

In silico analysis of chalcone synthetase (CHS) in *Cannabis sativa*

Narges Nemati¹, Masoumeh Fallah Ziarani² 

1. M.Sc. Biotechnology and Molecular Genetic, College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran.

2. Ph.D. in Plant Biotechnology, Faculty of Life Science and Biotechnology, Shahid Beheshti University, Tehran, Iran.

Correspondence:
Masoumeh Fallah Ziarani
Email: masoumeh_fallah@yahoo.com

Received: 3, Jan. 2025
Accepted: 14, Mar. 2025

How to cite:

Nemati, N., & Fallah Ziarani, M. (2025). *In silico* analysis of chalcone synthetase (CHS) in *Cannabis sativa*. *Crop Biotechnology*, 14 (3), 43-52. (DOI: [10.30473/cb.2025.73286.1997](https://doi.org/10.30473/cb.2025.73286.1997))

ABSTRACT

The chalcone synthetase (CHS) gene is a key gene in the flavonoid biosynthesis pathway in plants. This gene encodes the enzyme chalcone synthase, which plays a crucial role in the early stages of the flavonoid biosynthesis pathway. In addition to its primary role in flavonoid biosynthesis, the CHS gene may also be involved in other metabolic pathways, such as the production of compounds that are precursors to cannabinoids (like THC and CBD). Due to the importance of this gene in *Cannabis sativa*, it was studied using various bioinformatics databases and software tools. A search of the nucleotide sequence of accession NM_001397940 from *Cannabis sativa* in the NCBI database revealed 100 matches. The highest Max Score (805) belonged to the alignment with NP_001384869.1 from the *Cannabis* plant itself, with 100% coverage and 100% similarity. The lowest Max Score (746) corresponded to the alignment with QRV61377.1 from the species *Litchi Chinensis* with 99% coverage and 92.01% similarity. The E-Value in this section was also used to determine the significance of the matches. Protein sequence alignment with accession number NP_001384869.1 from *Cannabis sativa* in the NCBI database also showed 100 matches. Phylogenetic analysis of the top 20 species that showed homology with the CHS gene in the BLAST n analysis created a dendrogram, classifying the species into six distinct groups. Genetic diversity analysis using MEGA 6 software indicated that approximately 31% of the CHS gene sequence has undergone substitution mutations and only 8% of these regions contain informative polymorphisms. Analysis of the protein's primary structure revealed that its isoelectric point (PI) is 6.04, and instability index is 37.13, indicating high stability of the protein under cellular conditions. Stable proteins are capable of withstanding environmental changes such as temperature, pH, and pressure, allowing them to maintain their function under stressful or harsh conditions. The amino acid composition analysis revealed that leucine (11%) had the highest abundance, while Pyrrolysine and Selenocystein (0%) were the least abundant. The GRAVY index was -0.09, indicating the hydrophilic nature of the protein. Secondary structure analysis of the protein, using Phyre2 software, identified six similar structures in the NCBI database. Furthermore, GORIV software analysis showed that the CHS protein contains approximately 40.8% alpha-helical structures which it helps the protein's stability. The 3D modeling of this protein in *Cannabis sativa* was performed using the PDB database which had a high similarity to the Chalcone Synthase protein from *Medicago sativa*. Studying the protein's structure can help in understanding its function, and examining its structural details may be useful in studies related to the active site of the protein and docking.

KEYWORDS

Chalcone synthase gene, Tetrahydrocannabinol, Cannabidiol, Cannabis Sativa, Flavonoid biosynthesis pathway.



مقاله پژوهشی»

آنالیز *in silico* ژن چالکون سنتاز (CHS) در گیاه *Cannabis sativa*نرگس نعمتی^۱، معصومه فلاح زيارانی^۲ 

۱. کارشناسی ارشد، گروه بیوتکنولوژی و ژنتیک مولکولی، دانشکده‌گان کشاورزی و منابع طبیعی، دانشگاه تهران، کرج، ایران.
۲. دکترای بیوتکنولوژی گیاهی، دانشگاه شهید بهشتی، تهران، ایران.

نویسنده مسئول:

معصومه فلاح زيارانی

ایمانامه: masoumeh_fallah@yahoo.com

تاریخ دریافت: ۱۴۰۳/۱۰/۱۴

تاریخ پذیرش: ۱۴۰۳/۱۲/۲۴

استناد به این مقاله:

فلاح زيارانی، معصومه و نعمتی، نرگس (۱۴۰۴). آنالیز *in silico* ژن چالکون سنتاز (CHS) در گیاه شاهدانه، فصلنامه علمی زیست فناوری گیاهان زراعی، ۱۴ (۳)، ۴۳-۵۲.

(DOI: [10.30473/cb.2025.73286.1997](https://doi.org/10.30473/cb.2025.73286.1997))

چکیده

ژن چالکون سنتاز (CHS) یکی از ژن های کلیدی در مسیر بیوسنتز فلاونوئیدها در گیاهان است. این ژن کد کننده آنزیم چالکون سنتاز می باشد که در مرحله ابتدایی مسیر بیوسنتز فلاونوئیدها نقش مهمی دارد. ژن CHS علاوه بر نقش اصلی خود در بیوسنتز فلاونوئیدها ممکن است در مسیرهای متابولیکی دیگری مانند تولید ترکیباتی که پیش ساز کاناบินوئیدها (مانند THC و CBD) هستند نیز دخالت داشته باشند. به دلیل اهمیت این ژن در گیاه شاهدانه این ژن به کمک پایگاه های مختلف اطلاعاتی و نرم افزارهای گوناگون بیوانفورماتیک مطالعه گردید. جست و جوی توالی نوکلئوتید شماره دسترسی NM_001397940 گیاه شاهدانه در پایگاه اطلاعاتی NCBI توانست ۱۰۰ تطابق را نشان دهد. همردیفی توالی پروتیینی با شماره دسترسی NP_001384869.1 در گیاه شاهدانه مقابل پایگاه اطلاعاتی NCBI توانست ۱۰۰ تطابق را نشان دهد. بیشترین امتیاز (805) Max Score متعلق به همردیفی با NP_001384869.1 از خود گیاه شاهدانه با ۱۰۰٪ پوشش و ۱۰۰٪ شباهت بود. کمترین (746) Max Score مربوط به همردیفی با QRV61377.1 از گونه *Litchi chinensis* با ۹۹٪ پوشش و ۹۲،۰۱٪ شباهت بود. مقدار E-Value در این بخش نیز برای تشخیص معناداری تطابق ها مورد استفاده قرار گرفت. آنالیز فیلوژنتیکی ۲۰ گونه اول که در آنالیز BLAST n با ژن CHS همولوژی نشان داده بود دندرومی ایجاد کرد که گونه های مورد مطالعه را در ۶ گروه مجزا دسته بندی کرد. آنالیز تنوع ژنتیکی توسط نرم افزار MEGA 6 نشان داد که در حدود ۳۱ درصد از طول توالی ژن CHS جهش های نوع جایگزینی رخ داده است و تنها ۸ درصد از این نواحی حاوی چند شکلی های آگاهی بخش هستند. با بررسی خصوصیات ساختار اولیه پروتیین مشخص شد که نقطه ایزوالکتریک (Pi) آن ۶،۰۴ و شاخص ناپایداری ۳۷،۱۳ که حاکی از پایداری بالای پروتیین در شرایط سلول است. پروتیین های پایدار قادرند در برابر تغییرات محیطی مانند دما PH و فشار مقاومت کنند. این ویژگی به آن ها امکان می دهد که عملکرد خود را در شرایط سخت یا تنش را حفظ کنند. ترکیب اسید آمینه های نشان داد لوسین (۱۱٪) بیشترین و پیرولیزین و سلسوسیتین (۰٪) کمترین فراوانی را دارند. شاخص GRAVY مقدار ۰،۰۹- را نشان داد که بیانگر ماهیت هیدروفیلیک پروتیین است. بررسی ساختار ثانویه پروتیین با استفاده از نرم افزار 6Phyre2 ساختار مشابه در دیتابیس NCBI برای این پروتیین پیدا شد. همچنین با نرم افزار GORIV مشخص شد که پروتیین CHS حاوی حدود ۴۰،۸٪ ماریچ آلفا هلیکس می باشد که به پایداری آن کمک می کند. مدل سازی سه بعدی این پروتیین در شاهدانه با استفاده از پایگاه PDB انجام گرفت که شباهت بالایی با پروتیین Chalcone synthase از *Medicago Sativa* داشت. مطالعه ساختار پروتیین می تواند به درک عملکرد پروتیین کمک کند و بررسی جزئیات ساختار آن می تواند در مطالعات جایگاه فعال پروتیین و داکینگ سودمند باشد.

واژه های کلیدی

ژن چالکون سنتاز (CHS)، تترا هیدروکانابینول (THC)، کانابیندیول (CBD)، گیاه شاهدانه (*Cannabis sativa*)، مسیر بیوسنتز فلاونوئیدها.

مقدمه

شاهدانه (*Cannabis sativa*) گونه‌ای از گیاهان است که تاریخچه‌ای غنی و پیچیده از استفاده انسانی، هم به صورت تفریحی و هم به صورت دارویی دارد. هزاران سال است که انسان‌ها این گیاه را به خاطر ویژگی‌های منحصر به فردش پرورش داده و استفاده کرده‌اند از الیاف مقاوم آن گرفته تا ترکیبات روان گردان قوی این گیاه (دهنوی و همکاران، ۲۰۲۲). در سال‌های اخیر علاقه زیادی به درک پایه ژنتیکی بیوشیمی شاهدانه به ویژه تولید کannabinoidهایی مانند THC و CBD به وجود آمده است. این ترکیبات دارای طیف گسترده‌ای از خواص درمانی از تسکین درد تا ضد التهاب هستند و توجه زیادی از سوی جوامع پزشکی و داروسازی به خود جلب کرده‌اند (دهنوی و همکاران، ۲۰۲۲).

اما چگونه گیاه شاهدانه این ترکیبات منحصر به فرد را تولید می‌کند؟ و چه ژن‌هایی در این فرایند دخیل هستند؟ یکی از ژن‌های کلیدی که در بیوسنتز کannabinoidها نقش دارد ژن چالکون سنتاز (CHS) است. چالکون سنتاز یک آنزیم است که نقش حیاتی در مسیر بیوسنتز فلاونوئیدها در گیاهان ایفا می‌کند. این مسیر مسئول تولید طیف وسیعی از ترکیبات است که به گیاهان رنگ می‌بخشد، آن‌ها را از نور فرابنفش محافظت می‌کنند، و دارای خواص آنتی اکسیدانی هستند (دائو و همکاران، ۲۰۱۱). بر اساس تحقیقات گزارش شده ژن CHS متعلق به ابر خانواده سنتاز پلی کتید (PKS) در گیاهان می‌باشد. همچنین، بیان ژن CHS تحت تاثیر عوامل محیطی مختلف مانند شدت نور، دما، تنش‌های زیستی و غیر زیستی قرار دارد و در مراحل مختلف رشد گیاه، بیان آن در بافت‌های مختلف متغیر است. تنوع و چند شکلی این ژن در گونه‌های مختلف، نشان دهنده‌ی تطبیق پذیری آن با شرایط محیطی متنوع است (وانگ و یو، ۲۰۱۸).

ژن CHS نه تنها در سنتز آنتوسیانین‌ها نقش دارد، بلکه به عنوان یک تنظیم کننده منفی درونی در انتقال اکسین عمل می‌کند و بر رشد اندام‌های گل تاثیر می‌گذارد. علاوه بر این، بیان اعضای خانواده ژنی CHS

به صورت بافتی-ویژه است و سطح بیان آن‌ها در مراحل مختلف رشد گیاه متفاوت می‌باشد (لیو و همکاران، ۲۰۲۵). اکثر گیاهان حداقل یک ژن CHS را در ژنوم خود دارند. در برنج (*Oryza sativa*) ۲۷ عضو، در ذرت (*Zea mays*) ۱۴ عضو و در توت سیاه (*Morus atropurpurea* Roxb.) ۵ عضو از این ژن شناسایی شده است (سو و همکاران، ۲۰۲۵).

در گیاه شاهدانه آنزیم CHS در تولید اسید اولیووتولیک که پیش ساز THC و CBD است نقش دارد. با کاتالیز واکنش ترکیبی malonyl-CoA و coumaroyl-CoA زنجیره‌ای از واکنش‌ها را به حرکت در می‌آورد که در نهایت منجر به تولید این کannabinoidهای شناخته شده می‌شود (حسین و همکاران، ۲۰۲۳). با این حال درک عملکرد و تنظیم ژن CHS در گیاه شاهدانه نیازمند بیش از تنها توالی‌یابی ژن است. یکی از روش‌های موثر تحلیل به کمک رایانه (In Silico) است که از ابزارهای محاسباتی و تکنیک‌های بیوانفورماتیک برای تفسیر داده‌های بیولوژیکی از جمله توالی‌های DNA و پروتیین استفاده می‌کند. ویژگی‌های ساختاری نقش‌های عملکردی و تاریخچه تکاملی آن را بررسی می‌کند. این تحلیل می‌تواند شامل تکنیک‌های مختلفی از جمله هم ردیفی توالی‌ها و آنالیز عملکردی تا مدل‌سازی ساختاری پیشرفته و مطالعات فیلوژنتیک باشد.

در این مقاله به تحلیل به کمک ابزارهای بیوانفورماتیکی (In Silico) ژن CHS در گیاه شاهدانه خواهیم پرداخت و به بررسی ابزارها و تکنیک‌های پیشرفته محاسباتی که برای کشف پیچیدگی‌های بیوسنتز کannabinoidها استفاده می‌شوند خواهیم پرداخت. ما به بررسی توالی ژن CHS تاریخچه تکاملی آن ساختار عملکرد پیش بینی شده‌اش و نحوه تنظیم آن خواهیم پرداخت. با انجام این کار درک عمیق‌تری از نحوه مشارکت این ژن کلیدی در تولید کannabinoidها در شاهدانه و چگونگی بهره‌برداری از آن در آینده به دست خواهیم آورد.

روش‌شناسی پژوهش

BLAST P و BLAST n

BLAST p و BLAST n در جهت جستجوی توالی مشابه با توالی ژنی و پروتئینی ژن چالکون سنتاز (CHS) گیاه شاهدانه به عنوان query انجام شد.

یافته‌های پژوهش و نتیجه‌گیری

نتایج BLAST n

جستجوی توالی نوکلوتیدی شماره دسترسی NM_001397940 گیاه شاهدانه در پایگاه اطلاعاتی NCBI توانست ۱۰۰ تطابق را نشان دهد که بالاترین امتیاز شاخص Max Score (2911) مربوط به تطابق با شماره دسترسی NM_001397940.1 از گونه ی گیاهی *Cannabis Sativa* با ۱۰۰٪ پوشش و با ۱۰۰٪ شباهت بود در حالی که کمترین امتیاز شاخص Max Score (۷۲۳) مربوط به تطابق با اکسشن CP039336.1 در گونه *Cicer arietinum* با ۶۲٪ پوشش و ۷۹,۹۸٪ شباهت بود. رای تعیین بهترین تطابق، از مقدار E-Value نیز استفاده شد که مقدار کمتر آن نشان‌دهنده انطباق با احتمال بالاتر و معناداری بیشتر است. مقدار E-Value کمتر از ۰,۰۱ نشان‌دهنده معنی‌داری تطابق است.

نتایج BLAST p

همردیفی توالی پروتئینی با شماره دسترسی NP_001384869.1 در گیاه شاهدانه مقابل پایگاه اطلاعاتی NCBI توانست ۱۰۰ تطابق را نشان دهد که بالاترین امتیاز شاخص Max score (۸۰۵) مربوط به همردیفی با شماره دسترسی NP_001384869.1 گیاه شاهدانه با ۱۰۰ درصد پوشش و ۱۰۰ درصد شباهت بود؛ در حالی که کمترین شاخص Max score (۷۴۶) مربوط به همردیفی با شماره دسترسی QRV61377.1 گونه گیاهی *Litchi chinensis* با ۹۹ درصد پوشش و ۹۲,۰۱ درصد شباهت بود. برای تشخیص انطباق از مقدار E-value هم می‌توان استفاده کرد. مقدار E-Value نشان‌دهنده تعداد تطابق‌های مورد انتظار در یک جایگاه داده تصادفی با همان اندازه است. مقدار E-Value کوچکتر نشان‌دهنده امتیاز بالاتر و معتبر بودن انطباق با احتمال بالاتر است، در حالی که مقادیر بسیار کوچک، به ویژه کمتر از ۰,۰۱، نشان‌دهنده

آنالیزهای فیلوژنتیکی

جهت هم‌ردیفی توالی ژن مورد مطالعه از نرم افزار T-Coffee استفاده شد. همچنین برای رسم درخت فیلوژنتیک از نرم افزار MEGA6 استفاده گردید (فلاح زیارانی و همکاران، ۲۰۱۷؛ فلاح زیارانی و همکاران، ۲۰۱۳؛ فلاح زیارانی و توحیدفر، ۲۰۱۸).

بررسی خصوصیات ساختار اولیه پروتئین

به منظور بررسی خصوصیات ساختار اولیه پروتئین از برنامه PROTPARM استفاده شد. برای تعیین محل استقرار و نقش مولکولی پروتئین در سلول از پایگاه داده UNIPROT بهره گرفته شد.

بررسی ساختار دوم پروتئین

ساختار ثانویه پروتئین و تعیین موقعیت صفحات آلفا و بتا با استفاده از ابزار آنالیز Phyre2 شناسایی گردید (فلاح زیارانی و همکاران، ۲۰۱۷؛ فلاح زیارانی و همکاران، ۲۰۲۱؛ فلاح زیارانی و همکاران، ۲۰۲۲).

بررسی ساختار سه بعدی پروتئین

مدل‌سازی سه‌بعدی پروتئین با استفاده از روش همولوژی مدلینگ و پایگاه داده Swiss-Model انجام شد. الگوی مناسب برای مدل‌سازی، با دقت از پایگاه داده PDB و بر اساس بالاترین میزان شباهت ساختاری انتخاب گردید. به منظور اعتبارسنجی ساختار سه‌بعدی مدل ترسیم‌شده، نمودار رامچاندوران مورد استفاده قرار گرفت و زوایای دی‌هدرال برای بررسی صحت استریوشیمیایی محاسبه شد. تحلیل دقیق ساختار پروتئین به درک عملکرد مولکولی آن کمک شایانی می‌کند. همچنین مطالعه جزئیات ساختاری پروتئین، به‌ویژه

الگوریتم‌های دیگر کاربرد دارد. تاجیما D یکی از معروف‌ترین آزمون‌های فرضیه خنثی می‌باشد که با مقایسه دو نوع تنوع ژنتیکی در یک جمعیت (تفاوت‌های جفتی و تعداد جایگاه‌های چندشکلی) انجام می‌شود. این آزمون براساس مقایسه تنوع نوکلئوتیدی حاصل از فراوانی آللی جایگاه‌های چند شکلی است (فرر، ۲۰۰۵). به طور کلی مقادیر مختلف تاجیما D می‌تواند نشان دهنده فرآیندهای مختلفی باشد به این صورت که تاجیما D مثبت نشان دهنده تعادل انتخابی است و ممکن است بیانگر انتخاب متعادل یا گزینش در جمعیت باشد. تاجیما منفی نشان دهنده یک گزینش مثبت یا فشار انتخابی است که باعث شده یک آلل سودمند به سرعت در جمعیت گسترش یابد و تاجیما D نزدیک به صفر بیانگر این است که تغییرات ژنتیکی در جمعیت به صورت تصادفی و مطابق با فرضیه خنثی رخ داده‌اند. آزمون تاجیمای مثبت در ژن مورد مطالعه نشان داد که ژن‌هایی که تنوع فراوانی آللی آن‌ها بالا است تاجیما D مثبت (بزرگتر از صفر) دارند و با انتخاب متعادل در جمعیت همراه هستند (جدول ۱).

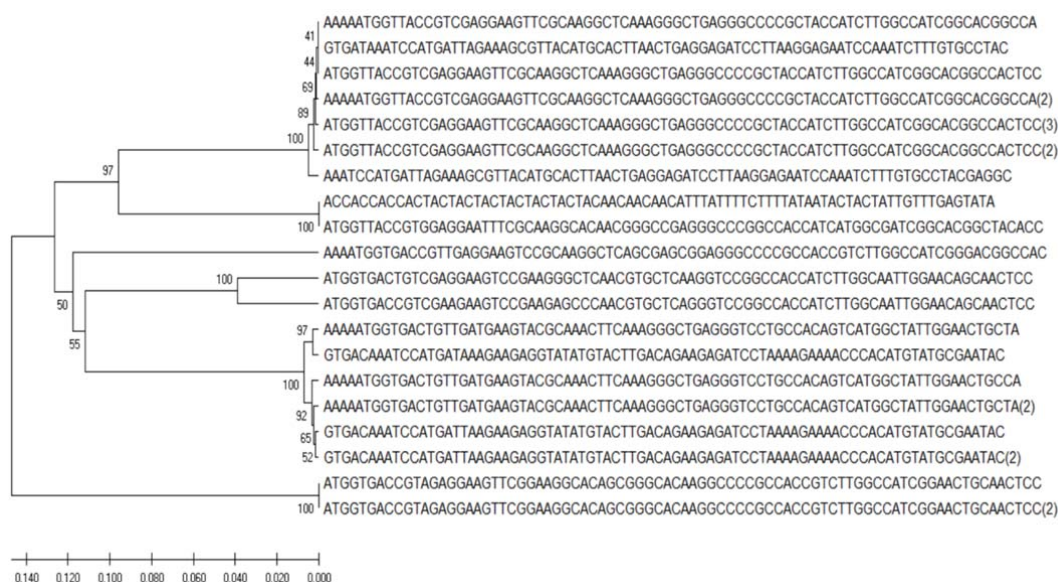
خصوصیات ساختار اولیه پروتئین و بررسی خواص فیزیکوشیمیایی

برای بررسی خواص فیزیکوشیمیایی پروتئین مورد بررسی از ساختار اولیه پروتئین و از برنامه PROTPARAM استفاده شد. نقطه ایزوالکتریک (PI) پروتئین مورد نظر ۶,۰۴ و بار خالص آن ۴- تعیین شد (زیرا تعداد اسیدآمینه‌های بار مثبت ۴۴ و تعداد اسیدآمینه‌های بار منفی ۴۹ می‌باشد). از نظر ترکیب اسیدآمینه بالاترین فراوانی (۱۱ درصد) متعلق به لوسین و کمترین فراوانی (صفر درصد) متعلق به دو اسید آمینه پیرولیزین و سلنوسیتئین می‌باشد. در نقطه ایزوالکتریک، یک اسید آمینه یا پروتئین به دلیل از دست دادن بار الکتریکی خالص خود، در یک میدان الکتریکی با جریان مستقیم، قادر به حرکت نیست. آگاهی از نقطه ایزوالکتریک پروتئین می‌تواند در انتخاب و بهینه‌سازی روش‌های خالص‌سازی مانند کروماتوگرافی تبادل یونی و ایزوالکترافورز فوکوسینگ بسیار مفید باشد.

انطباق‌های معنادار از نظر آماری هستند. به طور کلی هر چه E-Value به صفر نزدیک‌تر باشد، میزان اطمینان به ارتباط بیولوژیکی بین توالی‌های بیشتر خواهد بود.

آنالیزهای فیلوژنتیکی

برای این منظور ۲۰ گونه اول که در آنالیز BLAST n با ژن CHS همولوژی نشان دادند انتخاب شدند. در گام اول این گونه‌ها با نرم افزار T_Coffee هم‌ردیف‌سازی شده و سپس برای انجام آنالیزهای فیلوژنتیکی از نرم افزار MEGA6 استفاده شد. آزمون شاخص نابرابری که با استفاده از نرم افزار MEGA6 انجام شد نشان داد که ژن CHS همولوژی بالایی با هم گونه‌های مورد بررسی دارد. آنالیز فیلوژنتیکی با روش UPGMA و انتخاب مدل جایگزینی WAG+G دندروگرامی ایجاد کرد که گونه‌های مورد مطالعه را در ۶ گروه کاملاً مجزا دسته‌بندی کرد (شکل ۱). مدل جایگزینی توسط نرم افزار MEGA6 بدست آمد و با مقایسه تفاوت توالی‌ها و انتخاب بهترین مدل جایگزینی بهترین درخت فیلوژنتیکی رسم شد. در ماتریس تشابه در مقایسه بین ۲۰ گونه مورد بررسی کمترین تشابه مربوط به عدد ۳۱/۰ می‌باشد. آنالیز تنوع ژنتیکی توسط نرم افزار MEGA6 نشان داد که در حدود ۳۱ درصد از طول توالی ژن CHS جهش‌های نوع جایگزینی رخ داده که در این میان تنها حدود ۸ درصد از طول توالی ژن دارای چندشکلی‌های آگاهی بخش می‌باشد. تنوع نوکلئوتیدی در جمعیت مورد مطالعه پایین و در حدود ۹ درصد می‌باشد (جدول ۱). آزمون تاجیما نشان داد که جمعیت مورد (۲۰ گونه مورد بررسی) مطالعه در تعادل قرار دارند. تست فرضیه خنثی به بررسی اختلافات مشاهده شده بین جفت‌های توالی ژنتیکی که با الگوی جایگزینی مشابه تکامل یافته‌اند می‌پردازد. اساس این تست مقایسه فرکانس نوکلئوتید یا آمینواسید در توالی‌های جفتی و تعداد تفاوت‌های مشاهده شده بین این توالی‌ها است. در زیست‌شناسی مدل جایگزینی فرایندی را توصیف می‌کند که در آن تغییرات یک صفت در طی زمان تحت تاثیر فشارهای انتخابی یا فرایندهای تصادفی (مانند جهش یا رانش ژنتیکی) رخ می‌دهد. این مدل در تفسیر درخت تکاملی در فیلوژنتیک و شاخه بندی و شبیه‌سازی توالی با تست روش‌ها و



شکل ۱. دندروگرام بدست آمده از گونه‌های مورد مطالعه با روش UPGMA با استفاده از داده‌های توالی NM_001397940.

اعداد روی شاخه‌ها مقادیر بوت استرپ (%) را نشان می‌دهند. مقیاس افقی فاصله ژنتیکی را نمایش می‌دهد.

Figure 1. The dendrogram of the studied species was constructed using the UPGMA method based on the NM_001397940 sequence data. The numbers on the branches represent bootstrap values(%). The horizontal scale indicates genetic distance.

جدول ۱. نتیجه تست تاجیما برای ژن CHS در موجودات مورد بررسی. آزمون تاجیما مثبت نشان داد که ژن‌هایی با تنوع آلی بالا دارای مقدار D مثبت بوده و تحت انتخاب متعادل در جمعیت قرار دارند.

Table 1. Tajima test result for the CHS gene in the studied organisms. The positive Tajima test indicates that genes with high allelic diversity have a positive D value, and are subject to balancing selection in the population.

شخص D تاجیما	شخص Π	شخص Θ	Ps تنوع آلی	تعداد جایگاه های متغیر S	تعداد نمونه M
0.437801	0.098479	0.089140	0.316245	475	20

لوله آزمایش در برابر حرارت می‌باشد؛ و این به نوع اسیدآمینه های موجود بستگی دارد. پروتئین‌هایی که شاخص ناپایداری آنها کوچک‌تر از ۴۰ می‌باشد با ثبات و در غیر این صورت بی ثبات خواهند بود. بنابراین پروتئین CHS جز پروتئین‌های پایدار می‌باشد. پروتئین‌های پایدار در شرایط تنش مثل تنش حرارتی می‌تواند ساختار سوم خود و در نتیجه عملکردشان را در شرایط تنش حفظ کنند (کرسی و همکاران، ۲۰۲۰).

با برنامه PROTSKALE نیز مشخص شد که پروتئین مورد مطالعه طبق روش کیت و دولیتل در انتهای N و همچنین نزدیک به انتهای C خود دارای آب گریزی معنی داری می‌باشد (با توجه به خط برش ۱ در شکل ۲).

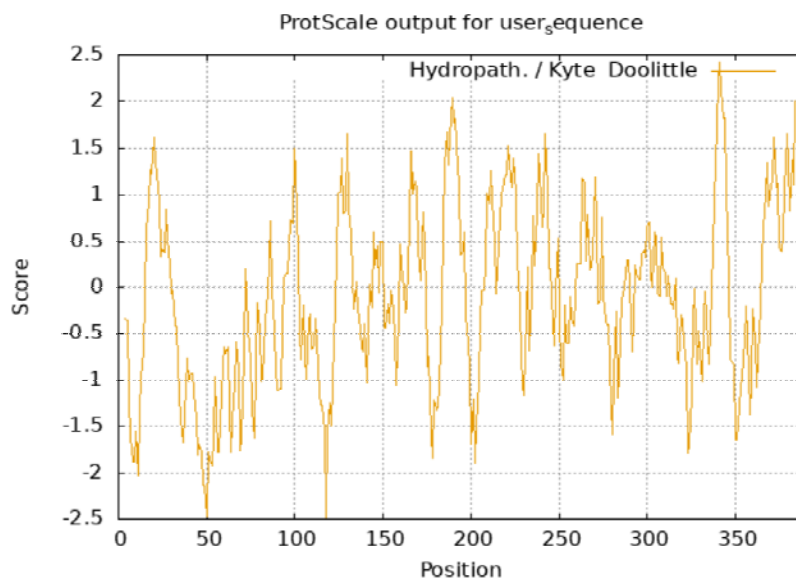
علاوه بر این، پروتئین‌هایی که در محدوده pH نقطه ایزوالکتریک خود قرار دارند، به دلیل تعادل بارهای مثبت و منفی، قابلیت نگهداری آب را از دست داده و به ناچار رسوب می‌کنند. از طرفی میانگین آب گریزی کل پروتئین مورد مطالعه (GRAVY) ۰/۰۹- می‌باشد. بررسی شاخص ناپایداری (Instability index) نشان می‌دهد که این پروتئین عمر بالایی دارد زیرا شاخص ناپایداری آن ۳۷،۱۳ است (جدول ۲)؛ این موضوع حاکی از آن است که این پروتئین در شرایط سلولی پایدار بوده و به راحتی تجزیه نمی‌شود.

شاخص ناپایداری پروتئین‌ها بیانگر میزان پایداری آنها در

جدول ۲. بررسی شاخص ناپایداری پروتئین CHS. مقدار شاخص ناپایداری ۳۷,۱۳ نشان دهنده پایداری نسبی پروتئین CHS در شرایط سلولی است. مقدار میانگین کلی آب گریزی منفی (-۰,۰۹۰) بیانگر ماهیت آب دوست بودن این پروتئین می باشد.

Table 2. CHS protein instability index review, the instability index of 37.13 indicates the relative stability of the CHS protein under cellular conditions. The negative GRAY value (-0.090) suggests the hydrophilic nature of this protein

شاخص ناپایداری Instability Index	شاخص آلیفاتیک Aliphatic Index	میانگین کل آبگریزی GRAVY
37.13	93.50	-۰,۰۹۰



شکل ۲. آنالیز آبگریزی پروتئین CHS با برنامه PROTSCALE. محور افقی موقعیت اسیدهای آمینه و محور عمودی میزان آبگریزی را نشان می دهد. به طوری که مقادیر مثبت نشان دهنده نواحی آبگریز هستند.

Figure 2. Hydropobicity analysis of CHS protein with PROTSKALE program. The X-axis shows amino acid positions, and the Y-axis represents the hydropobicity score, with positive values indicating hydrophobic regions

ساختار دوم پروتئین

با استفاده از نرم افزار آنالیز 2 Phyre، ساختار دوم پروتئین CHS پیش بینی شد. نتایج حاصل حاکی از آن است که ۶ ساختار مشابه در دیتابیس NCBI برای این پروتئین پیدا شد، یکی از این ساختارها مربوط به ساختار پروتئین Chalcone synthase از alfalfa است که دارای شباهت ۱۰۰ با پروتئین CHS گیاه شاهدانه می باشد (شکل ۳). همچنین با نرم افزار GORIV مشخص شد که پروتئین CHS حاوی حدود ۴۰,۸۷٪ مارپیچ آلفا هلیکس است و این اطلاعات به منظور بررسی پایداری پروتئین دارای اهمیت می باشد (شکل ۴).

مارپیچ آلفا معمولاً در سطح هسته های پروتئینی قرار دارد و به عنوان یک رابط برای تعامل با محیط آبی خارجی عمل می کند. قسمت داخلی مارپیچ آلفا تمایل به داشتن آمینواسیدهای آب گریز و قسمت بیرونی آن تمایل به داشتن آمینواسیدهای آب دوست دارد. به همین دلیل، در طول مارپیچ، از هر چهار آمینواسید، سه تای آنها آب گریز هستند. آمینواسیدهای موجود در هسته پروتئین یا داخل غشای سلولی نیز معمولاً خاصیت آب گریزی دارند. به طور کلی، مارپیچ های آلفایی که در سطح قرار دارند، تعداد کمتری آمینواسید آب گریز دارند. این ویژگی می تواند در پیش بینی ساختار پروتئین ها مفید باشد. به

می‌شود صفحات بتا سختی بیشتری داشته و انعطاف‌پذیری کمتری داشته باشند. پیوندهای هیدروژنی بین رشته‌ای که بین گروه‌های CO یک رشته بتا و گروه NH رشته بتای مجاور ایجاد می‌شوند، پایداری بیشتری به صفحات بتا می‌بخشند و ظاهری زیگزاگ به آن‌ها می‌دهند. هرچه میزان مارپیچ‌های آلفا و بتا در پروتئین بیشتر باشد، انعطاف‌پذیری پروتئین کمتر و پایداری آن بیشتر خواهد بود. نتایج نشان داد که پایداری پروتئین CHS نسبتاً زیاد است.

عنوان مثال، نواحی‌ای که حاوی مقادیر بالایی از آلانین، گلوتامیک اسید، لوسین و متیونین هستند و مقادیر کمتری از پرولین، گلیسین، تیروزین و سرین دارند، تمایل به تشکیل مارپیچ آلفا دارند. از سوی دیگر، ساختار صفحات بتا بسیار کشیده و چین‌دار است. یکی از تفاوت‌های مهم بین صفحات بتا و مارپیچ آلفا این است که آمینواسیدهایی که در ساختار اولیه پروتئین با فاصله زیادی از هم قرار دارند، در تشکیل صفحات بتا در مجاورت یکدیگر قرار می‌گیرند. این ویژگی باعث



شکل ۳. ساختار پیشنهادی برای پروتئین CHS با برنامه Phyre2

Figure 3. Proposed structure for the CHS protein with the Phyre2 program

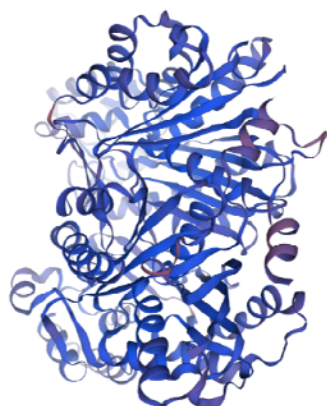
7/25/24, 9:55 AM

NPS@ : GOR4 SECONDARY STRUCTURE PREDICTION RESULTS

```
GOR4 :
Alpha helix (Hh) : 159 is 40.87%
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) : 59 is 15.17%
Beta turn (Tt) : 0 is 0.00%
Bend region (Ss) : 0 is 0.00%
Random coil (Cc) : 171 is 43.96%
Ambiguous states (?) : 0 is 0.00%
Other states : 0 is 0.00%
```

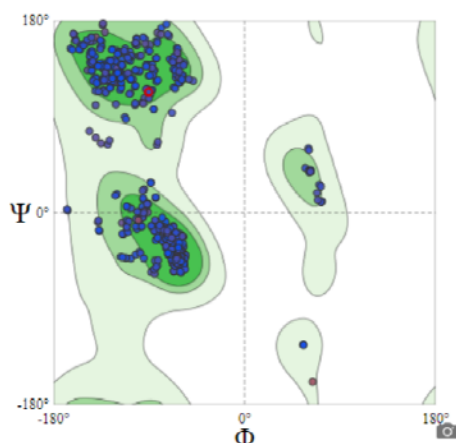
شکل ۴. آنالیز ساختار دو بعدی پروتئین CHS با نرم افزار GORIV

Figure 4. Analysis of the two-dimensional structure of CHS protein with GORIV software



شکل ۶. مدل سه بعدی پروتئین هدف با استفاده از Swiss Model براساس مدل سازی همولوژی ایجاد شده است. این مدل ساختار کلی، عناصر ثانویه و محل های احتمالی عملکردی پروتئین را نشان می دهد.

Figure 6. 3D structural model of the target protein was generated using Swiss-Model based on homology modeling. This model illustrates the overall structure, secondary elements, and potential functional sites of the protein.

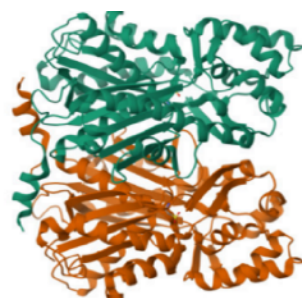


شکل ۷. نمودار رامچاندرا کیفیت مدل پیش بینی شده در Swiss Model را ارزیابی می کند و توزیع زوایای دی هیدرال (Psi و Phi) را در ساختار پروتئین نشان می دهد. حضور بیشتر اسیدهای آمینه در نواحی مجاز نشان دهنده پایداری ساختاری مدل است.

Figure 7. The Ramachandran plot evaluates the quality of the predicted model and illustrates the distribution of dihedral angles (Psi and Phi) in the protein structure. A higher concentration of amino acids in the allowed regions indicates structural stability of the model.

ساختار سوم پروتئین

مدل سازی ساختار سه بعدی براساس انتخاب یک الگوی با شباهت بالا به پروتئین هدف، با استفاده از پایگاه داده PDB انجام گرفت (شکل ۵). الگوی منتخب برای مدل سازی پروتئین CHS در گیاه شاهدانه (1JWX)، دارای ۳۸۹ آمینواسید است و در سال ۲۰۰۲ با قدرت تفکیک ۱٫۶۳ آنگستروم از گیاه *Medicago sativa* با استفاده از روش کریستالوگرافی اشعه ایکس (X-ray) کشف شده است. این الگو وزن مولکولی ۴۲٫۷۸۷ کیلو دالتون دارد و شامل دو جهش است. مدل سازی ساختار سه بعدی پروتئین CHS گیاه شاهدانه با استفاده از سرور Swiss-Model پیش بینی شد. ساختار براساس نزدیک ترین الگو ایجاد و با تحلیل های مختلف اعتبار سنجی شد (شکل ۶). نمودار رامچاندرا نشان داد که ۹۷٫۱۴٪ از اسیدهای آمینه در نواحی مطلوب قرار دارند و تنها ۰٫۳۹٪ به عنوان خارج از محدوده طبقه بندی شدند. امتیاز MolProbity برابر با ۰٫۹۷ بود که کیفیت بالای مدل را تایید می کند همچنین هیچ پیوند نامناسبی مشاهده نشد (شکل ۷). این مدل برای مطالعات بیشتر روی عملکرد پروتئین CHS و تجزیه و تحلیل های اتصال لیگاند مناسب است.



شکل ۵. مدل سه بعدی ساختار پروتئین CHS که با استفاده از روش همولوژی مدلینگ پیش بینی شده است. در این مدل، نواحی مختلف پروتئین با رنگ های متفاوت مشخص شده اند و ساختار کلی در قالب PBD نمایش داده شده است.

Figure 5. Three-dimensional modeled structure of CHS. predicted using homology modeling. Different regions of the protein are highlighted in distinct colors, and the overall structure is presented in PBD format.

تعارض منافع

هیچ‌گونه تعارض منافع توسط نویسندگان وجود ندارد.

References

- Dao, T.T.H., Linthorst, H.J.M., & Verpoorte, R. (2011). Chalcone synthase and its functions in plant resistance. *Phytochemistry Reviews*, 10, 397-412. DOI: 10.1007/s11101-011-9211-7.
- Fallah Ziarani, M., Masoud, T., Mirjalili, M. H. (2021b). Optimizing Culture Condition for High Frequency Regeneration *In Vitro* Tissue Culture in *Daucus carota* L. and Genetic Stability Assessment. *Journal of Plant Bioinformatics and Biotechnology*, 1, 84-96
- Fallah Ziarani, M., Tohidfar, M., & Mirjalili, M. H. (2022) Optimization of Nisin Gene Transfer by *Agrobacterium* in Carrot Plant. *Journal of Plant Research (Iranian Journal of Biology)*, 35 (1), 13-21.
- Fallah Ziarani, M., & Tohidfar, M. (2028) Genome editing for Change the color of the flower using crispr technology. *Crop Biotech*, 21, 71-79.
- Fallah Ziarani, M., Raheem, H., Ghasemali, G., & Mokhtar, J. (2013). *Agrobacterium*-mediated transformation of cotyledonary leaf of lettuce (*Lactuca sativa* L.) by the GCHI gene. *Iranian Journal of Genetics and Plant Breeding*, 2(2): 47-55.
- Fallah Ziarani, M., Tohidfar, M., & Aminfar, Z. (2017). Bioinformatic analysis of Acyl Carrier Protein (ACP) in eukaryotes and prokaryotes. *Crop Biotechnology*, 7(17) 15-29.
- Ferrer, J.L. et al. (1999). Three-dimensional structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature Structural Biology*, 6, 775-784.
- Hussain, S., Hussain, A., Ahmad, I., Wahid, F., & Sajid, M. (2023). Bioinformatics-based characterization of the Chalcone synthase (CHS) family genes in flowering plants. *SABRAO J. Breed. Genet*, 55(6), 1950-1962.
- Kearsey, L.J., Prandi, N., Karuppiiah, V., Yan, C., Leys, D., Toogood, H., & Scrutton, N.S. (2020). structure of the *Cannabis sativa* olivetol-producing enzyme reveals cyclization plasticity in type III polyketide synthases. *The FEBS Journal*, 287(8), 1511-1524.
- Mostafaei Dehnavi, M., Ebadi, A., Peirovi, A., Taylor, G., & Salami, S. A. (2022). THC and CBD fingerprinting of an elite cannabis collection from Iran: Quantifying diversity to underpin future cannabis breeding. *Plants*, 11(1), 129.
- Su, S., Xuan, X., Tan, J., Yu, Z., Jiao, Y., Zhang, Z., & Ramakrishnan, M. (2025). Analysis of the CHS Gene Family Reveals Its Functional Responses to Hormones, Salinity, and Drought Stress in Moso Bamboo (*Phyllostachys edulis*). *Plants*, 14(2), 161.
- Wang, Z., Yu, Q., Shen, W., El Mohtar, C. A., Zhao, X., & Gmitter, F. G. (2018). Functional study of CHS gene family members in citrus revealed a novel CHS gene affecting the production of flavonoids. *BMC Plant Biology*, 18, 1-13.
- Xie, Z., Yang, L., Fan, M., Xuan, S., Jia, X., Zhang, Z., ... & Li, J. (2025). Genome-wide identification, characterization and expression analysis of the chalcone synthase gene family in Chinese cabbage. *BMC genomics*, 26, 168.