۰/۵۷۶
PRLH	۶ (۱–۶)	۱۱ (۷–۱۷)	۵ (۱۸–۲۲)	GAA-MF	۰/۶۸۲	۰/۸۷۹
TGFB2	۳ (۱–۳)	۱۰ (۴–۱۳)	۷ (۱۴–۲۰)	ALS-MF	۰/۹۱۳	۰/۶۵۹
INSL3	۶ (۱–۶)	۸ (۷–۱۴)	۶ (۱۵–۲۰)	NO	–	۰/۴۷۳
KL	۱۶ (۱–۱۶)	۱۰ (۱۷–۲۶)	۷ (۲۷–۳۳)	LRA-MF	۰/۷۹۷	۰/۵۵۳
MEP1B	۶ (۱–۶)	۱۱ (۷–۱۷)	۵ (۱۸–۲۲)	NO	–	۰/۴۰۴
PRH1	۲ (۱–۲)	۹ (۳–۱۱)	۵ (۱۲–۱۶)	SSA-MF	۰/۸۶۵	۰/۸۷۸
CXCL8	۴ (۱–۴)	۱۱ (۵–۱۵)	۵ (۱۶–۲۰)	ALC-MF	۰/۸۷۵	۰/۵۹۳
TGFB1	۸ (۱–۸)	۱۲ (۹–۲۰)	۹ (۲۱–۲۹)	TPG-RP	۰/۳۷۰	۰/۸۱۹
HSPG2	۶ (۱–۶)	۱۰ (۶–۱۶)	۵ (۱۷–۲۱)	LLA-MF	۰/۸۰۸	۰/۸۷۸
SERPINA1	۸ (۱–۸)	۱۱ (۹–۱۹)	۵ (۲۰–۲۴)	SLA-MF	۰/۹۵۵	۰/۶۵۸
PLG	۵ (۱–۵)	۹ (۶–۱۴)	۵ (۱۵–۱۹)	GQG-MF	۰/۹۵۵	۰/۶۹۵
PROC	۳ (۱–۳)	۱۰ (۴–۱۳)	۵ (۱۴–۱۸)	ISG-MF	۰/۶۳۰	۰/۵۶۰
REG3A	۱۰ (۱–۱۰)	۱۱ (۱۱–۲۱)	۵ (۲۲–۲۶)	VQG-MF	۰/۹۸۳	۰/۸۵۳
ADCYAP1	۸ (۱–۸)	۱۰ (۹–۱۸)	۶ (۱۹–۲۴)	NO	–	۰/۳۰۴
NPY	۱۱ (۱–۱۱)	۱۱ (۱۲–۲۲)	۶ (۲۳–۲۸)	AEA-MF	۰/۶۱۰	۰/۸۱۴
XYLT1	۱۶ (۱–۱۶)	۱۷ (۱۷–۳۳)	۱ (۳۴)	NO	–	۰/۳۳۵
GNAS	۲۹ (۱–۲۹)	۱۳ (۳۰–۴۲)	۴ (۴۳–۴۶)	NO	–	۰/۱۸۷
TGFB3	۷ (۱–۷)	۹ (۸–۱۶)	۷ (۱۷–۲۳)	SLS-MF	۰/۸۲۸	۰/۵۶۰
QPCT	۹ (۱–۹)	۱۱ (۱۰–۲۰)	۸ (۲۱–۲۸)	SRG-MF	۰/۸۸۲	۰/۷۹۶
TAC1	۲ (۱–۲)	۱۲ (۳–۱۴)	۵ (۱۵–۱۹)	NO	–	۰/۴۹۶
DEFB4A	۳ (۱–۳)	۱۲ (۴–۱۵)	۸ (۱۶–۲۳)	VFG-MF	۰/۹۹۱	۰/۵۰۵
CXCL5	۱۷ (۱–۱۷)	۱۱ (۱۸–۲۸)	۸ (۲۹–۳۶)	IAS-MF	۰/۷۹۷	۰/۷۵۸
GAL	۵ (۱–۵)	۹ (۶–۱۴)	۵ (۱۵–۱۹)	AAA-LS	۰/۷۴۴	۰/۷۶۵
CD5L	۲ (۱–۲)	۱۱ (۳–۱۳)	۶ (۱۴–۱۹)	GFL-AM	۰/۴۴۶	۰/۷۰۰
GNRH2	۷ (۱–۷)	۹ (۸–۱۶)	۷ (۱۷–۲۳)	SEA-MF	۰/۹۵۷	۰/۸۸۴
PTH2	۱۳ (۱–۱۳)	۱۱ (۱۴–۲۴)	۶ (۲۵–۳۰)	VRT-MF	۰/۹۵۶	۰/۵۵۳
NTS	۷ (۱–۷)	۱۱ (۸–۱۸)	۵ (۱۹–۲۳)	SLC-MF	۰/۷۰۵	۰/۵۹۸
GRP	۷ (۱–۷)	۱۱ (۸–۱۸)	۵ (۱۹–۲۳)	GRA-MF	۰/۹۹۲	۰/۸۴۶
MLN	۶ (۱–۶)	۱۱ (۷–۱۷)	۸ (۱۸–۲۵)	TEA-MF	۰/۸۱۷	۰/۷۶۲
NPPC	۵ (۱–۵)	۱۳ (۶–۱۸)	۵ (۱۹–۲۳)	SEA-MF	۰/۹۶۵	۰/۸۷۵
QRFP	۶ (۱–۶)	۱۰ (۷–۱۶)	۲ (۱۷–۱۸)	NO	–	۰/۳۴۹
VSTM4	۲ (۱–۲)	۱۱ (۳–۱۳)	۱۰ (۱۴–۲۳)	VCA-AM	۰/۶۰۵	۰/۶۲۸
GALP	۷ (۱–۷)	۱۰ (۸–۱۸)	۶ (۱۹–۲۴)	SLA-ET	۰/۵۶۳	۰/۸۴۴
PYY3	۱۰ (۱–۱۰)	۱۰ (۱۱–۲۰)	۳ (۲۱–۲۳)	NO	–	۰/۳۷۹
FDCSP	۳ (۱–۳)	۹ (۴–۱۲)	۱۷ (۱۳–۲۳)	NO	–	۰/۴۳۷
C10orf99	۳ (۱–۳)	۱۲ (۴–۱۵)	۹ (۱۶–۲۴)	TEG-MF	۰/۷۰۳	۰/۸۸۰
LEAP2	۵ (۱–۵)	۱۲ (۶–۱۷)	۵ (۱۸–۲۲)	IDG-MF	۰/۹۵۱	۰/۹۲۹
SPINK4	۶ (۱–۶)	۱۲ (۷–۱۸)	۹ (۱۹–۲۶)	VAA-MF	۰/۹۵۴	۰/۶۱۵
UTS2B	۹ (۱–۹)	۱۱ (۱۰–۲۰)	۸ (۲۱–۲۸)	VHG-MF	۰/۹۲۴	۰/۸۴۶
GH1	۸ (۱–۸)	۱۰ (۹–۱۸)	۸ (۱۹–۲۶)	GSA-MF	۰/۹۲۱	۰/۷۶۴

No: توالی راهنما دارای جایگاه برش در انتهای ناحیه‌ی c نمی‌باشد.

مشکلات پس از بیان ژن هدف در E. coli را افزایش می‌دهند و ممکن است ساختار سه بعدی پروتئین هورمون رشد انسانی را تغییر دهند که می‌تواند موجب ایجاد مشکلاتی در عملکرد و کارایی پروتئین شود. در نهایت، بر اساس نتایج حاصل از پایگاه داده‌ی SignalP، تعداد ۳۲ توالی راهنما واجد جایگاه برش صحیح و قدرت برش مناسب بودند که در ادامه‌ی بررسی‌ها ارزیابی شدند (جدول ۲).

در مطالعه‌ی زمانی و همکاران، از پایگاه داده‌ی SignalP برای پیش‌بینی جایگاه برش توالی‌های راهنما استفاده شد و ۲۲ توالی راهنمای پروکاریوتی و توالی راهنمای خود هورمون رشد انسانی در اتصال با این هورمون جهت ترشح بررسی شدند. نتایج مطالعه‌ی آن‌ها نشان داد که تنها ۱۷ توالی راهنما واجد شرایط مناسب برای ادامه‌ی بررسی بودند (۷).

جدول ۳. تعیین ویژگی‌های فیزیکوشیمیایی توالی‌های راهنما توسط ProtParam و SOLpro

ردیف	توالی‌های راهنما	طول اسید آمینه	وزن ملکولی	PI	بار خالص	GRAVY	شاخص آلیفاتیک	شاخص ناپایداری	قابلیت حل شدن	درصد قابلیت حل شدن
۱	GNRH1	۲۳	۲۵۰۴/۱	۷/۸۱	۰/۸۲	۰/۸۹۱	۱۳۵/۶۵	۹/۵۳	INSOLUBLE	۰/۷۶
۲	CALCA	۲۵	۲۶۸۹/۲	۸/۵۲	۱/۱۱	۱/۱۳۲	۱۳۲/۸۰	۵۵/۶۲	INSOLUBLE	۰/۸۱
۳	RLN3	۲۵	۲۸۱۷/۴	۴/۵۳	−۱/۰۰	۱/۱۳۲	۱۴۸/۴۰	۲۵/۳۶	INSOLUBLE	۰/۸۰
۴	GUCA2B	۲۶	۲۵۹۸/۱	۸/۰۰	۰/۹۱	۱/۲۱۵	۱۵۰/۰۰	۶۰/۴۶	INSOLUBLE	۰/۷۲
۵	CALCB	۲۵	۲۷۵۸/۳	۹۹/۹	۲/۰۰	۰/۸۷۶	۱۱۷/۲۰	۵۵/۶۲	INSOLUBLE	۰/۸۲
۶	SCG5	۲۶	۲۸۶۰/۴	۹/۵۰	۱/۰۰	۱/۰۳۸	۹۳/۸۵	۲۸/۱۴	INSOLUBLE	۰/۷۹
۷	PRLH	۲۲	۲۴۱۳/۱	۱۰/۸۶	۲/۹۱	۱/۵۲۳	۱۷۳/۱۸	۲۹/۸۴	INSOLUBLE	۰/۸۴
۸	TGFB2	۲۰	۲۲۳۰/۷	۶/۶۸	۰/۱۳	۱/۸۴۵	۱۷۰/۵۰	۶۱/۷۳	INSOLUBLE	۰/۸۲
۹	KL	۳۳	۳۵۱۳/۳	۱۲/۷۰	۶/۰۰	−۰/۰۴۵	۱۱۲/۴۲	۱۱۵/۲۶	INSOLUBLE	۰/۸۰
۱۰	PRH1	۱۶	۱۶۶۲/۱	۵/۲۸	۰/۰۰	۲/۴۵۰	۲۰۷/۵۰	۳۳/۴۵	INSOLUBLE	۰/۸۱
۱۱	CXCL8	۲۰	۲۰۰۷/۵	۷/۹۸	۰/۹۱	۱/۹۷۵	۱۶۱/۵۰	۳/۸۱	INSOLUBLE	۰/۸۱
۱۲	HSPG2	۲۱	۲۲۱۶/۷	۱۲/۰۰	۲/۱۱	۱/۲۸۶	۱۷۲/۳۸	۶/۴۲	INSOLUBLE	۰/۸۳
۱۳	SERPINA1	۲۴	۲۴۳۰/۰	۵/۲۷	−۰/۱۸	۱/۷۶۲	۱۵۸/۳۳	۶۸/۱۸	INSOLUBLE	۰/۷۲
۱۴	PLG	۱۹	۲۱۵۴/۶	۶/۵۱	۰/۱۱	۰/۶۷۴	۱۵۳/۶۸	۱۳/۵۰	INSOLUBLE	۰/۶۶
۱۵	PROC	۱۸	۲۰۲۳/۴	۵/۲۸	۰/۰۰	۱/۱۸۳	۱۳۰/۰۰	−۹/۳۸	INSOLUBLE	۰/۷۴
۱۶	REG3A	۲۶	۲۸۴۶/۵	۵/۲۷	−۰/۰۹	۱/۱۷۷	۱۳۱/۱۵	۷۴/۶۵	INSOLUBLE	۰/۸۱
۱۷	NPY	۲۸	۲۷۹۸/۴	۷/۹۸	۰/۹۱	۱/۲۳۲	۱۶۳/۹۳	۳۰/۵۶	INSOLUBLE	۰/۷۹
۱۸	TGFB3	۲۳	۲۵۵۶/۱	۱۱/۰۰	۲/۱۱	۱/۱۵۲	۱۵۲/۶۱	۵۶/۲۸	INSOLUBLE	۰/۸۱
۱۹	QPCT	۲۸	۳۰۰۸/۶	۱۲/۴۸	۴/۲۲	۰/۵۱۸	۱۲۸/۹۳	۴۴/۰۶	INSOLUBLE	۰/۷۸
۲۰	DEFB4A	۲۳	۲۷۲۱/۴	۸/۵۰	۱/۰۰	۱/۸۶۵	۱۴۳/۹۱	۳۳/۷۷	INSOLUBLE	۰/۷۷
۲۱	CXCL5	۳۶	۳۶۳۳/۴	۱۰/۳۵	۱/۹۱	۱/۰۱۴	۱۴۶/۳۹	۵۷/۲۸	INSOLUBLE	۰/۷۰
۲۲	GNRH2	۲۳	۲۴۱۹/۹	۹/۳۷	۱/۱۱	۰/۷۱۷	۱۴۸/۷۰	۶۵/۹۵	INSOLUBLE	۰/۷۴
۲۳	PTH2	۳۰	۳۵۱۵/۳	۱۲/۱۸	۴/۰۰	۰/۶۷۰	۱۶۵/۳۳	۴۸/۷۸	INSOLUBLE	۰/۷۰
۲۴	NTS	۲۳	۲۵۷۷/۲	۷/۸۲	۰/۸۲	۱/۵۴۸	۱۲۳/۰۴	۳۲/۵۳	INSOLUBLE	۰/۷۹
۲۵	GRP	۲۳	۲۵۱۸/۱	۱۱/۵۲	۲/۹۱	۰/۸۳۹	۱۵۶/۹۶	۲۰/۷۵	INSOLUBLE	۰/۸۴
۲۶	MLN	۲۵	۲۵۶۷/۱	۸/۵۲	۱/۱۱	۱/۱۱۶	۱۳۲/۸۰	۳۲/۷۱	INSOLUBLE	۰/۸۳
۲۷	NPPC	۲۳	۲۴۹۴/۰	۶/۵۰	۰/۰۲	۱/۰۷۰	۱۶۵/۶۵	۹۵/۴۴	INSOLUBLE	۰/۷۶
۲۸	C10orf99	۲۴	۲۶۷۲/۳	۵/۷۵	−۰/۱۸	۱/۸۲۵	۱۷۴/۵۸	۳۰/۵۲	INSOLUBLE	۰/۸۱
۲۹	LEAP2	۲۲	۲۵۲۸/۲	۶/۴۹	۰/۰۲	۱/۵۸۶	۱۷۷/۲۷	۱۲/۵۹	INSOLUBLE	۰/۷۷
۳۰	SPINK4	۲۶	۲۷۷۵/۳	۵/۸۲	۰/۰۰	۱/۴۳۸	۱۶۸/۸۵	۲۴/۶۴	INSOLUBLE	۰/۸۳
۳۱	UTS2B	۲۸	۳۰۰۸/۶	۸/۰۰	۱/۰۲	۱/۱۹۳	۱۴۲/۵۰	۵۲/۴۶	INSOLUBLE	۰/۸۱
۳۲	GH1	۲۶	۲۷۳۶/۲	۵/۷۵	−۰/۰۹	۰/۷۷۷	۱۱۶/۵۴	۵۶/۹۰	INSOLUBLE	۰/۷۴

PI: Isoelectric point; GRAVY: Grand average of hydropathy

جدول ۴. مرتب کردن توالی‌های راهنما بر اساس شاخص آلیفاتیک، GRAVY، طول ناحیه‌ی h و D-score به همراه ارزیابی نوع و فعالیت ترشحی توالی‌های راهنما

ردیف	توالی‌های راهنما	شاخص آلیفاتیک	GRAVY	نمره‌ی D	طول ناحیه‌ی h	مکانیسم	نمره‌ی Sec	نمره‌ی اطمینان
۱	PRH1	۲۰۷/۵۰	۲/۴۵۰	۰/۸۷۸	۹	Sec	۰/۹۱۱	۰/۹۲۶
۲	LEAP2	۱۷۷/۲۷	۱/۵۸۶	۰/۹۲۹	۱۲	TM	۰/۹۱۳	۰/۸۸۳
۳	C10orf99	۱۷۴/۵۸	۱/۸۲۵	۰/۸۸۰	۱۲	Sec	۰/۹۲۹	۰/۹۸۳
۴	PRLH	۱۷۳/۱۸	۱/۵۲۳	۰/۸۷۹	۱۱	Sec	۰/۹۲۱	۰/۹۹۴
۵	HSPG2	۱۷۲/۳۸	۱/۲۸۶	۰/۸۷۸	۱۰	Sec	۰/۹۲۱	۰/۹۷۱
۶	TGFB2	۱۷۰/۵۰	۱/۸۴۵	۰/۶۵۹	۱۰	TM	۰/۸۸۴	۰/۷۸۶
۷	SPINK4	۱۶۸/۸۵	۱/۴۳۸	۰/۶۱۵	۱۲	Sec	۰/۸۰۳	۰/۹۷۲
۸	NPPC	۱۶۵/۶۵	۱/۰۷۰	۰/۸۷۵	۱۳	Sec	۰/۹۰۸	۰/۹۹۳
۹	PTH2	۱۶۵/۳۳	۰/۶۷۰	۰/۵۵۳	۱۱	TM	۰/۸۳۵	۰/۷۷۳
۱۰	NPY	۱۶۳/۹۳	۱/۲۳۲	۰/۸۱۴	۱۱	Sec	۰/۹۱۹	۰/۹۹۴
۱۱	CXCL8	۱۶۱/۵۰	۱/۹۷۵	۰/۵۹۳	۱۱	Sec	۰/۹۱۴	۰/۹۴۴
۱۲	SERPINA1	۱۵۸/۳۳	۱/۷۶۲	۰/۶۵۸	۱۱	Sec	۰/۹۰۷	۰/۹۷۸
۱۳	GRP	۱۵۶/۹۶	۰/۸۳۹	۰/۸۴۶	۱۱	Sec	۰/۸۶۵	۰/۹۸۹
۱۴	PLG	۱۵۳/۶۸	۰/۶۷۴	۰/۶۹۵	۹	TM	۰/۸۴۹	۰/۸۲۸
۱۵	TGFB3	۱۵۲/۶۱	۱/۱۵۲	۰/۵۶۰	۹	Sec	۰/۹۰۷	۰/۹۳۸
۱۶	GUCA2B	۱۵۰/۰۰	۱/۲۱۵	۰/۸۵۹	۱۰	Sec	۰/۸۸۹	۰/۹۹۳

در مطالعه‌ی میرحسینی و همکاران، از این پایگاه داده برای پیش‌بینی جایگاه برش توالی‌های راهنمای پروتئین اینترفرون بتای انسانی استفاده شد. توالی‌های راهنمایی که واجد شرایط لازم برای ادامه‌ی بررسی بودند، حفظ و باقی توالی‌های راهنما حذف شدند (۱۹).

طول توالی‌های راهنمای انسانی که در این تحقیق بررسی شدند، در محدوده‌ی ۳۶-۱۶ آمینواسید قرار داشتند. همچنین، مقدار GRAVY که حاصل جمع مقادیر هیدروپاتی همه‌ی آمینواسیدها تقسیم بر تعداد اسیدآمینه در توالی است، برای توالی‌های راهنمای مختلف محاسبه شد (۲۰).

ویژگی‌هایی نظیر بار خالص و pH ایزوالکتریک توالی‌های راهنمای منتخب نیز توسط ابزار Protein property calculator سایت Genescript بررسی شدند. بار خالص توالی‌های راهنمای ارزیابی شده، در محدوده‌ی ۱/۰۰- برای RLN3 تا ۶/۰۰ برای KL متغیر بود.

مطالعات قبلی پیشنهاد نمودند که با افزایش سطوح هیدروفوبیسیتی (Hydrophobicity) و طول ناحیه‌ی h، نرخ ترشح پروتئین بهبود می‌یابد (۲۱). شاخص آلیفاتیک، عامل نشان دهنده‌ی مقدار آب‌گریزی توالی مورد بررسی است و با استفاده از حجم نسبی اشغال شده توسط زنجیره‌های جانبی آلیفاتیک اسیدهای آمینه‌ی آلانین، لوسین، ایزولوسین و والین در توالی آمینواسیدی مورد نظر، تعریف می‌شود. همچنین، این شاخص می‌تواند به عنوان عامل مثبتی برای افزایش پایداری دمایی پروتئین‌ها محسوب شود؛ به طوری که شاخص آلیفاتیک پروتئین‌های باکتری‌های ترموفیل به طور قابل توجهی از پروتئین‌های معمولی بالاتر است (۲۲). در این مطالعه، توالی‌های دارای مقادیر بالاتر از ۱۵۰ انتخاب شدند. شاخص ناپایداری که پایداری دمایی پروتئین‌ها را در حالت طبیعی بر اساس ساختمان اولیه پروتئین‌ها برآورد می‌نماید، برای توالی‌های منتخب محاسبه شد (جدول ۳). پروتئین‌های دارای شاخص ناپایداری کمتر از ۴۰، پایدار و پروتئین‌های بالاتر از ۴۰ به عنوان ناپایدار پیش‌بینی می‌شوند (۲۳).

سطوح هیدروفوبیسیتی توالی‌های راهنما می‌تواند با شاخص آلیفاتیک و GRAVY نشان داده شود. بنابراین، مهم‌ترین شاخص برای انتخاب توالی راهنمای مناسب، هیدوروفوبیسیتی است که شامل شاخص آلیفاتیک، GRAVY و طول ناحیه‌ی h است. از این رو، رتبه‌بندی داده‌های مربوط به آن‌ها در کنار شاخص‌های مهم دیگری نظیر D-score، امکان مقایسه‌ی مناسب آن‌ها را فراهم می‌کند.

نتایج پیش‌بینی قابلیت حل شدن نشان داد که hGH متصل به همه‌ی توالی‌های راهنمای منتخب، نامحلول است. این یافته، با توجه به ماهیت توالی اسیدآمینه‌ی hGH قابل انتظار و در تأیید نتایج حاصل از مطالعه‌ی قبلی ارزیابی توالی‌های راهنما در اتصال با هورمون رشد انسانی که به صورت نامحلول پیش‌بینی شدند بود (۷).

در پیش‌بینی قابلیت ترشحی، مقادیر بالاتر از ۰/۵ به معنی وجود قابلیت ترشح برای پروتئین مورد نظر است (۱۳). در این پژوهش، فعالیت ترشحی تعداد ۱۶ توالی راهنمای حاصل از گزینش بر اساس امتیاز شاخص آلیفاتیک، توسط پایگاه داده‌ی SecretomeP ارزیابی شد. بیشینه و کمینه‌ی این شاخص در محدوده‌ی ۰/۹۲۹–۰/۸۰۳ قرار داشت (جدول ٤). نتایج این تجزیه و تحلیل نشان داد که تمامی ۱۶ توالی راهنمای منتخب، دارای میزان مناسب قابلیت ترشح پروتئین هستند.

نتایج پیش‌بینی مکانیسم انتقال پروتئین ترشحی نشان داد که ٤ توالی راهنما از نوع تراغشایی (trans-membrane یا TM) و فاقد قابلیت ترشح به فضای پری‌پلاسمیک بودند (جدول ٤). همچنین، این نتایج نشان دادند که هیچ کدام از توالی‌های راهنما از نوع Tat نبودند؛ چرا که ساز و کار Tat نیازمند داشتن موتیفی با ۲ عدد Arg، طولانی‌ترین ناحیه‌ی n و ناحیه‌ی h آب‌گریز کوتاه‌تر در مقایسه با سایر توالی‌های راهنما است (۲٤). نتایج این ارزیابی نشان داد که بقیه‌ی توالی‌های راهنما از نوع ترشحی Sec بودند و اغلب دارای درصد اطمینان بالاتر از ۹۰ درصد بودند. این نتایج، بیانگر احتمال ترشح بسیار بالای آن‌ها در E. coli بود.

نتایج این مطالعه نشان دادند که سه توالی راهنمای Proline rich protein HaeIII subfamily 1 (PRH1)، C10orf99 و Prolactin-releasing hormone (PRLH) در مقایسه با سایر توالی‌های راهنمای بررسی شده در اتصال با هورمون رشد انسانی، پتانسیل ترشحی بهتری در E. coli دارند. همچنین، از بررسی نتایج استنباط شد که توالی راهنمای LEAP2 با وجود داشتن ویژگی‌های مناسب نظیر شاخص آلیفاتیک بالا، به دلیل ساز و کار ترشحی از نوع تراغشایی، فاقد قابلیت ترشح به فضای پری‌پلاسمیک است. در نتیجه، از دیدگاه تئوری، برای بیان ترشحی هورمون رشد در E. coli مناسب نمی‌باشد و فاقد پتانسیل لازم برای استفاده در E. coli است. در مطالعه‌ی قبلی که توالی‌های راهنمای مختلف پروکاریوتی در اتصال با هورمون رشد انسانی بررسی شدند، در نهایت، ۳ توالی راهنما به عنوان بهترین توالی‌های راهنما در اتصال با این هورمون جهت افزایش ترشح آن به فضای پری‌پلاسمیک انتخاب شدند. این توالی‌های راهنمای منتخب، همگی منشأ پروکاریوتی داشتند (۷).

نتیجه‌گیری نهایی این که مطالعه‌ای اختصاصی بر روی توالی‌های راهنما با منشأ انسانی جهت به کارگیری در تولید پری‌پلاسمیک هورمون رشد انسانی نوترکیب در E. coli و ارزیابی کارایی و بازده احتمالی آن انجام نشده بود. به دلیل زمان‌بر و پرهزینه بودن مطالعات تجربی هر یک از این توالی‌های راهنما با منشأ انسانی، از ابزارهای بیوانفورماتیکی متعدد و معتبر برای ارزیابی آن‌ها به صورت فیوژن استفاده شد که علاوه بر کاهش زمان و هزینه‌ها، افزایش دقت پژوهش‌های تجربی را نیز در پی دارد. نتایج نشان می‌دهند که از لحاظ تئوری، امکان به کارگیری توالی‌های راهنمای جدید با پتانسیل بهتر در مطالعات تجربی مرتبط با بیان پروتئین در میزبان‌های بیانی وجود دارد. همچنین، پیش‌بینی می‌شود نتایج این تحقیق بتواند روند پژوهش‌ها را در حوزه‌های مرتبط با بیان ترشحی انواع پروتئین‌های دارویی تسریع کند و حتی برای سایر پروتئین‌ها و توالی‌های راهنما با بازده بالایی مورد استفاده قرار گیرد. بر اساس نتایج به دست آمده، توالی‌های راهنمای مناسب به صورت فیوژن با پروتئین هورمون رشد انسانی که به صورت تئوری از پتانسیل و قابلیت بیان بهتری برخوردار بودند، به ترتیب فسفوپروتئین بزاقی غنی شده با پرولین (PRH1)، پروتئین ترشح شده‌ی C10orf99 و پپتید رها کننده‌ی پرولاکتین (PRLH) هستند. مطالعات آزمایشگاهی بیشتر در آینده، می‌توانند نتایج بیوانفورماتیکی این تحقیق را ارزیابی و بررسی نماید و با به کارگیری توالی‌های راهنمای تأیید شده در آزمایش‌های تجربی بعدی، دقت و صحت نتایج حاصل از مطالعه‌ی In silico مورد بررسی و کاوش دقیق‌تر قرار گیرد.

تشکر و قدردانی

این مقاله حاصل پایان‌نامه‌ی دوره‌ی کارشناسی ارشد ژنتیک مولکولی در دانشگاه گیلان می‌باشد.

References

1. Reh CS, Geffner ME. Somatotropin in the treatment of growth hormone deficiency and Turner syndrome in pediatric patients: A review. Clin Pharmacol 2010; 2: 111-22.
2. Franklin SL, Geffner ME. Growth hormone: the expansion of available products and indications. Pediatr Clin North Am 2011; 58(5): 1141-65, x.
3. Kim MJ, Park HS, Seo KH, Yang HJ, Kim SK, Choi JH. Complete solubilization and purification of recombinant human growth hormone produced in Escherichia coli. PLoS One 2013; 8(2): e56168.
4. Yoon SH, Kim SK, Kim JF. Secretory production of recombinant proteins in Escherichia coli. Recent Pat Biotechnol 2010; 4(1): 23-9.
5. Thanassi DG, Hultgren SJ. Multiple pathways allow protein secretion across the bacterial outer membrane. Curr Opin Cell Biol 2000; 12(4): 420-30.
6. Sockolosky JT, Szoka FC. Periplasmic production via the pET expression system of soluble, bioactive human growth hormone. Protein Expr Purif 2013; 87(2): 129-35.
7. Zamani M, Nezafat N, Negahdaripour M, Dabbagh F,

Ghasemi Y. In silico evaluation of different signal peptides for the secretory production of human growth hormone in E. coli. Int J Pept Res Ther 2015; 21(3): 261-8.
8. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 2004; 340(4): 783-95.
9. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. Nat Methods 2011; 8(10): 785-6.
10. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2007; 2(4): 953-71.
11. Garcia dlT, Carrasco B, Harding SE. SOLPRO: Theory and computer program for the prediction of SOLution PROperties of rigid macromolecules and bioparticles. Eur Biophys J 1997; 25(5-6): 361-72.
12. Magnan CN, Randall A, Baldi P. SOLpro: Accurate sequence-based prediction of protein solubility. Bioinformatics 2009; 25(17): 2200-7.
13. Bendtsen JD, Jensen LJ, Blom N, von HG, Brunak S. Feature-based prediction of non-classical and leaderless protein secretion. Protein Eng Des Sel 2004; 17(4): 349-56.
14. Bagos PG, Nikolaou EP, Liakopoulos TD, Tsirigos KD. Combined prediction of Tat and Sec signal peptides with hidden Markov models. Bioinformatics 2010; 26(22): 2811-7.
15. Baneyx F, Mujacic M. Recombinant protein folding and misfolding in Escherichia coli. Nat Biotechnol 2004; 22(11): 1399-408.
16. Baradaran A, Sieo CC, Foo HL, Illias RM, Yusoff K, Rahim RA. Cloning and in silico characterization of two signal peptides from Pediococcus pentosaceus and their function for the secretion of heterologous protein in Lactococcus lactis. Biotechnol Lett 2013; 35(2): 233-8.
17. Choi JH, Lee SY. Secretory and extracellular production of recombinant proteins using Escherichia coli. Appl Microbiol Biotechnol 2004; 64(5): 625-35.
18. Xiong E, Zheng C, Wu X, Wang W. Protein subcellular location: The gap between prediction and experimentation. Plant Mol Biol Rep 2016; 34(1): 52-61.
19. Mirhoseini Z, Pezeshkian Z, Ghovvati S. Phylogenetic and in silico analysis of interferon beta-1b protein. J Mazandaran Univ Med Sci 2017; 26(145): 70-82. [In Persian].
20. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol 1982; 157(1): 105-32.
21. Chen H, Kim J, Kendall DA. Competition between functional signal peptides demonstrates variation in affinity for the secretion pathway. J Bacteriol 1996; 178(23): 6658-64.
22. Ikai A. Thermostability and aliphatic index of globular proteins. J Biochem 1980; 88(6): 1895-8.
23. Guruprasad K, Reddy BV, Pandit MW. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. Protein Eng 1990; 4(2): 155-61.
24. Hinsley AP, Stanley NR, Palmer T, Berks BC. A naturally occurring bacterial Tat signal peptide lacking one of the 'invariant' arginine residues of the consensus targeting motif. FEBS Lett 2001; 497(1): 45-9.

Journal of Isfahan Medical School
Received: 07.06.2017
Vol. 35, No. 440, 4th Week, September 2017
Accepted: 18.06.2017

A Comprehensive Bioinformatic Assessment of Different Signal Peptides for Secretory Expression of Human Growth Hormone in Escherichia Coli: An In Silico Study

Aref Doozandeh-Juibari[1], Hamid Reza Vaziri[2], Shahrokh Ghovvati[3], Mohammad Mahdi Sohani[4]

Original Article

Abstract

Background: Nowadays, by genetic engineering and bioinformatics, large scale production of pharmacological recombinant proteins in Escherichia coli (E. coli) bacteria, which has unique expression properties, becomes a routine and economic imperative. In this study, periplasmic production of human growth hormone was investigated using bioinformatics methods.

Methods: The aim of this study was bioinformatic evaluation of 48 human signal peptides by reliable servers for expression analysis of human growth hormone in Escherichia coli. Accuracy and precision of 48 signal peptides were evaluated via powerful SignalP server. Physicochemical properties of remaining signal peptides were investigated using Genescript and Protparam servers. Solubility of protein, secretory activity of signal peptides after expression, and transmission mechanism of signal peptides were investigated using Solpro, ProtCompB and PRED-TAT, respectively.

Findings: Theoretically, proline rich protein HaeIII subfamily 1 (PRH1), C10orf99, and prolactin-releasing hormone (PRLH) signal peptides were predicted as the most proper signal peptides in fusion of human growth hormone protein, respectively.

Conclusion: Secretory expression instead of cytoplasmic expression provides benefits. This study results indicated that by examining different signal peptide sequences in fusion with human growth hormone protein, achieving signal peptides with potential and capability for high expression is possible. The accuracy of these results can be verified in future studies and experiments.

Keywords: Escherichia coli, Human growth hormone, Periplasm, Bioinformatics

Citation: Doozandeh-Juibari A, Vaziri HR, Ghovvati S, Sohani MM. **A Comprehensive Bioinformatic Assessment of Different Signal Peptides for Secretory Expression of Human Growth Hormone in Escherichia Coli: An In Silico Study.** J Isfahan Med Sch 2017; 35(440): 890-9.

1- MSc Student, Department of Molecular Cell Biology, School of Applied Science, University of Guilan, Rasht, Iran
2- Associate Professor, Department of Molecular Cell Biology, School of Applied Science, University of Guilan, Rasht, Iran
3- Assistant Professor, Department of Biotechnology, School of Agriculture, University of Guilan, Rasht, Iran
4- Associate Professor, Department of Biotechnology, School of Agriculture, University of Guilan, Rasht, Iran
Corresponding Author: Shahrokh Ghovvati, Email: ghovvati@guilan.ac.ir