

ORIGINAL ARTICLE

Bioinformatic Investigation of Protein Stability (SQS) in Prokaryotes and Eukaryotes

Masoud Tohidfar^{1*}, Maryam Mehdizadeh Hakkak², Mohammad Hossein Mirjalili³

¹Department of Cell and Molecular Biology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran.

²Department of Cell and Molecular Biology, Faculty of Life Sciences and Biotechnology, Shahid Beheshti University, Tehran, Iran.

³Department of Agricultural Engineering, Research Institute of Medicinal Plants and Raw Materials, Shahid Beheshti University, Tehran, Iran.

ABSTRACT

Squalane is an unsaturated triterpene that has wide applications in pharmaceuticals. In this research, the production of squalene and its bioinformatic analysis in four species of unicellular and multicellular eukaryotes and prokaryotes were investigated in order to determine the difference of this gene in eukaryotes and prokaryotes. The results of phylogenetic analysis showed that algae as multicellular eukaryotes, yeast as single-celled eukaryote and bacteria as single-celled prokaryote were placed in one group and plants were placed in a separate group. GC percentage of SQS protein was evaluated by GC Content Calculator, as well as aliphatic index and instability index by protparam. The results showed that the SQS gene is in a range from unstable to stable. The analysis of the presence of signal sequences and the analysis of the detection of the final location of the protein showed that the possibility of transferring the SQS protein to the mitochondria, chloroplast and secretory pathway is very low and it is not among the signal proteins. It was also found in *Gymnema sylvestre* that this protein has three protected domains. The comparison of the secondary structure of the protein confirmed the existence of alpha sheets. 3D modeling of this protein in plant was done by homology modeling method and using Swiss Model database after selecting a suitable model with high similarity which was extracted from PDB database. In order to validate the structure of the drawn three-dimensional model and stereochemical analysis, the Ramachandran diagram was drawn and the dihedral angles were calculated. The results of structural quality evaluation showed that the proposed models have good quality and stability. The study of the protein structure can help to understand the function of the protein, and studying the details of its structure can be useful in the studies of the active site of the protein and docking.

KEYWORDS

Aliphatic Index, 3D modeling, GC Percentage, Phylogenetic Analysis, SQS protein.

Correspondence

Masoud Tohidfar

Email: m_tohidfar@sbu.ac.ir

How to cite

Tohidfar, M., Mehdizadeh Hakkak, M., & Mirjalili, M. H. (2023). Bioinformatic Investigation of Protein Stability (SQS) in Prokaryotes and Eukaryotes. *Crop Biotechnology*, 12(41), 1-9.

نشریه علمی

زیست فناوری گیاهان زراعی

«مقاله پژوهشی»

بررسی بیوانفورماتیکی پروتئین (SQS) در پروکاریوت‌ها و یوکاریوت‌ها

مسعود توحیدفر^{۱*}، مریم مهدیزاده حاکک^۲، محمدحسین میرجلیلی^۳

چکیده

اسکوالن یک تری‌ترین غیر اشباع با کاربردهای گسترده در داروسازی است. در این تحقیق تولید اسکوالن و بررسی بیوانفورماتیکی آن در چهار گونه یوکاریوت تک‌سلولی و پرسلولی و پروکاریوت به منظور تفاوت این ژن در یوکاریوت‌ها و پروکاریوت‌ها مورد بررسی قرار گرفت. نتایج آنالیز فیلوژنتیکی نشان داد که جلبک به عنوان یوکاریوت پرسلولی، مخمر به عنوان یوکاریوت تک‌سلولی و باکتری به عنوان پروکاریوت تک‌سلولی در یک گروه و گیاه در گروه مجزا قرار گرفت. درصد GC پروتئین SQS توسط GC Content Calculator و همچنین شاخص آلیفاتیک و شاخص ناپایداری توسط protparam بررسی شد. نتایج نشان داد که ژن SQS در دامنه‌ای از ناپایداری تا پایدار قرار دارد. آنالیز وجود توالی‌های سیگنال و آنالیز رهگیری محل استقرار نهایی پروتئین نشان داد که احتمال انتقال پروتئین SQS به میتوکندری، کلروپلاست و مسیر ترشحی بسیار پایین است و جزء پروتئین‌های سیگنال نیست. همچنین در گیاه *Gymnema sylvestre* مشخص شد که این پروتئین دارای سه دومین حافظت شده است. مقایسه‌ی ساختار ثانویه پروتئین وجود صفحات آلفا را تایید کرد. مدل‌سازی سه بعدی این پروتئین در گیاه به روش همولوژی مدلینگ و با استفاده از پایگاه داده Swiss Model پس از انتخاب الگوی مناسب با میزان شباهت بالا که از پایگاه داده PDB استخراج شد، انجام گرفت. جهت اعتبارسنجی ساختاری مدل ترسیم شده سه بعدی و آنالیز استرئوشیمیایی، نمودار رامچاندرا ترسیم و زوایای دی‌هیدرال محاسبه شدند. نتایج ارزیابی کیفیت ساختاری نشان داد که مدل‌های پیشنهادی دارای کیفیت و پایداری مناسبی می‌باشند. مطالعه ساختار پروتئین می‌تواند به درک عملکرد پروتئین کمک کند و بررسی جزئیات ساختار آن می‌تواند در مطالعات جایگاه فعال پروتئین و داکینگ سودمند باشد.

واژه‌های کلیدی

آنالیز فیلوژنتیکی، پروتئین SQS، درصد GC، شاخص آلیفاتیک، مدل‌سازی سه بعدی.

^۱گروه زیست‌شناسی سلولی و مولکولی، دانشکده علوم و فناوری زیستی، دانشگاه شهید بهشتی، تهران، ایران.
^۲گروه زیست‌شناسی سلولی و مولکولی، دانشکده علوم و فناوری زیستی، دانشگاه شهید بهشتی، تهران، ایران.
^۳گروه مهندسی کشاورزی، پژوهشکده گیاهان و مواد اولیه دارویی، دانشگاه شهید بهشتی، تهران، ایران.

نویسنده مسئول:

مسعود توحیدفر

رایانامه: m_tohidfar@sbu.ac.ir

استناد به این مقاله:

توحیدفر، مسعود، مهدیزاده حاکک، مریم و میرجلیلی، محمدحسین (۱۴۰۲). بررسی بیوانفورماتیکی پروتئین (SQS) در پروکاریوت‌ها و یوکاریوت‌ها. فصلنامه علمی زیست‌فناوری گیاهان زراعی، ۱۲(۴۱)، ۹-۱.

مقدمه

اسکوالن یک هیدروکربن تری ترین است که در روغن زیتون، پالم، تاج‌خروس و سیوس برنج یافت می‌شود. این ترکیب غیراشباع و شامل شش پیوند دوگانه است که غنی‌ترین منبع تولید آن، روغن کبد کوسه ماهیان می‌باشد که به‌صورت سنتی استفاده‌های متعددی دارد (آچیمویچ و رزمن، ۲۰۱۳). در انسان اسکوالن در سیوم و غدد ترشحی پوست یافت می‌شود. این ترکیب یک حدواسط بسیار مهم جهت سنتز استرول‌ها و هورمون‌ها در سلول است. همچنین اسکوالن پیش‌ماده مهم برای سنتز تری‌ترین‌ها است (مرتضوی، زارع نژاد و علویان، ۲۰۱۶).

پیشینه پژوهش

از کاربردهای اسکوالن و تولید تری‌ترین در گیاهان دارویی مانند سوبه بومی آئورانتیوکیتریوم، شیرین بیان بومی ایران، لاله واژگون می‌توان به استفاده از آن به‌عنوان ادجوانت واکسن‌ها، مهارکننده سلول‌های سرطانی و خاصیت آنتی‌اکسیدانتی آن اشاره کرد. میکروارگانیزم‌های متعددی توانایی تولید اسکوالن را دارند که شامل باکتری‌ها، مخمرها، قارچ‌ها، ریزجلبک‌ها و ترائوستوکیتریدها هستند (ماتسورا، واتانابی و کایا، ۲۰۱۲). به دلیل اهمیت ژن اسکوالن سنتاز که در بالا اشاره شد در این تحقیق به بررسی بیوانفورماتیکی این ژن پرداخته می‌شود تا از نتایج آن در تحقیقات آینده استفاده شود. همچنین در این تحقیق به بررسی این ژن در یوکاریوت‌ها و پروکاریوت‌ها پرداخته می‌شود.

روش شناسی پژوهش

بررسی پایداری پروتئین اهمیت زیادی دارد. در این تحقیق ژن SQS در گیاه *Gymnema sylvestre* با اکسشن نامبر MT892813.1، جلبک *Chlorella variabilis* با اکسشن نامبر XM_005851928.1 به عنوان یوکاریوت‌های پرسلولی و در مخمر *Yarrowia lipolytica* با اکسشن نامبر AF092497.1 به عنوان یوکاریوت تک‌سلولی و ژن SQS در باکتری *Stigmatella aurantiaca* با اکسشن نامبر AJ494839.1 به عنوان پروکاریوت انتخاب شدند. توالی ژن و پروتئین آنها از سایت NCBI دریافت و آنالیزهای بیوانفورماتیکی بر روی این توالی‌ها انجام شد. از برنامه‌ی Bioedit2007 برای هم‌ردیفی چندگانه‌ی توالی‌ها استفاده شد (https://softfamous.com/bioedit/). برای رسم درخت

فیلوژنتیک نرم‌افزار MEGA5 به‌کار برده شد (http://www.megasoftware.net).

برای بررسی خصوصیات ژن SQS محتوی GC با استفاده از Genscan web server (http://hollywood.mit.edu/GENSCAN.html) به دست آمد. به منظور آنالیز ناحیه پروموتور ژن SQS از نرم‌افزار PromoterScan (http://159.149.160.88/pscan/) استفاده شد. هم‌ردیفی چندگانه‌ی توالی پروتئینی پروتئین SQS با استفاده از برنامه‌ی MultiAlign انجام شد. همچنین هم‌ردیفی چندگانه‌ی توالی پروتئین SQS با برنامه‌ی MultiAlign (به منظور تایید نتیجه‌ی توالی‌یابی با نرم‌افزار Bioedit) انجام شد. به منظور بررسی خصوصیات ساختار اولیه‌ی پروتئین از برنامه‌ی PROTPARM استفاده شد. برای بررسی محل استقرار و نقش مولکولی پروتئین در سلول از برنامه‌ی UNIPROT استفاده شد. آنالیز وجود توالی‌های سیگنال توسط SignalP مورد بررسی قرار گرفت (https://github.com/fteufel/signalp-6.0). محل‌های استقرار پروتئین توسط برنامه‌ی MitoProt (http://ihg.gsf.de/ihg/mitoprot.html) مشخص شد. از برنامه‌ی جامع InterProScan (www.ebi.ac.uk/interpro) به منظور شناسایی موتیف‌ها و دامین‌های حفاظت شده در پروتئین استفاده شد. از ابزار آنالیز Phyre2 (http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index) به منظور تعیین ساختار دوم پروتئین و مشخص شدن صفحات آلفا و بتا استفاده شد. مدل‌سازی سه بعدی این پروتئین به روش همولوژی مدلینگ و با استفاده از پایگاه داده‌ی Swiss Model پس از انتخاب الگوی مناسب با میزان شباهت بالا که از پایگاه داده‌ی PDB استخراج شد، انجام گرفت. جهت اعتبارسنجی ساختاری مدل ترسیم شده سه بعدی و آنالیز استرئوشیمیایی، نمودار رامانچاندرا ترسیم و زوایای دی‌هیدرال محاسبه شدند. مطالعه ساختار پروتئین می‌تواند به درک عملکرد پروتئین کمک کند و بررسی جزئیات ساختار آن می‌تواند در مطالعات جایگاه فعال پروتئین و Docking¹ سودمند باشد. آنالیز هضم توسط PeptideCutter (http://us.expasy.org/tools/peptidecutter) بررسی شد. تغییرات پس از ترجمه بوسیله‌ی ScanProsite (http://www.expasy.org/tools/scanprosite/) مشخص شد.

یافته های پژوهش و نتیجه گیری

آنالیزهای فیلوژنتیکی

برای انجام آنالیز فیلوژنتیکی از ۴ گونه‌ی انتخاب شده استفاده شد. توالی گونه‌های مورد نظر که از سایت NCBI دریافت شده بودند در گام اول با نرم‌افزار Bioedit2007 هم‌ردیف‌سازی شدند و سپس برای انجام آنالیزهای فیلوژنتیکی از نرم‌افزار MEGA5 استفاده شد. آنالیز فیلوژنی با روش UPGMA دندروگرامی ایجاد کرد که ۴ گونه‌ی مورد مطالعه را در ۲ گروه کاملاً مجزا دسته‌بندی کرد (شکل ۱). مدل جایگزینی با آنالیز توسط نرم‌افزار MEGA5 به دست آمد و با مقایسه‌ی تفاوت توالی‌ها بهترین درخت فیلوژنتیکی رسم شد (ماتسورا، واتانابی و کایا، ۲۰۱۲). نتایج نشان داد که به جز گیاه، جلبک به عنوان یوکاریوت پرسلولی، مخمر به عنوان یوکاریوت تک‌سلولی و باکتری به عنوان پروکاریوت تک‌سلولی در یک گروه قرار گرفتند و گیاه در یک گروه مجزا قرار گرفت (شکل ۱). درخت فیلوژنتیکی بر اساس شباهت توالی‌ها، زمان اشتقاق گونه‌ها را مشخص می‌کند و گونه‌ها را بر اساس شباهت و تفاوتشان در یک گروه یا گروه‌های مختلف قرار می‌دهد.

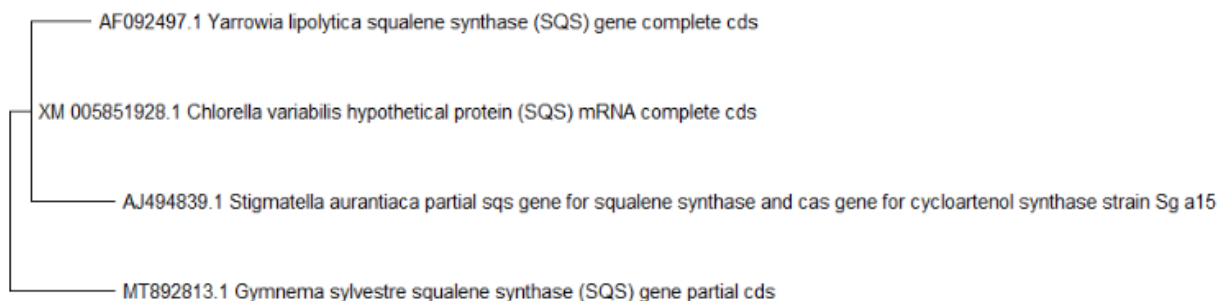
آنالیز درصد GC

درصد GC در مخمر *Yarrowia lipolytica* ۴۷، در گیاه *Chlorella variabilis* ۳۴/۲۵، در جلبک *Gymnema sylvestre* ۶۷/۸۱ و در باکتری *Stigmatella aurantiaca* ۶۵/۶۴ درصد می‌باشد. نتایج نشان داد که درصد GC در یوکاریوت‌های پرسلولی و تک‌سلولی و پروکاریوت‌ها در دامنه‌ی ای از ناپایدار تا پایدار قرار دارد و ثابت نیست (فلاح زیارانی، توحیدفر و امین فر، ۲۰۱۷). آنالیز محتوی GC مشخص کرد که ژن‌هایی با محتوی GC بالا پایدارتری در برابر حرارت نسبت به

ژن‌هایی با محتوی GC کمتر دارند (لاول و همکاران، ۲۰۰۳). تنوع در محتوی گوانین (G) و سیتوزین (C) مشاهده شده در بین گونه‌ها یکی از موضوعات مورد توجه در تکامل بیوانفورماتیکی است. چگونگی ارزش GC با محاسبه‌ی محتوی GC به کل ژنوم در موجودات مورد مطالعه می‌باشد. فرگوسون و همکاران در سال ۲۰۱۱ و کائو در سال ۲۰۱۱ اعلام کردند که تنوع در محتوی درصد GC در یوکاریوت‌ها کمتر می‌باشد (۸، ۷). علاوه بر این، آنالیز محتوی درصد GC به‌وسیله‌ی سرور ProtParam مشخص کرد که ژن‌هایی با درصد GC بالاتر در مقایسه با ژن‌هایی با درصد GC کمتر پایدارتر می‌باشند. این ممکن است ناشی از این حقیقت باشد که جفت‌های GC توسط سه باند هیدروژنی به هم متصل می‌شوند، اما جفت‌های AT بوسیله دو باند هیدروژنی باند می‌شوند، باندهای هیدروژنی به پایداری بیشتر محصولات ژن کمک می‌کنند و تعداد باندهای هیدروژنی بیشتر نتیجه‌اش پایداری بیشتر پروتئین می‌باشد. هر چه درصد GC بیشتر باشد بیشتر قادر به حفظ ساختار فضایی خود در دماهای بالاتر می‌باشد (بورگ و کارلین، ۱۹۹۷).

بررسی خصوصیات ساختار اولیه‌ی پروتئین

شاخص ناپایداری در مخمر ۴۷/۹۴، در گیاه ۳۳/۶۸، در جلبک ۵۳/۴۳، در باکتری ۴۸/۲۴ درصد می‌باشد. شاخص ناپایداری محاسبه شده برای پروتئین‌های مورد بررسی نشان می‌دهند که پروتئین‌هایی که شاخص ناپایداری آنها کوچکتر از ۴۰ باشند با ثبات و در غیر اینصورت بی‌ثبات خواهند بود. نتایج نشان داد که به جز گیاه، ژن SQS در باکتری، جلبک و مخمر شاخص ناپایداری بالاتر از ۴۰ دارند. پس ژن SQS دامنه‌ی متفاوتی از پایداری در یوکاریوت‌ها و پروکاریوت‌ها دارند. شاخص آلیفاتیک در مخمر ۲۵/۸۲، در گیاه ۲۷/۶۱، در جلبک ۱۵/۹۰، در باکتری ۱۵/۶۲ می‌باشد.



شکل ۱. دندروگرام به دست آمده از گونه‌های مورد مطالعه با روش UPGMA

شاخص آب‌گریزی پروتئین می‌باشد. در صورتی که GRAVY محاسبه شده برای یک پروتئین منفی باشد، بدین معنی است که آن پروتئین غیرقطبی است و در صورت مثبت بودن آن قطبی محسوب می‌شود (فلاح زیارانی و همکاران، ۲۰۲۱). GRAVY محاسبه شده برای پروتئین SQS در هر ۴ مورد بررسی مثبت می‌باشد و نشان‌دهنده‌ی قطبی بودن آن می‌باشد. حلالیت پروتئین‌های قطبی بیشتر از حلالیت پروتئین‌های غیرقطبی در مواد حلال می‌باشد. در برابر حلال‌های ویژه‌ی هر پروتئین، پروتئین‌های غیر قطبی در وسط و پروتئین‌های قطبی در قسمت بیرونی تجمع پیدا می‌کنند (موراکامی و همکاران، ۲۰۰۵). به دلیل اهمیت ژن SQS در گیاه، آنالیزهای بعدی بر روی ژن SQS در گیاه انجام شد.

آنالیز وجود توالی‌های سیگنال

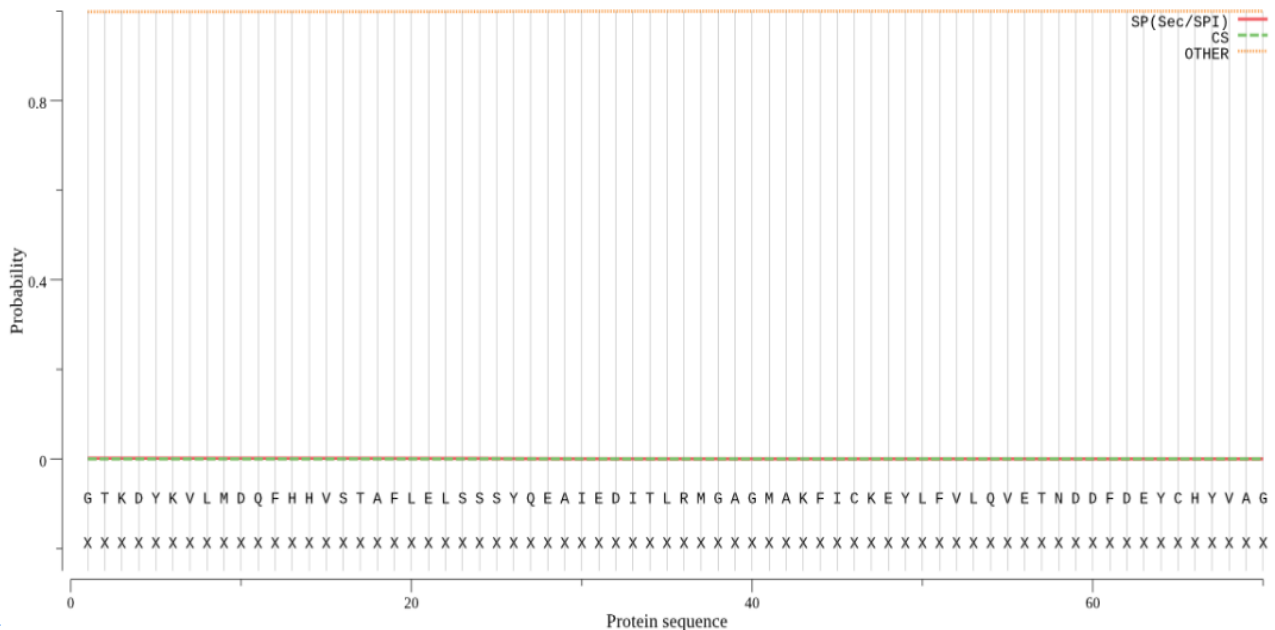
نتایج آنالیز سیگنال ژن SQS با استفاده از signal P در جدول زیر مشخص شده است (جدول ۱؛ شکل ۲). نتایج نشان داد که این پروتئین جزو پروتئین‌های سیگنال نمی‌باشد (جدول ۱؛ شکل ۲).

شاخص آلیفاتیک در واقع عبارت است از حجم نسبی پروتئین که با زنجیره‌های آلیفاتیک اشغال شده‌اند (A, V, I, L) و به عنوان یک فاکتور مثبت در افزایش ثبات حرارتی پروتئین‌های گلوبول مانند محسوب می‌شود و به همین دلیل در باکتری‌های مقاوم به گرما، پروتئین‌ها دارای شاخص آلیفاتیک بالایی می‌باشند. پروتئین‌هایی با شاخص آلیفاتیک بسیار بالا (بالتر از ۱۰۰) ممکن است در دامنه‌ی دمایی بسیار بالایی از خود ثبات نشان دهند. چون شاخص آلیفاتیک پروتئین مورد مطالعه در ۴ مورد بررسی کمتر از ۱۰۰ می‌باشد، این پروتئین جزو پروتئین‌های غیرپایدار است و به خوبی قادر به حفظ ساختار سه بعدی خود در دمای بالا نمی‌باشد (فلاح زیارانی و صفائی زاده، ۲۰۲۲). متوسط GRAVY (خاصیت آب‌گریزی پروتئین) در مخمر ۰/۷۸۱، در گیاه ۰/۵۰۵، در جلبک ۰/۸۷۹، در باکتری ۰/۸۱۱ به دست آمد. متوسط GRAVY کل محاسبه شده برای پروتئین‌ها از تقسیم مجموع هیدروپاتی محاسبه شده برای تمام اسیدهای آمینه در پروتئین بر تعداد کل اسیدهای آمینه‌ی آن پروتئین به دست می‌آید. متوسط هیدروپاتی کل (GRAVY) نشان‌دهنده‌ی

جدول ۱. آنالیز سیگنال ژن SQS با استفاده از signal P

Protein type	Signal peptide (sec/SPI)	Other
Likelihood	0.0013	0.9987

SignalP-5.0 prediction (Eukarya): Sequence



شکل ۲. آنالیز سیگنال ژن SQS با استفاده از signal P

آنالیز رهگیری محل استقرار نهایی پروتئین

آنالیز توالی پروتئین با برنامه TargetP نشان می‌دهد که احتمال انتقال پروتئین SQS به میتوکندری، کلروپلاست و مسیر ترشحی بسیار پایین است. بنابراین استنباط می‌شود که این پروتئین به میتوکندری منتقل نمی‌شود و این نتیجه یک بار دیگر تأییدکننده نتایج قبلی توسط برنامه‌ی Signal P مبنی بر داخل سلولی بودن پروتئین و نیز استقرار آن در هسته سلول می‌باشد (جدول ۲).

جدول ۲. آنالیز رهگیری محل استقرار نهایی پروتئین با استفاده از

Protein type	Other	Signal peptide	Mitochondrial transfer peptide	Chloroplast transfer peptide	Thylakoid luminal transfer peptide
Likelihood	0.9999	0.0001	0	0	0

شناسایی موتیف‌ها و دمین‌های حفاظت شده

به منظور شناسایی موتیف‌ها و دمین‌های حفاظت شده در پروتئین از برنامه جامع InterProScan استفاده شد. نتیجه‌ی آنالیز با این برنامه نشان داد که پروتئین SQS دارای سه دمین حفاظت‌شده‌ی Squalene/phytoene synthase (IPR002060) (تبدیل دو مولکول فarnesیل دی فسفات (FPP) به اسکوالن را کاتالیز می‌کند)، Trans-isoprenyl diphosphate synthases, eukaryotic-HH) type (IPR044844) (کاتالیز یک واکنش تراکم سر به سر (farnesyl-) Squalene synthase-like_1 (IPR006449) و (diphosphate farnesyltransferase activity) می‌باشد.

ساختار دوم پروتئین

در این تحقیق از نرم افزار Phyre2 برای بررسی ساختار ثانویه پروتئین SQS استفاده شد. نتایج نشان می‌دهد که شش ساختار مشابه در پایگاه داده پروتئین برای SQS یافت شد، یکی از این ساختارها، به نام ساختار کریستالی d1ezfa، شباهت ۹۲٪ با پروتئین SQS داشت (شکل ۳ الف). همچنین با نرم‌افزار GORIV مشخص شد که پروتئین 6-SFT حاوی حدود ۹۲٪ مارپیچ آلفا هلیکس است و این اطلاعات به منظور بررسی پایداری پروتئین دارای اهمیت می‌باشد (شکل ۸ ب). معمول‌ترین مکان برای مارپیچ آلفا سطح هسته‌های پروتئین

است که یک رابط را برای تعامل با محیط آبی بیرون ایجاد می‌کند. قسمت داخلی مارپیچ تمایل به داشتن آمینواسیدهای آب‌گریز و قسمت بیرون تمایل به داشتن آمینواسیدهای آب‌دوست دارد، بنابراین در طول مارپیچ از هر چهار آمینواسید، سه آمینواسید آب‌گریز خواهند بود. سایر آمینواسیدهای موجود در هسته پروتئین و یا داخل غشای سلولی خاصیت آب‌گریزی دارند. به طور کلی مارپیچ‌های قرار گرفته در سطح، تعداد کمتری آمینواسید آب‌گریز دارند. از این ویژگی می‌توان در پیش‌بینی ساختار پروتئین‌ها کمک گرفت. برای مثال نواحی با مقادیر بیشتر از آلانین، گلوتامین اسید، لوسین و متیونین و مقادیر کمتر از پرولین، گلیسین، تیروزین و سرین تمایل به تشکیل مارپیچ آلفا دارند. ساختار صفحه‌های بتا، ساختار دوم بسیار کشیده و چین‌دار می‌باشد. یکی از تفاوت‌های مهم صفحه‌های بتا با مارپیچ آلفا این است که اسیدهای آمینه‌ای که معمولاً در ساختار اول زنجیره پروتئینی با فاصله زیاد از هم قرار گرفته‌اند، برای تشکیل این ساختار در مجاورت یکدیگر قرار می‌گیرند، بنابراین صفحه‌های بتا تمایل به سختی داشته و انعطاف‌پذیری ناچیزی دارند. پیوندهای هیدروژنی بین‌رشته‌ای که میان گروه‌های CO یک رشته بتا و NH رشته‌ی بتای مجاور ایجاد می‌شوند، به صفحات بتا پایداری می‌بخشند و باعث می‌شوند که این صفحات ظاهری زیگزاگ داشته باشند. هر چه مارپیچ آلفا و بتا بیشتر باشد انعطاف‌پذیری پروتئین کمتر و پایداری آن بیشتر است. نتایج نشان داد پایداری پروتئین SQS زیاد نیست (شکل ۳ ب) (موراکامی و همکاران، ۲۰۰۵).

ساختار سوم پروتئین

مدلسازی ساختار سه بعدی بر اساس انتخاب الگوی با شباهت زیاد به پروتئین هدف با استفاده از پایگاه داده مدل سوئیس انجام شد (۱۲) (شکل ۴). مدل انتخاب شده برای مدل سازی پروتئین SQS در انسانی (FARNESYL-DIPHOSPHATE) (1EZF) (FARNESYLTRANSFERASE) حاوی ۳۴۰ اسید آمینه است و توسط پراش اشعه ایکس با وضوح ۲/۱۵ آنگستروم کشف شده است. شناسایی الگوی 3wei.1.A با پروتئین هدف ۳۰/۷۸ است. نتایج این تحقیق می‌تواند در تحقیقات آتی و داکینگ مولکولی مورد استفاده قرار گیرد و اطلاعات اولیه برای بررسی سایر گیرنده‌های ایمنی ارائه شود. بر اساس خصوصیات مدل پیش‌بینی شده، می‌توان خصوصیات پروتئین مورد مطالعه را پیش‌بینی کرد.

GOR4 :

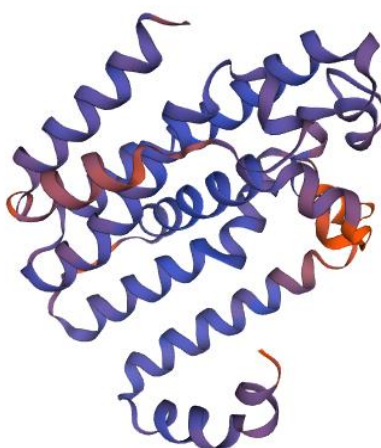
Alpha helix (Hh) :	92 is	38.02%
3 ₁₀ helix (Gg) :	0 is	0.00%
Pi helix (Ii) :	0 is	0.00%
Beta bridge (Bb) :	0 is	0.00%
Extended strand (Ee) :	67 is	27.69%
Beta turn (Tt) :	0 is	0.00%
Bend region (Ss) :	0 is	0.00%
Random coil (Cc) :	83 is	34.30%
Ambiguous states (?) :	0 is	0.00%
Other states :	0 is	0.00%

(الف)



(ب)

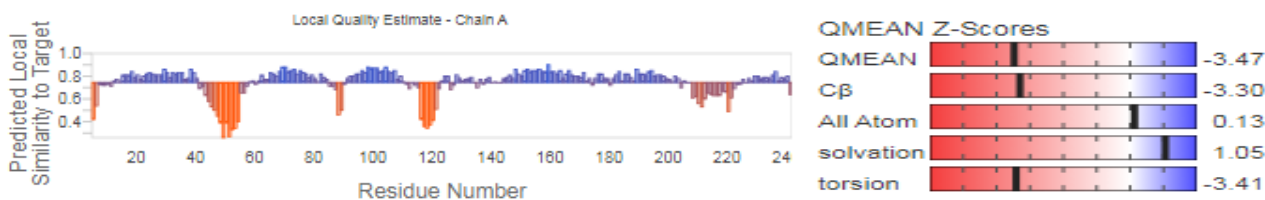
شکل ۳. الف: ساختار پیشنهادی برای پروتئین SQS با برنامه‌ی Phyre2 ب: آنالیز ساختار دوبعدی پروتئین SQS با نرم افزار GORIV



شکل ۴. ساختار مدل سازی شده SQS در Swiss Model در گیاه *Gymnema sylvestre*

جدول ۳. میزان همسانی اسیدآمینهای پروتئین هدف با الگو

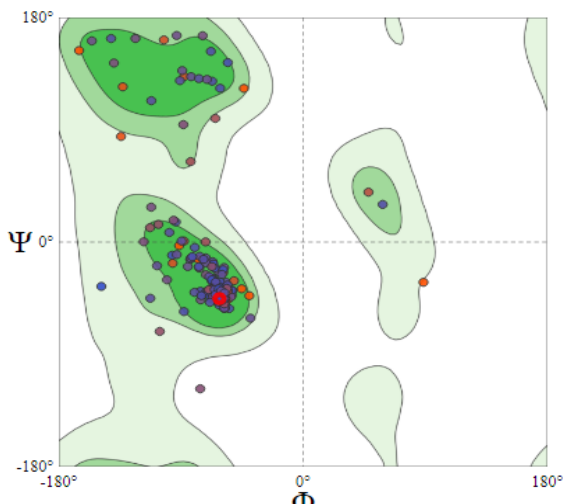
Template	Seq Identity	Oligo-state	Found by	Method	Resolution	Sq Similarity	Range	Coverage	Description
3wei.1.A	۳۰/۷۸%	Monomer	BLAST	Crystal structure	۲/۱۵Å	۰/۳۳	۱-۳۸۹	۹۷%	mutant in complex with presqualene pyrophosphate



شکل ۵. نمودار ارزیابی کیفی موضعی ساختار و نمودار

پروتئین‌های مدل سازی شده در شکل ۴ نشان داده شده است. ارزیابی کیفی مدل پیش‌بینی شده نیز به منظور بررسی کیفیت مدل انجام شد و مشخص شد مدل پیش‌بینی شده از کیفیت خوبی برخوردار است (شکل ۵).

اولین گام در انجام مطالعات بعدی جهت مطالعه‌ی جایگاه اتصال و یا لیگاند‌های مهاری و یا فعال‌کننده، قطعاً مدل سازی این پروتئین می‌باشد. با استفاده از مدل سازی انجام شده برای پروتئین SQS می‌توان خصوصیات آن را پیش‌بینی کرد. ساختار سه بعدی



شکل ۶. نمودار رامچاندردان ساختار SQS مدل‌سازی شده در گیاه *Gymnema sylvestre*. نواحی سبز پر رنگ به معنی قرارگیری ایده آل اسیدهای آمینه در نواحی سبز کم رنگ به معنی نواحی مجاز، و نواحی آبی کم رنگ نواحی غیر مجاز می‌باشند.

به‌علت سختی و پر هزینه بودن فرآیند کریستالوگرافی به‌خصوص در مورد پروتئین‌های مهمی مانند SQS که نقش‌های حیاتی در مسیرهای شناخته شده دارند، پیشگویی ساختار آن از طریق ابزار *In silico* به منزله‌ی میانبری برای مطالعات بعدی و بررسی جزئیات ساختاری آن خواهد بود. به‌عبارت دیگر طراحی یک لیگاند مؤثر برای فعال‌سازی یا مهار پروتئین در یک مسیر خاص، با شبیه‌سازی ساختار سوم آن پروتئین مقرون به صرفه بوده و روند مطالعات را تسریع می‌کند. با توجه به نتایج ارزیابی کیفی مدل‌های ایجاد شده، می‌توان با ضریب اطمینان بالایی از این مدل‌ها در آنالیزها و طراحی‌های بعدی استفاده نمود. نتایج این مقاله می‌تواند در مطالعه‌ی رفتار و کنش پروتئین SQS در گیاهان مختلف می‌تواند بسیار کمک‌کننده باشد و به فهم چگونگی نحوه‌ی برهمکنش این پروتئین با سایر پروتئین‌ها کمک کرده و زمینه‌ساز روشن شدن نحوه‌ی فعالیت آن در سلول‌ها باشد (لابادورف، لینک و راجرز، ۲۰۱۰، میسومی، یوشیدا و نیشیدا، ۲۰۱۱).

پیشنهاد

با توجه به اهمیت نقش ژن SQS در این تحقیق به بررسی بیوانفورماتیکی این ژن پرداخته شد. نتایج این تحقیق می‌تواند در مطالعات بعدی مانند بررسی برهمکنش بین پروتئین‌ها و داکینگ مولکولی استفاده کرد.

ارزیابی مدل‌های پیشنهاد شده در Swiss Model بر اساس پارامترهای QMEAN و GMQE می‌باشد. تابع QMEAN4 یک تابع حسابی مرکب برای ارزیابی کیفیت global (کل ساختار) و local (هر اسید آمینه) ساختار مدل‌سازی شده می‌باشد. نمودار کیفیت Local برای هر اسید آمینه در مدل (محور X) شباهت مورد انتظار به ساختار بومی (محور Y) را نشان می‌دهد. به طور معمول، اسید آمینه‌هایی که امتیاز زیر ۰/۶ را نشان بدهند، از کیفیت پایینی برخوردارند. تابع QMEAN4 شامل چهار توصیف ساختاری پتانسیل زاویه‌ی چرخش، پتانسیل فاصله‌ی اتم‌ها، اثر متقابل کربن بتا و پتانسیل حلال‌پوشی می‌باشد. GMQE یک ارزیابی کیفی است که خصوصیات هم‌ترازی مدل الگو را ترکیب می‌کند. نمودارهای ارزیابی کیفی موضعی ساختار و میزان Z-score در مورد تابع حسابی مرکب QMEAN4، تمامی اتم‌ها، کربن بتا، حلالیت و زاویه‌ی چرخشی در شکل ۵ نشان داده شده است (شکل ۵). ارزیابی مدل به منظور بررسی کیفیت مدل پیش‌بینی است و هر چه کیفیت مدل پیش‌بینی شده بالاتر باشد، خصوصیات آن به خصوصیت پروتئین مورد بررسی نزدیک‌تر است.

ارزیابی کیفیت ساختار مدل شده پس از بهینه‌سازی انرژی به وسیله نمودار رامچاندردان با استفاده از ابزار Rampage صورت گرفت (اشوید و همکاران، ۲۰۰۳). نمودار رامچاندردان یک روش بسیار پرکاربرد جهت پلات زاویه‌ی چرخشی پروتئین و ارزیابی کیفیت استرئوشیمیایی ساختار مدل‌سازی شده می‌باشد.

نمودار رامچاندردان با محاسبه درصد اسیدآمینه‌های قرار گرفته در نواحی مطلوب، مجاز و یا غیرمجاز کیفیت مدل ورودی را تعیین می‌نماید. چنانچه در شکل ۶ دیده می‌شود ۹۹/۲٪، ۰/۸٪ و ۵۰٪ اسید آمینه برای ساختار مدل شده در گیاه *Gymnema sylvestre* به ترتیب در مناطق مطلوب، مجاز و غیر مجاز قرار دارند و این بدان معنی است که هر مدل از کیفیت بالایی برخوردار است (شکل ۶). هرچه درصد اسیدهای آمینه در مناطق مطلوب بیشتر باشد، آن مدل کیفیت بهتری دارد و مدل‌سازی به درستی انجام شده است.

در روش مشابهت مدلینگ یا مدل‌سازی مقایسه‌ای، ساختمان پروتئین بر اساس مشابهت توالی با ساختمان‌های شناخته شده با روش‌های تجربی پیش‌بینی می‌شود. در واقع این روش بر این اصل استوار است که اگر دو پروتئین مشابهت توالی بالایی داشته باشند، احتمالاً ساختمان سه بعدی بسیار مشابهی دارند. ساختار سه بعدی پروتئین منبع مهم اطلاعاتی برای درک بهتر عملکرد پروتئین و برهمکنش آن با اجزای دیگر (لیگاند، پروتئین و ...) می‌باشد.

References

- Ačimovič, J., & Rozman, D. (2013). Steroidal triterpenes of cholesterol synthesis, *Molecules*, *18*, 4002-4017.
- Brown, AK. Sridharan, S. Kremer, L. Lindenberg, S. Dover, LG. Sacchettini, JC. & Besra, GS. (2005). Probing the mechanism of the Mycobacterium tuberculosis beta-ketoacyl-acyl carrier protein synthase III mtFabH: factors influencing catalysis and substrate specificity. *Journal of Biological Chemistry*, *18*, 54-63.
- Burge, C., & Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology*, *268*(1), 78-94
- Cao, H. (2011). Structure-function analysis of diacyl-glycerol acyltransferase sequences from 70 organism. *BMC research notes*, *4*, 249-279.
- Fallah Ziarani, M. & Safaeizadeh, M. (2022). In Silico Analysis of Sucrose: Fructan 6- Fructose Transferase Gene in *Triticum aestivum* L. *Crop Biotechnology*, *11*(4), 97-114.
- Fallah Ziarani, M., M. Tohidfar & Aminfar, Z. (2017). Bioinformatic analysis of Acyl Carrier Protein (ACP) in eukaryotes and prokaryotes. *Crop Biotechnology*, *17*, 15-29.
- Fallah Ziarani, M., M. Tohidfar, Mirjalili, M & Ahmadi Gavlighi, H. (2021) A Bioinformatics Analysis of Antibacterial Peptides in Five Species of Prokaryote and Eukaryote and the Evaluation of Antibacterial Effects of Nisin on Gram-positive and Gram-negative Bacteria. *Biological Journal of Microorganisms*, *10* (39), 28-50.
- Ferguson, AA. Jiang, N. & Pack, M. (2011). Recycling and reshaping genes through GC-biased acquisition. *Mobile Genetic Elements*, *15*, 2134-2141.
- Labadorf, A. Link, A. & Rogers, MF. (2010). Genome-wide analysis of alternative splicing in chlamydomonas reinhardtii. *BMC Genomics*, *11*, 114-124.
- Lovell, SC. Davis, IW. Arendall, WB. de Bakker, PI. Word, JM. Prisant, MG. & Richardson, DC. (2003). Structure validation by C α geometry: ϕ , ψ and C β deviation. *Proteins: Structure, Function, and Bioinformatics*, *50* (3), 437-450.
- Matsuura, H., Watanabe, M. M., & Kaya, K. (2012) Squalene quantification using octadecyl benzene as the internal standard," *Procedia Environmental Sciences*, *15*, 43-46.
- Mortazavi, M., Zarenezhad, M., Alavian, S. M., Gholamzadeh, S., Malekpour, A., Ghorbani, M., et al., 2016 Bioinformatic Analysis of Codon Usage and Phylogenetic Relationships in Different Genotypes of the Hepatitis C Virus, *Hepatitis Monthly*, *16*, 1-8.
- Misumi, C. Yoshida, Y. & Nishida, K. (2011). Genome analyses and its significance in four unicellular algae and Thalassiosira pseudonana. *J. Plant. Res*, *121*, 3-17.
- Murakami, M. Matsushika, A. Ashikari, M. Yamashino, T. & Mizuno, T. (2005). Circadian-associated rice pseudo response regulators (OsPRRs): insight into the control of flowering time. *Biosci. Biotechnol. Biochem*, *69*, 410- 414.
- Schwede, T. Kopp, J. Guex, N. & Peitsch, MC. (2003). SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research*, *31*, 3381-3385.