

Evaluation of Bioinformatics Servers for Predicting Mutations Affecting Recombinant Protein Stability

Shabnam Khamoushi¹, Maryam Peymani^{2*}, Kheirollah Yari³, Jafar Amani⁴, Abbas Doosti⁵

1- Department of Biology, Shk. C., Islamic Azad University, Shahrekord, Iran

2- Department of Biology, Shk. C., Islamic Azad University, Shahrekord, Iran

3- Medical Biology Research Center, Health Technology Institute, Kermanshah University of Medical Sciences, Kermanshah, Iran

4- Applied Microbiology Research Center, Biomedicine Technologies Institute, Baqiyatallah University of Medical Sciences, Tehran, Iran

5- Biotechnology Research Center, Shk.C., Islamic Azad University, Shahrekord, Iran

***Corresponding Author:** Maryam Peymani, Professor, Department of Biology, Shk.C., Islamic Azad University, Shahrekord, Iran; **Email:** m.peymani@iau.ir

| Received | 2024/08/20

| Revised | 2025/08/3

| Accepted | 2025/08/27

| Abstract |

Background and Aim: Bioinformatics tools play a crucial role in screening and optimizing amino acid mutations in recombinant proteins. Given the lack of a comprehensive Persian review on tools predicting the impact of mutations on protein stability, this study aimed to review and introduce these computational tools.

Materials and methods: In this study, Google Scholar, PubMed, and Scopus databases were searched to retrieve relevant articles published between 1971 and 2023. The search strategy utilized keywords including "bioinformatics methods", "stability", "mutation", and "predictive databases". Furthermore, protein stability prediction tools (such as PoPMuSiC, HotMuSiC, DUET, SDM, mCSM, CUPSAT, DDGun, STRUM, and I-Mutant) and 3D structure validation servers (including I-TASSER, QMEAN, Phyre2, PROCHECK, ProSA, and ERRAT) were analyzed and compared regarding their algorithms, functions, and features.

Results: The findings indicated that stability prediction tools assess the impact of amino acid mutations on protein structure and melting temperature (ΔT_m) by analyzing sequences, dynamic flexibility, entropy changes, and free energy of stability ($\Delta \Delta G$). Following modeling, validation servers facilitate the selection of the most stable 3D model by evaluating stereochemical quality, model accuracy, and structural errors. The finalized model can serve as a reference structure for gene construct design and experimental validation.

Conclusion: Utilizing bioinformatics tools in protein engineering significantly reduces laboratory time and costs, enabling the rational design of recombinant proteins with optimal stability, thereby facilitating advancements in applied biotechnology.

| Keywords |

Protein Conformation|
Mutation|
Protein Engineering|
Thermodynamics|
Protein Stability|
Protein Databases|

| Iau Science |

ارزیابی پایگاه‌های داده بیوانفورماتیکی پیش‌بینی‌کننده جهش‌های مؤثر در پایداری پروتئین‌های نو ترکیب |

شبنم خاموشی^۱، مریم پیمانی^{۲*}، خیراله یاری^۳، جعفر امانی^۴، عباس دوستی^۵

۱- گروه زیست‌شناسی، واحد شهرکرد، دانشگاه آزاد اسلامی، شهرکرد، ایران

۲- گروه زیست‌شناسی، واحد شهرکرد، دانشگاه آزاد اسلامی، شهرکرد، ایران

۳- مرکز تحقیقات بیولوژی پزشکی، پژوهشکده فناوری‌های سلامت، دانشگاه علوم پزشکی، کرمانشاه، ایران

۴- مرکز تحقیقات میکروبیولوژی کاربردی، پژوهشکده فناوری‌های زیست‌پزشکی، دانشگاه علوم پزشکی بقیه‌الله (عج)، تهران، ایران

۵- مرکز تحقیقات بیوتکنولوژی، واحد شهرکرد، دانشگاه آزاد اسلامی، شهرکرد، ایران

*نویسنده مسئول: مریم پیمانی، گروه زیست‌شناسی، واحد شهرکرد، دانشگاه آزاد اسلامی، شهرکرد، ایران؛ پست الکترونیکی: m.peymani@iaui.ir

| تاریخ دریافت | ۱۴۰۳/۰۵/۳۰ | | تاریخ بازنگری | ۱۴۰۴/۰۵/۱۲ | | تاریخ پذیرش | ۱۴۰۴/۰۶/۰۵ |

چکیده |

زمینه و هدف: ابزارهای بیوانفورماتیکی نقش کلیدی در غربالگری و بهینه‌سازی جهش‌های آمینواسیدی در پروتئین‌های نو ترکیب دارند. با توجه به عدم وجود مقاله جامع فارسی در زمینه ابزارهای پیش‌بینی‌کننده اثر جهش بر پایداری پروتئین، پژوهش حاضر با هدف بررسی و معرفی این ابزارها انجام شد.

مواد و روش‌ها: در این مطالعه، جهت یافتن مقالات مرتبط در بازه زمانی ۱۹۷۱ تا ۲۰۲۳، پایگاه‌های گوگل اسکالر، پاب‌مد و اسکوپوس جست‌وجو شدند. واژگان کلیدی شامل «روش‌های بیوانفورماتیک»، «پایداری»، «جهش» و «پایگاه‌های پیش‌بینی‌کننده» بودند. همچنین ابزارهای پیش‌بینی پایداری (مانند PoP، Hot MuSiC، MuSiC، DUET، SDM، mCSM، CUPSAT، DDGun، STRUM و I-Mutant) و پایگاه‌های داده اعتبارسنجی ساختار سوم پروتئین (شامل PROCHECK، Phyre2، QMEAN، I-TASSER، ProSA و ERRAT) مورد تجزیه، تحلیل و مقایسه قرار گرفتند.

نتایج: ابزارهای پیش‌بینی با آنالیز توالی، بررسی انعطاف‌پذیری دینامیکی و ارزیابی تغییرهای آنتروپی و انرژی آزاد ($\Delta\Delta G$)، تأثیر جهش‌های اسیدآمینو بر ساختار و دمای ذوب (T_m) پروتئین را پیش‌بینی می‌کنند. پس از مدل‌سازی نیز، سرورهای اعتبارسنجی نیز با سنجش کیفیت استریوشیمیایی و دقت مدل‌ها، فرایند انتخاب پایدارترین مدل سه‌بعدی را تسهیل می‌کنند؛ مدل نهایی می‌تواند به عنوان ساختار مرجع برای طراحی سازه ژنی و آزمایش‌های تجربی استفاده شود.

نتیجه‌گیری: ابزارهای بیوانفورماتیکی در مهندسی پروتئین، با کاهش چشمگیر هزینه و زمان آزمایشگاهی، امکان طراحی هوشمندانه پروتئین‌های نو ترکیب با پایداری بهینه را فراهم کرده و گامی مؤثر در توسعه بیوتکنولوژی کاربردی هستند.

واژگان کلیدی |

ساختار پروتئین |

جهش |

مهندسی پروتئین |

ترمودینامیک |

پایداری پروتئین |

پایگاه‌های داده پروتئینی |

| علوم Iau |

| ۱ مقدمه

پروتئین‌های نو ترکیب با پتانسیل درمانی بالا، در حوزه‌های زیست‌فناوری^۱، آنزیم، سایتوکاین و واکسن شناخته شده و برای اهداف مختلفی از جمله تشخیص، پیشگیری و یا درمان مورد استفاده قرار می‌گیرند (۱). همچنین قابلیت بیان بالا و امکان بررسی خصوصیت‌های بیوشیمیایی، به آن‌ها فرصت استفاده در فرایندهای صنعتی و محصولات تجاری را داده است. در طول فرایند بیان و خالص‌سازی و ذخیره‌سازی پروتئین‌های نو ترکیب، ناپایداری، یکی از موضوع‌های چالش برانگیز است. عواملی از جمله بروز جهش، دناتوراسیون یا خاتمه زودرس زنجیره پلی پپتیدی (۲)، برای پایداری درون سلولی یک پروتئین حائز اهمیت هستند؛ چراکه این عوامل احتمالاً به دلیل جلوگیری از تاخوردگی^۲ مناسب یا ایجاد اختلال در ساختار سوم^۳، باعث تسریع در تخریب پروتئین‌ها می‌شوند (۳).

در گذشته، جهت ارزیابی پایداری پروتئین، به مطالعه‌های آزمایشگاهی اکتفا می‌شد (۴) که دارای توان عملیاتی پایین^۴، زمان و هزینه بودند. امروزه در دسترس بودن مدل ساختاری پروتئین یکی از کلیدهای درک فرایندهای بیولوژیکی در سطح مولکولی است که نیازمند استفاده از سرورهای بیوانفورماتیکی جهت پیش‌بینی تأثیر جهش‌ها روی پایداری است (۵). استراتژی‌های مختلفی برای افزایش پایداری ارائه شده که شامل تغییرهای هدفمند در مولکول پروتئین یا بهینه‌سازی در شرایط بیان و خالص‌سازی است (۶).

در تکنیک‌های طراحی و آنالیز پروتئین، جهش‌های نقطه‌ای باعث ایجاد افزایش یا کاهش پایداری می‌شوند (۷). از طرفی طراحی منطقی پروتئین^۵ با پایداری کنترل شده در طیف وسیعی از کاربردها از اهمیت زیادی برخوردار است (۸). به‌طوری که می‌توان عملکرد یک پروتئین را برای اهداف بیولوژیکی و صنعتی مطلوب تغییر داد (۹).

^۵ Rational Protein Design

^۶ High-throughput

^۷ Single Point Mutation

^۸ Mutant Stability Calculations

^۱ Biotechnology

^۲ Folding

^۳ Tertiary Structure

^۴ Low-throughput

در این دوره از عصر اطلاعات با توان عملیاتی بالا^۶، توانایی پیش‌بینی دقیق تأثیرهای جهش نقطه‌ای^۷ بر پایداری پروتئین، روشی اساسی برای درک تأثیرهای تنوع ژنوم انسانی است (۱۰). چنین رویکردی در گسترش پروتئین‌های نو ترکیب با عملکردهای بهینه شده ارزشمند است (۱۱). اما، پیش‌بینی دقیق تأثیر جایگزینی اسیدآمینه بر پایداری پروتئین یک چالش حل نشده مهندسی پروتئین است و مطالعه‌های آزمایشگاهی لازم باید صورت گیرد (۸).

موفقیت این روش‌ها به کیفیت ساختار ورودی، نمونه‌برداری ساختاری و انرژی آزاد برای ارزیابی توالی‌های جهش یافته بستگی دارد (۱۲). با توجه به پتانسیل بالای ابزارهای محاسباتی در زمینه مهندسی ژنتیک، سرورهای بیوانفورماتیکی مختلفی جهت بررسی پایداری پروتئین‌ها از طریق پیش‌بینی جهش‌های بهینه وجود دارند که در این مطالعه، جزئیات و عملکرد پایگاه‌های داده و سرورهای بیوانفورماتیکی در دسترس را با هدف پیش‌بینی جهش‌های پایدارکننده در پروتئین‌ها، مورد ارزیابی و بررسی قرار داده شده است.

| ۲ مواد و روش‌ها

| ۱-۲ پایگاه‌های داده پیش‌بینی کننده پایداری

در این مطالعه از مقاله‌های در دسترس مرتبط با پایگاه‌های داده پیش‌بینی کننده جهش برای پایداری پروتئین، بررسی ساختار سه‌بعدی و سپس ارزیابی آن‌ها در بازه زمانی ۱۹۷۱ تا ۲۰۲۳ مورد بهره‌برداری قرار گرفت. این پایگاه‌های داده با الگوریتم‌های متفاوت و بررسی پارامترهای مختلف، پایداری پروتئین جهش یافته را مورد ارزیابی قرار داده و در نهایت بهترین آمینواسید بهینه را برای بهبود پایداری پروتئین، ارائه می‌دهند. آدرس پیش‌بینی کننده پایداری بیان شده در جدول ۱ است.

| ۱-۱-۲ پایگاه داده MuSiC^۸

پایگاه داده MuSiC به‌عنوان نتیجه ۱۵ سال تحقیق آکادمیک تیم دهوک^۱ و همکاران در زمینه آنالیز تأثیر جهش بر پایداری پروتئین، از سال ۲۰۰۹ توسعه یافته‌اند. این پایگاه داده به ارزیابی اثرات بیوفیزیکی جهش‌های اسید آمینه در پروتئین‌ها می‌پردازد و شامل دو برنامه PopMuSiC و HOTMuSiC می‌باشد که به ترتیب تغییرات در پایداری ترمودینامیکی^۲ و پایداری حرارتی^۳ را پیش‌بینی می‌کنند (۱۳).

در PopMuSiC طراحی منطقی پروتئین‌ها با پایداری کنترل‌شده در حوزه‌های بیوتکنولوژیکی و زیست‌محیطی، از جمله آنزیم‌های سنتزی کاتالیزوری، از اهمیت فوق‌العاده‌ای برخوردار است (۸). پایگاه داده HOTMuSiC توالی پروتئین را اسکن می‌کند و تأثیر جایگزینی تمام اسید آمینه‌های ممکن را بر دمای ذوب پروتئین تنها در چند دور ارزیابی می‌کند که جایگزین آزمایش‌های جهش‌زایی پرهزینه و زمان‌بر شده است. این ابزار برای تجزیه و تحلیل مکانیسم‌های مولکولی مرتبط با پایداری حرارتی پروتئین در طیف گسترده‌ای از فرایندهای بیولوژیکی، پیش‌بینی ساختار و طراحی پروتئین کاملاً کارآمد است (۱۴).

۲-۱-۲ | پایگاه داده DDGun^۴

برای پیش‌بینی تغییرهای پایداری پروتئین به‌صورت الگوریتم‌های تکی و چندتایی از این پایگاه داده استفاده می‌شود. استفاده از اطلاعات توالی و ساختاری پروتئین، روش‌هایی جهت پیش‌بینی تغییرهای انرژی آزاد گیبس^۵ بین پروتئین‌های وحشی^۶ و جهش‌یافته^۷ توسعه داده شده است. بیشتر روش‌های موجود، خاصیت پیش‌بینی مقارنی دارند و تضمین می‌کنند که مقدار $\Delta\Delta G$ پیش‌بینی‌شده برای یک تغییر، به‌طور ضرورت برعکس آن مقداری باشد که در تغییر معکوس پیش‌بینی می‌شود (۱۵).

۲-۱-۳ | پایگاه داده DynaMut^۸

پروتئین‌ها، مولکول‌های پویایی هستند و عملکرد آن‌ها ذاتاً با حرکات مولکولی آن‌ها مرتبط است. از این رو مطالعه ماهیت پویایی، انعطاف‌پذیری و ساختارهای قابل دسترسی در پروتئین‌ها ضروری است. به‌طوری که برای درک عملکرد آن‌ها و ارزیابی تأثیر تغییرهای ممکن در یک پروتئین بر ساختار، عملکرد، برهم‌کنش‌ها و فنوتیپ‌های مختلف آن مهم است (۱۶). پایگاه داده DynaMut شامل ۲۲۹۷ جهش است که به‌طور تصادفی از مجموعه ۲۶۴۸ جهش نقطه‌ای مختلف در ۱۳۱ پروتئین کروی انتخاب شده‌اند که با مقایسه مستقیم، تأثیر جهش‌ها را بر انرژی آزاد ممکن می‌سازد (۱۷).

۲-۱-۴ | پایگاه داده DUET^۹

این پایگاه داده توسط دو گلاس^{۱۰} و همکاران در سال ۲۰۱۴ طراحی گردید و یک رویکرد محاسباتی یکپارچه برای مطالعه جهش‌های بد معنا^{۱۱} در پروتئین‌ها، براساس دیدگاه‌های مختلف، با ترکیب انتخابی دو روش مکمل mCSM^{۱۲} و SDM^{۱۳} (۱۸) ساخته شد. یک روش یادگیری ماشینی^{۱۴} برای پیش‌بینی اثرهای جهش‌های اشتباه براساس فرم‌های ساختاری است (۱۹).

از یک نمایش نموداری از محیط رزیدوی وحشی برای استخراج الگوهای هندسی، فیزیکی و شیمیایی استفاده می‌کند که در کاربردهای مختلف، از جمله طبقه‌بندی ساختاری، پیش‌بینی عملکردی (۲۰) و همچنین پیش‌بینی لیگاند پروتئین مبتنی بر گیرنده در مقیاس بزرگ (۲۱) استفاده شده‌اند. SDM تابع انرژی پتانسیل آماری است و از فراوانی جایگزینی اسید آمینه خاص محیطی در خانواده پروتئینی همولوگ استفاده می‌شود که این امر، برای محاسبه امتیاز پایداری به‌عنوان نماینده‌ای از تفاوت انرژی آزاد بین پروتئین وحشی و جهش‌یافته به‌کار می‌رود (۱۸).

⁸ Dynamics of Mutations

⁹ Database of Unified Evolutionary Treatments

¹⁰ Douglas

¹¹ Missense

¹² mutation Cutoff Scanning Matrix

¹³ Site Directed Mutator

¹⁴ Machine Learning

¹ Dehouck

² Thermodynamic stability

³ Thermal stability

⁴ $\Delta\Delta G$ untrained

⁵ $\Delta\Delta G$; Gibbs free energy change

⁶ Wild-type

⁷ Mutant

عملکرد بهتری در جداسازی جهش‌های پایدارکننده تجربی نسبت به سایر روش‌هایی که صرفاً بر ساختارهای پروتئینی تجربی متکی هستند ارائه می‌دهد (۲۴).

۲-۲ | پایگاه‌های داده پیش‌بینی ساختار سه‌بعدی پروتئین

۱-۲-۲ | پایگاه داده I-TASSER^۹

پایگاه داده I-TASSER جهت پیش‌بینی ساختارهای سه‌بعدی پروتئینی که سال ۲۰۰۶ در مرکز بیوانفورماتیک KU منتشر شد. فرایندهای مدل‌سازی ساختار پروتئین با دانش تخصصی انسانی و اطلاعات بیوشیمیایی (عملکرد، جهش‌زایی، رزیدوهای کاتالیزوری و ...) می‌تواند در مونتاژ ساختاری و انتخاب مدل کمک کند. گسترش الگوریتم‌های خودکار مزیت بالایی در کاربرد بالقوه پیش‌بینی‌های ساختاری در مقیاس پروتئوم^{۱۰} دارد و این امکان را به همه افراد می‌دهد تا مدل‌های ساختاری را برای توالی‌های خود تولید کنند. معیار آن، شناسایی الگوهای همولوگ^{۱۱} با تشابه بیش از ۳۰ درصد است. برای پروتئین‌های چندزنجیره‌ای، یک جهت‌گیری نادرست ممکن است باعث تغییر در امتیاز تطابق ساختاری (TM-score^{۱۲}) و ریشه میانگین مربعات انحراف (RMSD^{۱۳}) شود که منجر به عدم تطابق امتیاز اطمینان (C-score^{۱۴}) با کیفیت کلی مدل می‌گردد (۲۵).

۲-۲-۲ | پایگاه داده QMEAN^{۱۵}

این پایگاه داده دسترسی به دو عملکرد امتیازدهنده را که در هشتمین دوره از ارزیابی سراسری پیش‌بینی ساختار پروتئین (CASP^{۱۶}) در سطح کلی با موفقیت آزمایش شده‌اند فراهم می‌کند. کاربر می‌تواند بین تابع امتیازدهی ترکیبی QMEAN که یک برآوردکننده کیفی براساس تحلیل هندسی مدل‌های تکی و تابع امتیازدهی مبتنی بر

۲-۱-۵ | پایگاه داده I-Mutant^۱

پایگاه داده‌ای است که به منظور پیش‌بینی تغییرهای پایداری پس از جهش در ساختار یا توالی پروتئینی مورد استفاده قرار می‌گیرد. پیش‌بینی‌ها یا از ساختار پروتئین یا از توالی پروتئین شروع می‌شوند. این روش بر روی مجموعه داده‌های مشتق شده از ProTherm، که از جامع‌ترین پایگاه داده‌های موجود از داده‌های تجربی ترمودینامیکی تغییرهای انرژی آزاد پایداری پروتئین بر اثر جهش در شرایط مختلف است، آموزش داده شد. این پایگاه داده می‌تواند هم برای پیش‌بینی نقش جهش بر روی پایداری پروتئین و هم به عنوان یک تخمین‌کننده رگرسیونی^۲ برای پیش‌بینی مقادیر $\Delta\Delta G$ مرتبط استفاده شود (۲۲).

۲-۱-۶ | پایگاه داده CUPSAT^۳

این پایگاه داده توسط پارتیبان^۴ و همکارانش در سال ۲۰۰۶ طراحی گردید که ابزاری جهت آنالیز پایداری پروتئین دانشگاه کلن است و برای پیش‌بینی تغییرات پایداری با جهش‌های نقطه‌ای استفاده می‌شود. مدل پیش‌بینی شده از پتانسیل‌های اسید آمینه-اتم و توزیع زاویه پیچش^۵ برای ارزیابی محیط اسید آمینه، قابلیت دسترسی به حلال و ویژگی ساختار ثانویه^۶ آن صورت می‌گیرد (۲۳).

۲-۱-۷ | پایگاه داده STRUM^۷

رویکردی به منظور پیش‌بینی تغییرهای پایداری در ساختار پروتئین با جهش تک‌نقطه‌ای است. جهت اعتباربخشی، مطالعه‌ای متشکل از ۱۵۰ پروتئین با ۳۴۲۱ جهش آزمایش شده که با استفاده از ۵ پایگاه داده مختلف، مورد ارزیابی قرار گرفت. ضریب همبستگی پیرسون^۸ جداسازی جهش‌های ایجاد شده و آزمایشی را از پروتئین‌های غیرهمولوگ کاهش می‌دهد و نیز نشان داد که STRUM

⁹ Iterative Threading ASsembly Refinement

¹⁰ Proteome

¹¹ Homolog

¹² Template Modeling score

¹³ Root-Mean-Square Deviation

¹⁴ Confidence score

¹⁵ Qualitative Model Energy Analysis

¹⁶ Critical Assessment of Structure Prediction

¹ Incremental Mutant

² Regression estimator

³ Cologne University Protein Stability Analysis Tool

⁴ Parthiban

⁵ Torsion angle

⁶ Secondary structure

⁷ Structure-based prediction of protein stability changes upon single-point Mutation

⁸ Pearson correlation coefficient

خوشه‌بندی QMEANclust که برآوردکننده کیفی کلی و موضعی براساس مقایسه وزنی از همه مدل‌های ارائه شده است انتخاب کند. روش‌های انتخاب بهترین مدل به دو دسته تقسیم می‌شوند: دسته اول توابع امتیازدهی متکی به تجزیه و تحلیل مدل‌های تکی براساس معیارهای تکاملی (۲۶) که فیزیکی یا شیمیایی هستند (۲۷)، دسته دوم از اطلاعات موجود در مجموعه‌ای از مدل‌ها برای یک توالی فرضی با استفاده از مقایسه ساختاری همه‌جانبه به منظور امتیازدهی کیفی استفاده می‌کنند (۲۸).

۲-۲-۳ | پایگاه داده^۱ Phyre

این پایگاه داده از روش همولوژی مدلینگ^۲ به منظور پیش‌بینی ساختار سه‌بعدی پروتئین و محل‌های اتصال لیگاند و آمینواسیدهای جهش‌یافته استفاده می‌کند که سال ۲۰۱۱ شروع به کار کرد. نکته متمایزکننده این پایگاه داده، سهولت استفاده از آن است که پس از ارسال یک توالی پروتئین، ساختار دوم و سوم، ترکیب زنجیره آن‌ها و کیفیت مدل را تفسیر می‌کند. اما همانند پایگاه‌های داده دیگر چند محدودیت دارد: اگر همولوژی بین توالی ورودی و پروتئین الگو تشخیص داده نشود، ساختار پیش‌بینی شده قابل اطمینان نخواهد بود و این توانایی را دارد تا اثر فنوتیپی^۳ جهش نقطه‌ای را پیش‌بینی کند، اما نمی‌تواند با دقت بالا این کار را انجام دهد (۲۹). آدرس این پایگاه‌های داده پیش‌بینی‌کننده ساختار سه‌بعدی پروتئین بیان شده در جدول ۲ است.

۲-۳-۲ | اعتبارسنجی ساختار سه‌بعدی پروتئین‌ها

مشکل عمده در زیست‌شناسی ساختاری، تشخیص خطاها در مدل‌های تجربی و نظری ساختارهای پروتئینی است. امروزه اعتبارسنجی توسط پایگاه‌های داده زیر، تبدیل به یک روش معمول شده است:

۲-۳-۱ | پایگاه داده^۴ PROSESS

ابزاری در جهت ارزیابی ساختارهای پروتئینی مشخص شده توسط کریستالوگرافی اشعه ایکس^۵ یا طیف‌سنجی NMR^۶ طراحی شده است که در آن انواعی از روش‌هایی که قبلاً شناخته شده و کاملاً آزمایش شده است، جهت ارزیابی کلی و مشخص رزیدوها استفاده می‌شود از جمله: کیفیت کووالانسی و هندسی، پیوندی/غیرپیوندی، زاویه پیچش و تغییر شیمیایی که جداول دقیق، تصاویر ساختاری و نمودارهایی را برای مقایسه ساختارهای پروتئینی با کیفیت ایجاد می‌کند (۳۰).

۲-۳-۲ | پایگاه داده^۷ VoroMQA

ابزاری رایج در سنجش کیفیت ساختاری پروتئین و انتخاب دقیق مدل‌های اعتبارسنجی ساختارهای آزمایشی است. ساختارهای ورودی ممکن است پروتئین‌های مونومر یا کمپلکس‌های الیگومر^۸ باشند و برای هر باقی‌مانده^۹ ورودی، امتیازهای کلی و موضعی را در نظر می‌گیرد. امتیازهای مکانی در طول زنجیره پروتئینی با ارائه ساختار ثانویه و اطلاعات دسترسی به حلال افزایش می‌یابد.

ویژگی مهم پایگاه داده این است که امکان ارزیابی مستقیم میان‌گش‌های پروتئین-پروتئین و بین زنجیره‌ای را دارد که امتیازهای کیفی مربوط به تخمین انرژی مشترک و امتیازهای مکانی برای رزیدوهای درگیر در تداخل‌های بین زنجیره‌ای را ارائه می‌دهد و نیز امتیازهای ارزیابی کیفی را در چهار سطح اتم، رزیدو، میان‌گش و کل ساختار ارائه می‌دهد. این پایگاه داده، عملکرد فوق‌العاده‌ای در انتخاب مدل و تخمین دقت مناطق ساختاری مکانی دارد و همچنین از هیچ اطلاعات ساختاری تکاملی (مانند حفظ توالی، ساختار ثانویه پیش‌بینی شده) که امکان تغییر در طی زمان دارد استفاده نمی‌کند، بلکه از ساختار کلی پروتئینی استفاده می‌کند (۳۱).

۲-۳-۲ | پایگاه داده^{۱۰} ResQ

یک ابزار اصلی برای ارائه مدل‌های ساختاری و بررسی عملکرد بیولوژیکی پروتئین‌ها با رویکردی برای تخمین کیفیت

⁷ Voronoi Tessellation-based Model Quality Assessment

⁸ Oligomer

⁹ Residue

¹⁰ Residue-specific Quality

¹ Protein Homology/analogy Recognition Engine

² Homology modeling

³ Phenotypic

⁴ Protein Structure Evaluation Suite and Server

⁵ X-ray crystallography

⁶ Nuclear Magnetic Resonance spectroscopy

مدل براساسِ ضریبِ کلی G، تعدادِ رزیدوها در هسته، مناطق مجاز و غیرمجاز انتخاب می‌گردد (۳۳).

۲-۳-۵ | پایگاه داده ProSA^۶

این پایگاه داده توسط سیپل^۷ و همکاران در سال ۱۹۹۳ برای تشخیص خطاها در ساختار سه‌بعدی پروتئینی طراحی شد که با استفاده از امتیاز^۸ Z نمودارهای انرژی، مشکلات بالقوه مشاهده‌شده در ساختارهای پروتئین و کیفیت آن در زمینه تمام ساختارهای پروتئینی مشخص می‌شود. کاربرد آن شامل تشخیص خطا در ساختارهای آزمایشی، مدل‌های نظری و مهندسی پروتئین است (۳۴، ۳۵).

۲-۳-۶ | پایگاه داده ERRAT^۹

این پایگاه داده توسط کولوس و بیتس^{۱۰} در سال ۱۹۹۳ طراحی و معرفی گردید که از طریق بررسی الگوهای برهم‌کنش‌های اتمی غیرپیوندی^{۱۱} به منظور تأیید پروتئینی مورد استفاده قرار می‌گیرد. در این روش، نواحی دارای تاخوردگی درست یا نادرست در ساختار پروتئین، براساس برهم‌کنش‌های اتمی غیرپیوندی مشخص می‌شوند؛ چرا که با توجه به توزیع غیرتصادفی اتم‌ها نسبت به یکدیگر در پروتئین‌های طبیعی، وجود خطا در ساختار مدل منجر به ایجاد توزیع‌های تصادفی‌تر میان اتم‌ها می‌گردد که می‌توان این نواحی خطادار را با روش‌های آماری از توزیع‌های صحیح متمایز نمود (۳۶). آدرس تمامی این پایگاه‌های داده اعتبارسنجی ساختار سه‌بعدی پروتئین در جدول ۳ بیان شده است.

۳ | بحث

با توجه به پتانسیل بالای پایگاه‌های داده در مهندسی ژنتیک و صرفه‌جویی در وقت و هزینه، رسیدن به پروتئین پایدار بسیار حائز اهمیت است. در واقع این پایگاه‌های داده با اسکن توالی، تجزیه و تحلیل حرکات و انعطاف‌پذیری پروتئین و پایداری ناشی از

اختصاصی هر رزیدو، پیش‌بینی ساختاری و مشخصات ذاتی فاکتور دمایی^۱ مربوط به تمام رزیدوها در طول زنجیره است که در بیشتر پیش‌بینی‌های ساختاری نادیده گرفته می‌شود. مزیت اصلی آن استفاده از رویکردهای حد واسط شبیه‌سازی ساختار سه‌بعدی مانند تغییرهای ساختاری مسیرهای شبیه‌سازی و استخراج داده مبتنی بر همولوژی است. میانگین اختلاف بین خطاهای اندازه‌گیری و مشاهده‌شده $1/4 \text{ \AA}$ است که با هدف‌های ساختاری آزمایش‌های CASP9-11 مورد ارزیابی قرار گرفته که کمتر یا قابل‌مقایسه با سایر پیش‌بینی‌کننده‌های پیشرفته است.

اختلاف بین رزیدوی جهش‌یافته و وحشی با میانگین خطای $2/4 \text{ \AA}$ پیش‌بینی‌شده، اما ضریب خطای آن برای مدل‌هایی با امتیاز $< 1/5$ و کیفیت کلی بهتر که به کاربردهای بیولوژیکی مرتبط‌تر هستند، به $1/4 \text{ \AA}$ کاهش می‌یابد. تحلیل داده‌ها نشان دادند که خطای آن با کیفیت مدل‌های ساختاری همبستگی بالایی دارد، به‌طوری که RSQ دقت بالاتری در مناطق حفاظت‌شده ساختارهای ثانویه منظم نسبت به نواحی نامنظم دارد. این پایگاه داده نسبت به روش‌های دیگر که از ویژگی‌های ساختارهای حد واسطه استفاده نمی‌کنند بهتر عمل می‌نماید. اگرچه هیچ ویژگی ساختاری حد واسطی استفاده نشود، اما پیش‌بینی تک مدل ResQ که براساس جست‌وجوی پایگاه‌های داده همولوگ مبتنی بر توالی و ساختار می‌باشد نیز ارزشمند است (۳۲).

۲-۳-۴ | پایگاه داده PROCHECK^۲

این پایگاه داده توسط لاسکووسکی^۳ و همکارانش طراحی و از سال ۱۹۹۳ در دسترس قرار گرفت. به‌منظور بررسی کیفیت استریوشیمیایی^۴ و دقت مدل‌های سه‌بعدی پیش‌بینی‌شده، با تجزیه و تحلیل نمودار راماچاندرا^۵ ارزیابی می‌گردد و بهترین

⁷ Sippl

⁸ Z-score

⁹ Error Analyzer for Protein Structures

¹⁰ Colovos & Yeates

¹¹ Non-bonded

¹ Temperature factor / B-factor

² PROtein structure validation suite

³ Laskowski

⁴ Stereochemical

⁵ Ramachandran plot

⁶ Protein Structure Analysis

تغییرهای آنترپپی، تأثیر جهش‌های اسیدآمینو بر ساختار و دمای ذوب آن را بررسی می‌کنند. سپس با پایگاه‌های داده اعتبارسنجی، ارزیابی کیفیت استریشیمیایی، دقت مدل‌ها و تشخیص خطاها در ساختار سه‌بعدی پروتئین انجام شده و در نهایت مدل انتخابی جهت سنتز سازه ژنی، بررسی‌های آزمایشگاهی و تأیید نهایی، مورد استفاده قرار می‌گیرد. به علاوه از پایگاه‌های داده پیش‌بینی‌کننده در دسترس، از جمله HotMuSiC, PoPMuSiC, I-STRUM, DDGun, CUPSAT, mCSM, SDM, DUET و I-Mutant برای جهش‌های پایدارکننده و پایگاه‌های داده I-TASSER, QMEAN, Phyre2, PROCHECK, ProSA و ERRAT جهت اعتبارسنجی و پیش‌بینی ساختار سه‌بعدی یک پروتئین استفاده گردید.

پایگاه داده PoPMuSiC برای شناسایی سریع فهرستی از جهش‌ها با پایداری مطلوب مفید است و با در نظر گرفتن بهینه بودن هر اسیدآمینو در توالی و شناسایی نقاطضعف ساختاری پروتئین که می‌توانند از نظر عملکردی و یا در نواحی حساس ساختاری باشند، در طراحی منطقی پروتئین استفاده می‌شود. ضریب همبستگی $R=0.79$ و ریشه میانگین مربعات خطای $\sigma_c=0.86$ kcal/mol بین تغییرهای پایداری اندازه‌گیری و پیش‌بینی شده، پس از اعتبارسنجی متقابل به دست می‌آید (۴). در این رابطه کابریتا^۱ و همکاران جهت بهبود پایداری ویروس تنباکو از پایگاه داده PoPMuSiC استفاده نمودند و از پارامترهای انرژی آزاد تاخوردگی و انرژی گیبس جهت بررسی پایداری پروتئین مورد نظر استفاده کردند. به طوری که جهش‌یافته‌هایی که دارای اختلاف انرژی $-1 \text{ Kcal/mol} \leq$ داشتند، به عنوان جهش‌یافته‌های پایدارکننده در نظر گرفتند و در نهایت نتایج آن‌ها نشان داد که پروتئین جهش‌یافته در مقایسه با پروتئین وحشی پایداری بیشتری داشتند که نشان‌دهنده پیش‌بینی درست این پایگاه داده در پایداری پروتئین می‌باشد (۳۷). پایگاه داده HotMuSiC توالی پروتئین را اسکن می‌کند و تأثیر جایگزینی تمام اسیدآمینوهای ممکن

را بر دمای ذوب پروتئین تنها در چند دور ارزیابی می‌کند که جایگزین آزمایش‌های جهش‌زایی پرهزینه و زمان‌بر شده است. در گذشته، تلاش زیادی برای تجزیه و تحلیل مکانیسم‌های مولکولی مرتبط با پایداری حرارتی پروتئین در طیف گسترده‌ای از فرایندهای بیولوژیکی، پیش‌بینی ساختار و طراحی پروتئین کاملاً کارآمد است (۱۴). همچنین، می‌توان از این قابلیت با هدف پیش‌بینی رابطه بین پایداری حرارتی پروتئین، تکامل طبیعی و مکانیسم‌های سازگاری با شرایط محیطی سخت استفاده نمود (۳۸). مطالعه‌های زیادی روی پایداری پروتئین‌های مختلف با این پایگاه‌های داده انجام شده است، از جمله در مطالعه وی^۲ و همکاران، جهت پیش‌بینی جهش‌ها به منظور افزایش پایداری آنزیم دکستراناز از این پایگاه‌های داده استفاده شده است. تجزیه و تحلیل ساختاری نشان داد که افزایش پایداری می‌تواند مربوط به وارد کردن آمینواسید پرولین و اثرهای هیدروفوبیک آن باشد (۳۹). در مطالعه دیگری، ژانگ^۳ و همکاران از این پایگاه‌ها جهت بهبود پایداری آنزیم آگاراز استفاده نمودند (۴۰).

نتایج این مطالعه‌ها نشان می‌دهند که این پایگاه‌های داده جهت بهبود پایداری آنزیم‌های مختلف به عنوان ابزاری دقیق استفاده می‌شوند. پایگاه داده DDGun بدون هیچ گونه اطلاعی از مجموعه داده‌های تجربی، عملکرد قابل توجهی دارد، به طوری که ضرایب همبستگی پیرسون بین مقادیر $\Delta\Delta G$ پیش‌بینی شده و اندازه‌گیری شده برای تغییرهای تکی و چندجایگاهی به ترتیب حدود ۰/۵ و ۰/۴ می‌رسد. در این رویکرد ترکیبی از اطلاعات تکاملی مورد استفاده قرار می‌گیرد که نقش مهمی در عملکرد بسیار بالای این پایگاه داده در پیش‌بینی $\Delta\Delta G$ از طریق جهش پروتئین دارد. با در نظر گرفتن اینکه، اختلاف انرژی آزاد آشکار بین پروتئین وحشی و جهش‌یافته، اولین فاکتور تجربی در مطالعه تأثیر جهش‌یافته‌های ناشناخته بر سلامت انسان است و احتمالاً در نهایت تاخوردگی پروتئین را منعکس سازد.

دو معیار در ارزیابی خاصیت نامتقارن پیش‌بینی‌کننده $\Delta\Delta G$ استفاده می‌شود: اولین معیار ضریب همبستگی پیرسون بین جهش‌یافته‌های مستقیم و معکوس و معیار دوم

³ Zhang¹ Cabrita² Wei

به منظور پیش‌بینی پایداری پروتئین‌های مختلف استفاده شده است (۴۱، ۴۲).

دوتسکو^۱ و همکاران در یک مطالعه سیستماتیک از ۱۶ پایگاه داده جهت پیش‌بینی پایداری آنزیم‌های سلولاز استفاده کردند که در بین آن‌ها DynaMut جز یکی از پایگاه‌های داده با دقت بالا در پیش‌بینی بود (۴۳). تأثیر جهش‌ها بر پایداری و برهم‌کنش‌های پروتئین، به‌ویژه در درک تأثیرهای جهش‌های مندلی و سوماتیک بر پیشرفت بیماری، از ارزش زیادی برخوردار است. در پایگاه داده DUET نتایج پیش‌بینی‌شده به‌صورت تغییر در انرژی آزاد گیبس بیان می‌شوند و مقادیر منفی نشان‌دهنده جهش‌های ناپایدارکننده است. اطلاعات تکمیلی مانند دسترسی به حلال نسبی رزیدو، بهینه بودن پیوند هیدروژنی زنجیره جانبی و ساختار ثانویه محاسبه نشان داده شده است. پژوهشگران برای طراحی پروتئینی با پایداری بهینه اغلب از جهش‌های نقطه‌ای هدفمند در همان جایگاه و با تکنیک‌های مشابه استفاده می‌کنند که زمان‌بر و پرهزینه است و اغلب نیاز به استفاده از روش‌های پیش‌بینی محاسباتی برای انتخاب بهترین ترادف ممکن را دارد (۷).

جهش‌های تصادفی در یک موقعیت مشخص ممکن است به تولید پروتئین‌های مقاوم یا حساس به حرارت کمک کند و عملکرد یک پروتئین را برای اهداف بیولوژیکی و صنعتی مطلوب تغییر دهد. با استفاده از پایگاه‌های داده DUET و SDM، mCSM ارزیابی اثر تغییرهای پایداری در کاربردهای مختلفی مشخص شد، از قبیل جهش در ژن‌های دخیل در سرطان انسانی (۴۴)، مهار /ینوزین-۵-مونوفسفات‌دهیدروژناز (۴۵)، جهش‌های سوماتیک فسفودی‌استراز در سرطان (۴۶) و پروتئین پرزینلین-۲ مربوط به بیماری آلزایمر (۴۷). پایگاه داده I-Mutant برای انجام چهار پیش‌بینی مختلف از جمله جهش تک‌نقطه‌ای از طریق ساختار سوم پروتئین، تغییرهای پایداری با اندازه‌گیری مقدار $\Delta\Delta G$ با جهش تک‌نقطه‌ای از طریق ساختار سوم پروتئین، تغییرهای پایداری پروتئین بر جهش تک‌نقطه‌ای از توالی پروتئین و تغییرهای پایداری با

ارزیابی غیرمستقیم، سوگیری میانگین است که میزان انحراف (برحسب kcal/mol) از یک مجموعه داده مشخص را تخمین می‌زند. در واقع، ضریب همبستگی بین تغییرهای معکوس مستقیم و متناظر تقریباً از ۰ تا ۰/۷۵- متغیر است، درحالی‌که سوگیری از ۰/۳۲- تا ۰/۹۹- است. بسیار مهم است که یک پیش‌بینی‌کننده با قوانین فیزیکی که باید شبیه‌سازی کند، مطابقت داشته باشد و با توجه به تغییرهای $\Delta\Delta G$ پروتئین، نتایج نامتقارنی داشته باشد که می‌توانند به‌خوبی پیش‌بینی تغییرهای پایداری پروتئین براساس تغییرها را انجام دهند. در این پایگاه داده ویژگی‌های ساده‌ای براساس اطلاعات توالی و ساختاری معرفی شده که از نظر ساختاری متقارن نبوده و در نهایت، معیار لازم را در تعیین توانایی پیش‌بینی کمیت دارد، به‌طوری که روش‌های پیش‌بینی جدیدی را تشکیل می‌دهد (۱۵).

از آنجایی که پروتئین‌ها مولکول‌های پویایی هستند و عملکرد آن‌ها ذاتاً با حرکات مولکولی آن‌ها مرتبط است؛ از این‌رو مطالعه ماهیت پویایی، انعطاف‌پذیری و ساختارهای قابل دسترسی در پروتئین‌ها ضروری است. به‌طوری که برای درک عملکرد آن‌ها و ارزیابی تأثیر تغییرهای ممکن در یک پروتئین بر ساختار، عملکرد و برهم‌کنش‌ها و فنوتیپ‌های مختلف آن مهم است (۱۶). پایگاه داده DynaMut از دو رویکرد برای تجزیه و تحلیل هوشمندانه حرکات پروتئین و ارزیابی تأثیر جهش‌ها بر پایداری و پویایی پروتئین ناشی از تغییرهای آنروپی ارتعاشی استفاده می‌کند. به‌منظور درک پیامدهای مولکولی جهش، در نظر گرفتن تغییرها در دینامیک پروتئین مهم است، از این‌رو با ترکیب این دانش با علائم مبتنی بر نمودار، می‌توان تأثیرهای جایگزینی جهش بدمعنا بر پایداری پروتئین را دقیقاً بررسی کرد. نتایج نشان‌دهنده اینست که این رویکرد از روش‌های جایگزین برای پیش‌بینی اثرهای جهش‌ها بر پایداری و انعطاف‌پذیری پروتئین ($p < 0.001$) بهتر عمل می‌کند و به یک همبستگی تا ۰/۷ در تست‌های کور دست می‌یابد (۱۳). همچنین در مطالعه‌های زیادی از این پایگاه داده

¹ Dotsenko

موتانت‌های حاصل با انجام داکینگ و شبیه‌سازی دینامیک مولکولی به تأیید رسید (۵۴).

در پایگاه داده STRUM ضریب همبستگی پیرسون، جداسازی جهش‌های ایجادشده و آزمایشی را از پروتئین‌های غیرهمولوگ که منعکس‌کننده همبستگی ذاتی در نمونه جهش فعلی است کاهش می‌دهد. باوجوداین، نتایج به‌طور قابل توجهی نسبت به سایر روش‌هایی که بر روی ساختارهای پروتئینی آزمایشی ساخته شده‌اند بهتر عمل می‌کند. مهم‌ترین ویژگی‌ها در این پایگاه داده، شرایط انرژی مبتنی بر فیزیک در مدل‌ها و امتیازهای حفاظت‌شده از هم‌ترازی‌های الگوی چندرشته‌ای می‌باشند. اما، تا زمانی که فولد کلی درست باشد دقت پیش‌بینی $\Delta\Delta G$ وابستگی کمی به دقت مدل‌های ساختاری پروتئین دارد. این قابلیت، امکان استفاده از مدل‌سازی ساختار با وضوح پایین را برای پیش‌بینی تغییر پایداری با دقت بالا بر جهش‌های نقطه‌ای نشان می‌دهند. این پایگاه داده جهت تعیین $\Delta\Delta G$ در چهار مجموعه داده دیگر (S2648, S350, Q306 و p35) به‌عنوان کنترل، از چهار الگوریتم پیشرفته از جمله I-Mutant, mCSM و PopMusic استفاده کرده از این‌رو عملکرد پیشرفته آن احتمالاً به‌دلیل ترکیبی از چندین ویژگی مکمل استخراج‌شده از طیف گسترده‌ای از پایگاه‌های داده است. نتایج نشان دادند که این پایگاه داده، با بیشتر روش‌هایی که بر روی ساختارهای آزمایشی ایجاد شده‌اند قابل مقایسه است یا عملکرد بهتری دارد (۲۴).

برای تخمین دقت پیش‌بینی‌های I-TASSER از دو معیار استفاده می‌شود: ۱- C-score که یک تابع امتیازدهی براساس چگالی نسبی ساختاری و امتیاز کلی از الگوهای متعدد پشت سرهم است. ۲- TM-score که مرتبط با اندازه‌گیری شباهت ساختاری با مقادیر [۰ و ۱] است، بدین صورت که یک همبستگی قوی بین C-score و TM-score با ضریب همبستگی ۰/۹۱ به‌وجود می‌آید. در مدل‌های توپولوژی صحیح مقیاس C-score ۱.۵- > می‌تواند هر دو مقدار مثبت و منفی کاذب با مقادیر زیر ۰/۱ را مشخص کند. با ترکیب C-score و طول پروتئین، دقت

اندازه‌گیری مقدار $\Delta\Delta G$ با جهش تک‌نقطه‌ای تنها از توالی پروتئین طراحی شده است. دقت پیش‌بینی مبتنی بر ساختار ۰/۸ است و بیشتر از دقت پیش‌بینی مبتنی بر توالی است. همچنین دقت کلی ۰/۷۷ است حتی زمانی که پیش‌بینی براساس توالی باشد. به‌طوری که دقت با اندازه‌گیری همبستگی بین مقادیر پیش‌بینی‌شده و مقادیر $\Delta\Delta G$ مشاهده‌شده ارزیابی می‌شود. همبستگی داده‌های پیش‌بینی‌شده و تجربی ۰/۷۱ است، با خطای استاندارد ۱/۳ kcal/mol زمانی که روش مبتنی بر ساختار است. اما هنگامی که براساس توالی است، همبستگی بین داده‌های مشاهده‌شده و پیش‌بینی‌شده ۰/۶۲ با خطای استاندارد ۱/۴۵ kcal/mol است. بنابراین Mutant2.0 I-به‌عنوان یک کمک‌کننده ارزشمند برای طراحی پروتئین معرفی می‌شود، حتی زمانی که ساختار پروتئین هنوز با وضوح اتمی شناخته نشده باشد (۲۲).

همچنین پایگاه داده CUPSAT نیز در پیش‌بینی $\Delta\Delta G$ از تفاوت انرژی آزاد پروتئین‌های وحشی و جهش‌یافته استفاده می‌کند که مقادیر منفی و مثبت به‌ترتیب به‌معنای اثر ناپایدار/پایدارکننده هستند. دقت پیش‌بینی نسبت به سایر آزمون‌های اعتبارسنجی بیش از ۸۰٪ است. از این‌رو، این برنامه به‌عنوان یک ابزار ارزشمند برای تجزیه و تحلیل طراحی و پایداری پروتئین عمل می‌کند (۲۳). آزمون‌های اعتبارسنجی متقاطع حداکثر ضریب همبستگی ۰/۷۷ با خطای استاندارد $1/0 > \text{kcal/mol}$ را نشان دادند. همچنین چندین مطالعه، ابزارهایی را برای این منظور با دقت پیش‌بینی متوسط توسعه داده‌اند (۲۲، ۴۸-۵۱).

در مطالعه‌های مختلفی از جمله توماس^۱ و همکاران جهت پایداری آنزیم بوتیریل کولینسترز (۵۲) و آنتاپاگودم^۲ و همکاران جهت پایداری P450 Cytochrome با استفاده از این پایگاه داده، پیش‌بینی جهش‌های پایدارکننده با هدف رسیدن به فرم پایدار این پروتئین‌ها استفاده شده است (۵۳). در مطالعه ساینی^۳ و همکاران به‌منظور پیش‌بینی پایداری پروتئین اینترلوکین-۴ که نقشی مهمی در تمایز، آپوپتوز و التهاب دارد نیز از این پایگاه داده استفاده کردند و پایداری

³ Saini

¹ Thomas

² Anthappagudem

همچنین VoroMQA نیز عملکرد فوق العاده‌ای در انتخاب مدل و تخمین دقت مناطق ساختاری مکانی دارد و قابلیت تجزیه و تحلیل پیشرفته نتایج از جمله تولید امتیازهای کیفی کلی و مکانی (ModFOLD, ProQ3D)، پردازش ساختارهای چندگانه (eQuant, ModFOLD) و ساختارهای چندزنجیره‌ای (SBROD, DFIRE)، ارائه ابزاری برای تفسیر امتیازهای کلی و مشاهده‌های تأثیرگذار از امتیازهای مکانی (eQuant, ModFOLD)، تحلیل کامل یک ساختار ورودی با اندازه متوسط (SBROD, DFIRE) را دارد (۳۱).

ResQ میانگین اختلاف بین خطاهای اندازه‌گیری و مشاهده شده با هدفهای ساختاری آزمایش‌های CASP9-11 مورد ارزیابی قرار می‌دهد که کمتر یا قابل مقایسه با سایر پیش‌بینی‌کننده‌های پیشرفته است. اختلاف بین رزیدوی جهش‌یافته و وحشی با میانگین خطای $2/4 \text{ \AA}$ پیش‌بینی شده، اما ضریب خطای آن برای مدل‌هایی با امتیاز $< -1/5$ و کیفیت کلی بهتر که به کاربردهای بیولوژیکی مرتبط‌تر هستند، به $1/4 \text{ \AA}$ کاهش می‌یابد. تحلیل داده‌ها نشان داد که خطای آن با کیفیت مدل‌های ساختاری همبستگی بالایی دارد، به‌طوری که RSQ دقت بالاتری در مناطق حفاظت‌شده ساختارهای ثانویه منظم نسبت به نواحی نامنظم دارد. این پایگاه داده نسبت به روش‌های دیگر که از ویژگی‌های ساختارهای حد واسطه استفاده نمی‌کنند بهتر عمل می‌نماید. اگرچه هیچ ویژگی ساختاری حد واسطی استفاده نشود، اما پیش‌بینی تک‌مدل ResQ که براساس جست‌وجوهای پایگاه داده همولوگ مبتنی بر توالی و ساختار می‌باشد نیز ارزشمند است (۳۲).

در Errat، یکی از محدودیت‌ها این است که به‌خوبی بین خطاهای این مقیاس و خطاهای بیشتر را تشخیص نمی‌دهد و به روش اصلاح اتمی حساس است، همچنین ساختارهای اصلاح‌نشده معمولاً امتیاز خوبی ندارند ($E > 95\%$ حد اطمینان). به‌علاوه تحلیل ارائه‌شده، به‌طور قابل اعتمادی مناطق خطا در ساختارهای کریستالی پروتئین را با بررسی آماری برهم‌کنش‌های اتمی زوجی شناسایی می‌کند (۳۶).

امتیاز Z (Z-Score) در پایگاه داده ProSA کیفیت کلی مدل و تمام امتیازهای زنجیره‌های پروتئینی آزمایشی موجود در بانک داده پروتئین (۵۵) را نشان می‌دهد که اگر خارج از محدوده تعیین شده باشد، نشان‌دهنده ساختارهای اشتباه است.

مدل‌های I-TASSER را می‌توان با خطای میانگین $0/08$ برای TM-score و $2 \text{ \AA}$ برای RMSD پیش‌بینی نمود (۲۵). در Qmean هر دو رویکرد ارزیابی مزایای خود را دارند که بسته به موقعیت و در دسترس بودن اطلاعات انتخاب می‌شوند. علاوه بر این روش‌های توافقی دیگری از جمله Pcons و ModFOLD وجود دارند که مزیت این پایگاه داده این است که به‌منظور کمک به فرایند خوشه‌بندی مدل‌ها، یک رتبه‌بندی اولیه از مدل‌های تکی به‌دست‌آمده را انجام می‌دهد. تفاوت عملکرد امتیازدهی آن با سایر پایگاه‌های داده ارزیابی کیفی مدل (ModFOLD و ProQ) در شرایط تشکیل‌دهنده آن، پتانسیل برهم‌کنش اتمی وابسته به فاصله دقیق‌تر و همچنین پتانسیل زاویه پیچشی در سه رزیدوی متوالی است. کاربر وب پایگاه داده این امکان را دارد که مدل را ارزیابی کند و به آن یک امتیاز کیفی جهت رتبه‌بندی ساختاری داده بدهد. تخمین خطای هر رزیدو به روش‌های مختلفی ارائه شده و این امکان را به کاربر می‌دهد تا مدل‌های پروتئین را با جزئیات بیشتری بررسی و مقایسه برآورد کیفی رویکردهای مختلف را فراهم می‌کند که به‌طور بالقوه قابلیت اطمینان پیش‌بینی را بهبود می‌بخشد (۲۸).

پایگاه داده PHYERE همانند پایگاه‌های داده دیگر چند محدودیت دارد: اول اینکه در صورتی که همولوژی بین توالی ورودی و پروتئین الگو تشخیص داده نشود، ساختار پیش‌بینی شده اطمینانی نخواهد بود. دوم اینکه این پایگاه داده دارای این توانایی است تا اثر فنوتیپی جهش نقطه‌ای را پیش‌بینی کند، اما نمی‌تواند با دقت بالا این کار را انجام دهد. گاهی کاربر از یک پروتئین با ساختار چندمری می‌خواهد استفاده کند که در این مورد نیز، قادر به پیش‌بینی ساختار سه‌بعدی برای پروتئین‌های چندمری نمی‌باشد (۲۹).

پیشگیری از تقلب علمی یک هدف مهم در اعتبارسنجی ساختار است، اما پیشگیری از اشتباه‌های ناخواسته نیز حائز اهمیت است. اگرچه پایگاه داده PROSESS اولین ابزار اعتبارسنجی ساختاری نیست، اما اعتقاد براینست که داده‌های ساختاری با رویکردی بهتر به‌منظور شناسایی و جلوگیری از خطاهای پرهزینه را ارائه می‌دهد (۳۰).

نمودار انرژی رزیدوها، انحراف انرژی کلی ساختار را با توجه به توزیع انرژی حاصل از ترکیب‌های تصادفی اندازه‌گیری می‌کند. بخش‌های مشکل‌ساز یک مدل با نموداری از امتیازهای کیفی محلی شناسایی می‌شوند و امتیازها با استفاده از گدهای رنگی بر روی نمایشگر ساختار سه‌بعدی ترسیم می‌شوند. نمودار انرژی، کیفیت مدل را با ترسیم انرژی‌ها به‌عنوان تابعی از موقعیت توالی اسیدآمینه نشان داده و به‌طور کلی، مقادیر مثبت مربوط به بخش‌های مشکل‌دار یک مدل است (۵۶).

۴ | نتیجه‌گیری

در این مطالعه، به بررسی جامع و ارزیابی پایگاه‌های داده که در پیش‌بینی جهش‌های پایدارکننده در پروتئین نقش دارند و همچنین پایگاه‌های داده تأییدکننده ساختار سه‌بعدی پروتئین‌ها پرداخته شد. از این‌رو، می‌توان گفت که روش‌های مبتنی بر ساختار، عملکرد بهتری در مقایسه با ابزارهای مبتنی بر توالی دارند، زیرا الگوریتم‌های شناخته‌شده مبتنی بر ساختار نشان‌دهنده فضای ساختاری رزیدوی جایگزین شده هستند که به‌همراه پتانسیل‌های فیزیکی و آماری، عملکرد بهتری ارائه خواهند داد. برخی از پایگاه‌های داده مبتنی بر ساختار، براساس اختلاف انرژی عمل کرده که سهم انرژی در پایداری پروتئین را نشان می‌دهند. از طرفی، ابزارهای مبتنی بر توالی این مزیت را دارند که در صورت عدم دسترسی به ساختار سه‌بعدی پروتئین قابل اجرا می‌باشند و از روش‌های یادگیری ماشینی استفاده می‌کنند. این پایگاه‌های داده بر روی مجموعه داده‌های متفاوتی پایه‌ریزی شده‌اند و ارزیابی عملکرد این روش‌ها کار دشواری است.

با این حال، زمانی که از مجموعه داده‌های مستقل برای ارزیابی آن‌ها استفاده می‌شود، ضرایب همبستگی پیرسون مربوط به عملکرد پیش‌بینی ابزارهای پیشرفته از ۰/۵ تا ۰/۷ می‌رسد. بیشتر پیش‌بینی‌کننده‌ها براساس زیرمجموعه‌های پایگاه داده^۱ ProTherm آموزش دیده می‌شوند که شامل مجموعه‌ای از مقادیر $\Delta\Delta G$ و سایر معیارهای ترمودینامیکی پایداری پروتئین است و از ترکیبی از اطلاعات ساختاری و مبتنی بر توالی بهره

می‌گیرند. به‌صورت کلی می‌توان گفت که ابزارهای موفق در پیش‌بینی جهش‌های نقطه‌ای پایدار و بهینه‌کننده پروتئین موردنظر به‌شمار می‌روند. برای طراحی کارآمدتر این ابزارها، شناخت درست از دو حوزه مهندسی ژنتیک و بیوانفورماتیک نیاز است. با توجه به پتانسیل بالای ابزارهای بیوانفورماتیک در زمینه مهندسی ژنتیک، همچنین مقرون به‌صرفه بودن، استفاده از این روش‌ها برای بررسی تغییرهای اعمال‌شده در جهت رسیدن به پروتئین بهینه ارزشمند است و این امکان وجود دارد که داده‌های به‌دست‌آمده برای ارائه ابزارهای پیش‌بینی پایداری در آینده قابل استفاده باشند. از این‌رو، به‌منظور تأیید نتایج بیوانفورماتیکی، لزوم مطالعه‌های آزمایشگاهی مربوطه نیز توصیه می‌گردد.

۵ | ملاحظات اخلاقی

نویسندگان اعلام می‌دارند که در این پژوهش از هیچ‌گونه نمونه انسانی یا جانوری استفاده نشده است. همچنین در تمامی مراحل نگارش مقاله، اصول صداقت علمی و عدم سرقت ادبی کاملاً رعایت گردیده است.

۶ | تشکر و قدردانی

بدین وسیله از مرکز تحقیقات بیولوژی پزشکی کرمانشاه و اعضای محترم آن، به‌ویژه ریاست محترم جناب آقای دکتر یاری که ما را در انجام این تحقیق همراهی نمودند، صمیمانه تشکر می‌نمایم.

۷ | تعارض منافع

از این‌رو نویسندگان اعلام می‌دارند که در نگارش این مقاله هیچ‌گونه تضاد منافع مالی یا غیرمالی وجود ندارد و این پژوهش بدون هیچ‌گونه سوگیری نسبت به پایگاه‌های داده و ابزارهای معرفی‌شده انجام شده است.

۸ | سهم نویسندگان

ایده‌پردازی، جست‌وجوی منابع، جمع‌آوری اطلاعات پایگاه‌های داده و نگارش پیش‌نویس اولیه مقاله: شبنم خاموشی

¹ Protein Thermodynamic Database

طراحی و ساختاردهی مطالعه، نظارت علمی بر فرایند نگارش،
بازبینی منتقدانه و ویرایش نهایی متن: دکتر خیراله یاری و
دکتر مریم پیمانی

در تحلیل و ارزیابی ابزارها، ارائه مشاوره‌های تخصصی در
تفسیر داده‌ها، بازخوانی و اصلاح مقاله: دکتر جعفر امانی و
دکتر عباس دوستی

همچنین تمام نویسندگان نسخه نهایی مقاله را مطالعه و تأیید
نمودند.

| ۹ کد اخلاق

این مطالعه از نوع مروری بوده و نیازی به اخذ کد اخلاق ندارد.

| ۹ هوش مصنوعی

ندارد.

جدول ۱- پایگاه‌های داده پیش‌بینی‌کننده جهش‌ها.

نام پایگاه داده	آدرس
Pop Music	/https://soft.dezyme.com
Hot Music	/https://soft.dezyme.com
DUET	/https://biosig.lab.uq.edu.au/duet
SDM	/https://biosig.lab.uq.edu.au/sdmv2
Mcsm	/https://biosig.lab.uq.edu.au/mcsm
Cupsat	/http://cupsat.tu-bs.de
DDGun	/https://folding.biofold.org/ddgun
STRUM	/https://zhanggroup.org/STRUM
I-Mutant2.0	/https://folding.biofold.org/i-mutant/i-mutant2.0.html

جدول ۲- پایگاه‌های داده پیش‌بینی‌کننده ساختار سه‌بعدی پروتئین.

نام پایگاه داده	آدرس
I-TASSER	/https://zhanggroup.org/I-TASSER
Qmean	/https://swissmodel.expasy.org/qmean
Phyre2	/https://www.sbg.bio.ic.ac.uk/phyre2

جدول ۳- پایگاه‌های داده اعتبارسنجی ساختار سه‌بعدی پروتئین‌ها.

نام پایگاه داده	آدرس
PROSESS	/http://www.prosess.ca
VoroMQA	/http://bioinformatics.ibt.lt/wtsam/voromqa
ResQ	/https://zhanggroup.org/ResQ
Procheck	/https://saves.mbi.ucla.edu
ProSA	/https://prosa.services.came.sbg.ac.at/prosa.php
Errat	/https://saves.mbi.ucla.edu

| **Extended Abstract****Evaluation of Bioinformatics Servers for Predicting Mutations Affecting Recombinant Protein Stability**

Background and Aim: Recombinant protein production frequently encounters limitations due to structural instability and misfolding. The rational design of amino acid mutations via computational screening offers a potent alternative to empirical trial-and-error approaches. Despite the growing global reliance on computational biology, a comprehensive Persian essay summarizing these analytical servers remains unavailable. Therefore, this study aimed to evaluate and categorize advanced bioinformatics tools tailored for predicting mutation impacts on protein stability and validating the resulting three-dimensional structures.

Materials and Methods: In this study, Google Scholar, PubMed, and Scopus databases were searched to retrieve relevant articles published between 1971 and 2023. The search strategy utilized keywords including "bioinformatics methods", "stability", "mutation", and "predictive databases". Furthermore, protein stability prediction tools (such as PoPMuSiC, HotMuSiC, DUET, SDM, mCSM, CUPSAT, DDGun, STRUM, and I-Mutant) and 3D structure validation servers (including I-TASSER, QMEAN, Phyre2, PROCHECK, ProSA, and ERRAT) were analyzed and compared regarding their algorithms, functions, and features.

Results: The comparative screening revealed that mutation-based stability tools accurately quantify changes in melting temperature (ΔT_m) and free energy of stability ($\Delta \Delta G$). Following modeling, validation servers facilitate the selection of the most stable 3D model by evaluating stereochemical quality, model accuracy, and structural errors. The finalized model can serve as a reference structure for gene construct design and experimental validation.

Conclusion: Utilizing *in-silico* bioinformatics toolkits in structural biology drastically minimizes wet-lab screening times and experimental costs. By providing accurate mathematical modeling of structural thermodynamics, these predictive servers facilitate the design of highly stable recombinant proteins, establishing a robust foundation for targeted advancements in industrial and medical biotechnology.

| **Keywords**

Protein Conformation, Mutation, Protein Engineering, Thermodynamics, Protein Stability, Databases, Protein.

1. Baldwin MR, Bradshaw M, Johnson EA, Barbieri JT. The C-terminus of botulinum neurotoxin type A light chain contributes to solubility, catalysis, and stability. *Protein expression and purification*. 2004;37(1):187-95.
2. Goldberg AL, St. John A. Intracellular protein degradation in mammalian and bacterial cells: Part 2. *Annual review of biochemistry*. 1976;45(1):747-804.
3. Guruprasad K, Reddy BB, 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 Engineering, Design and Selection*. 1990;4(2):155-61.
4. Dehouck Y, Kwasigroch JM, Gilis D, Rooman M. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC bioinformatics*. 2011;12(1):1-12.
5. Chance MR, Bresnick AR, Burley SK, Jiang J-S, Lima CD, Sali A, et al. Structural genomics: a pipeline for providing structures for the biologist. *Protein science: a publication of the Protein Society*. 2002;11(4):723.
6. Savchenko A, Yee A, Khachatryan A, Skarina T, Evdokimova E, Pavlova M, et al. Strategies for structural proteomics of prokaryotes: Quantifying the advantages of studying orthologous proteins and of using both NMR and X-ray crystallography approaches. *Proteins: Structure, Function, and Bioinformatics*. 2003;50(3):392-9.
7. Shirley BA. *Protein stability and folding: Theory and practice*: Springer; 1995.
8. Goldenzweig A, Fleishman SJ. Principles of protein stability and their application in computational design. *Annual review of biochemistry*. 2018;87(1):105-29.
9. Lőrincz AT, Reed SI. Sequence analysis of temperature-sensitive mutations in the *Saccharomyces cerevisiae* gene CDC28. *Molecular and cellular biology*. 1986;6(11):4099-103.
10. Capriotti E, Nehrt NL, Kann MG, Bromberg Y. Bioinformatics for personal genome interpretation. *Briefings in bioinformatics*. 2012;13(4):495-512.
11. Socha RD, Tokuriki N. Modulating protein stability-directed evolution strategies for improved protein function. *The FEBS journal*. 2013;280(22):5582-95.
12. Davey JA, Chica RA. Optimization of rotamers prior to template minimization improves stability predictions made by computational protein design. *Protein Science*. 2015;24(4):545-60.
13. Pucci F, Kwasigroch JM, Rooman M. Protein thermal stability engineering using HoTMuSiC. *Structural Bioinformatics: Springer*; 2020. p. 59-73.
14. Lapidus LJ. Protein unfolding mechanisms and their effects on folding experiments. *F1000Research*. 2017;6.
15. Montanucci L, Capriotti E, Frank Y, Ben-Tal N, Fariselli P. DDGun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. *BMC bioinformatics*. 2019;20:1-10.
16. Albanaz AT, Rodrigues CH, Pires DE, Ascher DB. Combating mutations in genetic disease and drug resistance: understanding molecular mechanisms to guide drug design. *Expert opinion on drug discovery*. 2017;12(6):553-63.
17. Rodrigues CH, Pires DE, Ascher DB. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic acids research*. 2018;46(W1):W350-W5.
18. Pires DE, Ascher DB, Blundell TL. DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. *Nucleic acids research*. 2014;42(W1):W314-W9.
19. Pires DE, Ascher DB, Blundell TL. mCSM: predicting the effects of mutations

- in proteins using graph-based signatures. *Bioinformatics*. 2014;30(3):335-42.
20. Pires DE, de Melo-Minardi RC, dos Santos MA, da Silveira CH, Santoro MM, Meira W, editors. Cutoff Scanning Matrix (CSM): structural classification and function prediction by protein inter-residue distance patterns. *BMC genomics*; 2011: Springer.
21. Pires DE, de Melo-Minardi RC, Da Silveira CH, Campos FF, Meira Jr W. aCSM: noise-free graph-based signatures to large-scale receptor-based ligand prediction. *Bioinformatics*. 2013;29(7):855-61.
22. Capriotti E, Fariselli P, Casadio R. I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic acids research*. 2005;33(suppl_2):W306-W10.
23. Parthiban V, Gromiha MM, Schomburg D. CUPSAT: prediction of protein stability upon point mutations. *Nucleic acids research*. 2006;34(suppl_2):W239-W42.
24. Quan L, Lv Q, Zhang Y. STRUM: structure-based prediction of protein stability changes upon single-point mutation. *Bioinformatics*. 2016;32(19):2936-46.
25. Zhou X, Zheng W, Li Y, Pearce R, Zhang C, Bell EW, et al. I-TASSER-MTD: a deep-learning-based platform for multi-domain protein structure and function prediction. *Nature Protocols*. 2022;17(10):2326-53.
26. Chen H, Kihara D. Estimating quality of template-based protein models by alignment stability. *Proteins: Structure, Function, and Bioinformatics*. 2008;71(3):1255-74.
27. Benkert P, Tosatto SC, Schomburg D. QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics*. 2008;71(1):261-77.
28. Benkert P, Künzli M, Schwede T. QMEAN server for protein model quality estimation. *Nucleic acids research*. 2009;37(suppl_2):W510-W4.
29. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*. 2015;10(6):845-58.
30. Berjanskii M, Liang Y, Zhou J, Tang P, Stothard P, Zhou Y, et al. PROSESS: a protein structure evaluation suite and server. *Nucleic acids research*. 2010;38(suppl_2):W633-W40.
31. Olechnovič K, Venclovas Č. VoroMQA web server for assessing three-dimensional structures of proteins and protein complexes. *Nucleic acids research*. 2019;47(W1):W437-W42.
32. Yang J, Wang Y, Zhang Y. ResQ: an approach to unified estimation of B-factor and residue-specific error in protein structure prediction. *Journal of molecular biology*. 2016;428(4):693-701.
33. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*. 1993;26(2):283-91.
34. Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins: Structure, Function, and Bioinformatics*. 1993;17(4):355-62.
35. Wiederstein M, Sippl MJ. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*. 2007;35(suppl_2):W407-W10.
36. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein science*. 1993;2(9):1511-9.
37. Cabrita LD, Gilis D, Robertson AL, Dehouck Y, Rooman M, Bottomley SP. Enhancing the stability and solubility of TEV protease using in silico design. *Protein science*. 2007;16(11):2360-7.
38. Pucci F, Rooman M. Improved insights into protein thermal stability: from the molecular to the structurome scale. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and*

- Engineering Sciences. 2016;374(2080):20160141.
39. Wei Z, Chen J, Xu L, Liu N, Yang J, Wang S. Improving the thermostability of GH49 dextranase AoDex by site-directed mutagenesis. *AMB Express*. 2023;13(1):1-10.
 40. Zhang P, Zhang J, Zhang L, Sun J, Li Y, Wu L, et al. Structure-based design of agarase AgWH50C from *Agarivorans gilvus* WH0801 to enhance thermostability. *Applied microbiology and biotechnology*. 2019;103:1289-98.
 41. Khan MA, Varma AK. In silico and structure-based assessment to classify VUS identified in the α -helical domain of BRCA2. *Journal of Biomolecular Structure and Dynamics*. 2023;41(19):9879-89.
 42. Kumar S, Kumari K, Azad GK. An immunoinformatics approach to study the epitopes of SARS-CoV-2 helicase, Nsp13. *Vacunas*. 2023.
 43. Dotsenko A, Denisenko J, Osipov D, Rozhkova A, Zorov I, Sinitsyn A. Testing and improving the performance of protein thermostability predictors for the engineering of cellulases. *Journal of Bioinformatics and Computational Biology*. 2023;21(02):2330001.
 44. Malhotra S, Alsulami AF, Heiyun Y, Ochoa BM, Jubbs H, Forbes S, et al. Understanding the impacts of missense mutations on structures and functions of human cancer-related genes: a preliminary computational analysis of the COSMIC Cancer Gene Census. *PLoS One*. 2019;14(7):e0219935.
 45. Singh V, Pacitto A, Donini S, Ferraris DM, Boros S, Illyés E, et al. Synthesis and structure–activity relationship of 1-(5-isoquinolinesulfonyl) piperazine analogues as inhibitors of *Mycobacterium tuberculosis* IMPDH. *European journal of medicinal chemistry*. 2019;174:309-29.
 46. Munir A, Kumar N, Ramalingam SB, Tamilzhalagan S, Shanmugam SK, Palaniappan AN, et al. Identification and characterization of genetic determinants of isoniazid and rifampicin resistance in *Mycobacterium tuberculosis* in southern India. *Scientific reports*. 2019;9(1):10283.
 47. Maryam A, Vedithi SC, Khalid RR, Alsulami AF, Torres PHM, Siddiqi AR, et al. The molecular organization of human cGMP specific phosphodiesterase 6 (PDE6): structural implications of somatic mutations in cancer and retinitis pigmentosa. *Computational and structural biotechnology journal*. 2019;17:378-89.
 48. Cheng J, Randall A, Baldi P. Prediction of protein stability changes for single-site mutations using support vector machines. *Proteins: Structure, Function, and Bioinformatics*. 2006;62(4):1125-32.
 49. Gilis D, Rooman M. PoPMuSiC, an algorithm for predicting protein mutant stability changes. Application to prion proteins. *Protein engineering*. 2000;13(12):849-56.
 50. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *Journal of molecular biology*. 2002;320(2):369-87.
 51. Zhou H, Zhou Y. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein science*. 2002;11(11):2714-26.
 52. Thomas B, Boopathy R. Essential amino acids for the stability of human butyrylcholinesterase as predicted by CUPSAT server. In *Silico Biology*. 2008;8(5-6):517-29.
 53. Anthappagudem A, Enaganti S, Bhukya B. Integration of mutational and molecular docking studies: An in silico approach to assess the stability and binding potential of CYP3A4. *Journal of Applied Biology and Biotechnology*. 2022;11(1):161-70.
 54. Saini S, Jyoti-Thakur C, Kumar V, Suhag A, Jakhar N. In silico mutational analysis and identification of stability centers in human interleukin-4. *Molecular Biology Research Communications*. 2018;7(2):67.

55. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, et al. The protein data bank. *Acta Crystallographica Section D: Biological Crystallography*. 2002;58(6):899-907.
56. Hooda V, babu Gundala P, Chinthala P. Sequence analysis and homology modeling of peroxidase from *Medicago sativa*. *Bioinformation*. 2012;8(20):974.