

## ارزیابی فولدینگ RNA، روابط تکاملی، و مدل‌سازی مولکولی برهمکنش‌های سوبسترا- آنزیم در لوسیفراز *Luciola sp.* با داکینگ مولکولی و آنالیزهای بیوانفورماتیک

مجتبی مرتضوی<sup>۱\*</sup>، مسعود ترک زاده ماهانی<sup>۱</sup>، مهدی رحیمی<sup>۱</sup> و علی ریاحی مدوار<sup>۲</sup>

<sup>۱</sup> ایران، کرمان، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، پژوهشگاه علوم و تکنولوژی پیشرفته و علوم محیطی، پژوهشگاه علوم محیطی، گروه بیوتکنولوژی

<sup>۲</sup> ایران، بجنورد، دانشگاه کوثر بجنورد، دانشکده علوم پایه، گروه بیولوژی مولکولی و سلولی

تاریخ دریافت: ۱۴۰۱/۰۳/۰۳ تاریخ پذیرش: ۱۴۰۱/۱۰/۱۲

### چکیده

آنزیم‌های لوسیفراز، آنزیم‌هایی هستند که در انتشار نور نقش دارند و به طور گسترده در زمینه بیوتکنولوژی صنعتی و پزشکی استفاده می‌شوند. ما قبلاً کلون‌سازی ژن لوسیفراز را از کرم شب تاب گونه ایرانی *Luciola sp.* گزارش کرده بودیم. با استفاده از روش‌های محاسباتی، لایه پنهان داده‌های ژنتیکی و ساختاری را در *Luciola sp.* تجزیه و تحلیل کردیم. برای این منظور با کمک نرم افزار ACUA، ENC و CBI مورد مطالعه قرار گرفت و محتوای AT1، GC1، AT2، GC2، AT3 و GC3 محاسبه شد. در ادامه از سرورهای UNAFold و Mfold برای پیش‌بینی تاخوردگی RNA استفاده شد. از روش ChExVis برای استخراج، ذخیره و آنالیز کانال‌های لوسیفراز استفاده شد. در نهایت، داکینگ سوبسترا توسط نرم افزار AutoDock Vina انجام شد. این تجزیه و تحلیل نشان می‌دهد که کدون‌های نادر Pro، Arg، Ile و Leu در توالی این ژن وجود دارد. داکینگ ۵-[O-N]-adenosine (DehydroLuciferyl)-sulfamoyl با لوسیفراز انجام شد و شبکه‌ای از میانکنش‌ها شناسایی شد. در پایگاه محاسباتی RNAfold، بهترین ساختار ثانویه دارای انرژی آزاد ۳۹۷/۷۰- کیلوکالری بر مول و انرژی آزاد ترمودینامیکی ۴۲۷/۲۸- کیلوکالری بر مول است. در این مطالعه، برخی اطلاعات بیولوژیکی پنهان از کد ژنتیکی شناسایی شد که ممکن است در فرایند فولدینگ و اتصال به سوبسترا نقش داشته باشند و نشان دهد که ترکیبات کدونی ممکن است در فولدینگ مناسب این آنزیم نقش داشته باشند.

واژه‌های کلیدی: لوسیفراز، داکینگ، کدون نادر، فولدینگ RNA، AutoDock Vina.

\*نویسنده مسئول، پست الکترونیکی: [mortezavimm@gmail.com](mailto:mortezavimm@gmail.com)

### مقدمه

بدون سلول گزارشگر ژنتیکی و تعیین فعالیت فسفاتاز استفاده می‌شوند (۱۰، ۱۵، ۲۵، ۲۸، ۳۰). در این راستا، تعدادی از لوسیفرازهای ایرانی نیز قبلاً کلونینگ و توالی‌یابی شده‌اند (۱۱، ۱۹).

از چالش‌های مهم در بیان آنزیم لوسیفراز بهینه‌سازی کدونی و تغییر کدون‌هاست. جهش‌های مترادف یکی از چالش‌های مهم در حوزه بیوتکنولوژی و بیوتکنولوژی پزشکی است. این جهش‌ها توالی‌کد کننده mRNA را

واکنش نشر نور کرم شب تاب و سوسک‌های کلیک توسط آنزیم لوسیفراز کاتالیز می‌شود. در ابتدا یک اسید انیدرید بین گروه لوسیفرین و COO<sup>-</sup> از AMP در حالت برانگیخته الکترونیکی تشکیل می‌شود که با بازگشت به حالت پایه نور ساطع می‌شود (۹). سنجش‌های مبتنی بر لوسیفراز غیر رادیواکتیو و حساس هستند و در زمینه‌های مختلف بیوتکنولوژی پزشکی مانند تشخیص زنده‌مانی میکروبی، تشخیص سرطان، تشخیص ATP، سنتز پروتئین

هنوز سؤالات زیادی در مورد مکانیسم‌هایی که این اثرات را تنظیم می‌کنند وجود دارد.

اخیراً، ما کلونینگ ژن لوسیفراز کرم شب تاب ایرانی را از گونه *Luciola sp.* گزارش کرده‌ایم (۱۹). با توجه به کاربرد گسترده این آنزیم در زمینه بیوتکنولوژی، مطالعه دقیق ویژگی‌های کدونی و ساختاری این آنزیم اهمیت زیادی دارد. این داده‌ها در فرایندهای بعدی مهندسی این آنزیم مانند انتخاب جهش جهت تغییر فعالیت، انتخاب میزبان بیانی و بیان آنزیم فعال بسیار اهمیت دارد. چراکه انتخاب نامناسب کدونها می‌تواند منجر به تولید آنزیم غیرفعال گردد. ما برای اولین بار ویژگی‌های فولدینگ RNA، اتصال سوبسترا، کانال‌ها و کدون‌ها را در این لوسیفراز ایرانی بررسی کردیم. همچنین، ویژگی‌های ژنتیکی از جمله تعداد موثر کدون (ENC)، شاخص سوگیری کدون (CBI) و شاخص تطبیق کدونی (CAI) ارزیابی شد. علاوه بر این، وضعیت کدون‌ها توسط سرور RaCC و سایر وب سرورهای محاسباتی مورد مطالعه قرار گرفت (۷، ۴۷). در این مطالعه، برخی از ویژگی‌های کدونها مشخص شد، و در ساختار سه بعدی لوسیفراز *Luciola sp.* مطالعه شدند. این کدون‌ها با استفاده از نرم افزار SPDBV (Swiss PdbViewer) (۱۸) و PyMOL (۸) مورد ارزیابی و تجزیه و تحلیل قرار گرفتند. ساختارهای ثانویه RNA توسط سرورهای UNAFold و Mfold پیش‌بینی شد (۲۶، ۵۴).

مطالعه رونویسی mRNA ممکن است مکانیسم فولدینگ ساختار لوسیفراز را روشن کند. روش ChExVis روشی کارآمد است که برای استخراج، ذخیره و آنالیز کانال‌های پروتئینی استفاده می‌شود (۲۹). در نهایت، شبیه‌سازی اتصال مولکولی مولکول سوبسترای (5'-O-[N-DehydroLuciferyl)-sulfamoyl]) و آنزیم با استفاده از (31) AutoDock Vina انجام شد (۴۹). این ارزیابی‌های داده‌های پنهان می‌تواند تکامل مولکولی آنزیم و بینش‌هایی درباره تاریخچه ژن را نشان دهد. تجزیه و تحلیل این داده‌ها می‌تواند روند شناسایی نقش استفاده از کدون در

تغییر می‌دهند، اما توالی اسیدهای آمینه را حفظ می‌کنند. بنابراین، از نظر تاریخی، این جهش‌ها از نظر فنوتیپی خاموش در نظر گرفته شده‌اند و اغلب در مطالعات مختلف ژنتیک انسانی نادیده گرفته می‌شوند (۲۱، ۵۱). قبلاً تصور می‌شد که جهش‌های مترادف خاموش هستند و اهمیت کمی دارند، اما این موضوع توجه فزاینده‌ای از سوی فارماکولوژیست‌ها و ژنتیک‌دانان به سمت خود جلب کرده است (۱۶، ۴۲). مطالعات قبلی نشان می‌دهد که جهش‌های مترادف ممکن است بر مکان‌های اتصال miRNA، پایداری mRNA و تغییر ساختار و عملکرد پروتئین تأثیر بگذارد (۳۶، ۴۴). مطالعات دیگر جهش‌های مترادف نشان داده است که در پروتئین‌ها، این جهش‌ها عملکردی و از نظر ساختاری مهم هستند و تأثیرات قابل توجهی بر عملکرد و بیان پروتئین ایجاد می‌کنند (۱۷، ۴۰). علاوه بر این، سایر مطالعات نشان داده‌اند که یکی از عوامل موثر بر ترکیب پروتئین نهایی و فرآیند تاخوردگی همزمان، تغییر در سرعت ترجمه است (۲۲).

مطالعات نشان می‌دهد که جهش‌های مترادف ویژگی سوبسترا و محل اتصال سوبسترا را در ژن P-گلیکوپروتئین تغییر می‌دهند (۲۰). از سوی دیگر، فرض بر این است که در مرحله اولیه تکامل سلولی، یکی از ویژگی‌های ماشین ترجمه، خطای بالای آن بود که در ابتدا تنها چند اسید آمینه را کدگذاری می‌کرد و در ادامه، اسیدهای آمینه جدیدی در طول تکامل به این کدها اضافه شدند (۳۹). این نتایج با این مفهوم سازگار است که در طول تکامل کدون‌های مترادف در بیان و عملکرد پروتئین ترجمه شده بسیار مؤثر بوده‌اند (۴۵). تاکنون مطالعات ژنتیکی، بیوفیزیکی و بیوشیمیایی زیادی انجام شده است که از طریق آن ویژگی‌ها و اثرات دستگاه ترجمه مشخص می‌شود (۳۷). اگر نتایج قابل مقایسه‌ای در دسترس باشد، باید در دیدگاه خود در مورد ضریب تأثیر جهش‌های مترادف تجدید نظر کنیم (۱). با این حال، جهش‌های مترادف بر جنبه‌های مختلف پروتئین‌ها تأثیر می‌گذارند و

مشابه از پایگاه داده NCBI به دست آمد و تراز چند توالی آنها (MSA) توسط برنامه ClustalW2 و MEGA 7 انجام شد. تجزیه و تحلیل شامل ۱۶ توالی اسید آمینه بود.

#### مدل سازی ساختارهای ثانویه RNA: از سرورهای

UNAFold و Mfold برای پیش‌بینی ساختارهای ثانویه RNA استفاده شد (۲۷، ۵۵). در این سرورها، پارامترهای ترمودینامیکی، از جمله حداقل انرژی آزاد (MFE) و تابع پاریتیشن به طور خودکار برای فولدینگ ساختار محاسبه می‌شوند. در این سرورها از پارامترهای پیش فرض استفاده شده است.

#### بررسی کانال های مولکولی: در ابتدا، ویژگی های کدون

در PyMOL و SPDBV تجزیه و تحلیل شد. سپس آنالیزهای استریوشیمیایی این مدل در سرور نمودار رامچاندرا انجام شد. در مطالعه توابع مولکولی، درک ساختار کانال هایی که به مکان های فعال منتهی می شوند مهم است. در این راستا از روش ChExVis برای بررسی کانال ساختاری درون آنزیم استفاده شد. لوسیفراز حشرات شب تاب حاوی جایگاه فعال در دخل ساختار هستند و شناسایی و مشاهده کانال هایی که به محل فعال منتهی می شوند بسیار ارزشمند و مهم است.

#### داکینگ مولکولی: برای مطالعه داکینگ مولکولی، از

AutoDock Vina (نسخه ۱،۱،۲) استفاده شده است. O-۵-[N-(DehydroLuciferyl)-sulfamoyl] عنوان لیگاند و گیرنده پروتئینی در نظر گرفته می شوند. مدل مولکولی سه بعدی این آنزیم در ابزار MGL (نسخه ۱،۵،۴) به فرمت PDBQT تبدیل شد (۳۱). فرمت SDF سوپسترا از پایگاه داده PubChem تهیه و فرمت آن از PDB در Open Babel (نسخه ۲،۳،۱) به PDBQT تبدیل شد (۳۸). این قالب PDB در ابزار MGL به قالب PDBQT تغییر یافت. مکان مشابهی از اتصال سوپسترا در آنزیم *Luciola cruciata* luciferase برای فضای جستجوی اتصال سوپسترا استفاده

لوسیفراز کرم شب تاب را تسهیل کند. تعیین ویژگی های خاص این ژن لوسیفراز می تواند توسعه بیوتکنولوژی پزشکی این آنزیم حیاتی را تسهیل کند.

#### مواد و روشها

##### تجزیه و تحلیل توالی: قبلاً، توالی DNA کد کننده لوسیفراز

از گونه *Luciola sp.* شناسایی شده بود. کسر، فراوانی و تعداد کدون ها در توالی ژن مورد مطالعه قرار گرفت و کدون هایی که ترجیح داده شدند، در پایگاه محاسباتی Sequence Manipulation Suite شناسایی شدند (۴۶) پایگاه محاسباتی Sequence Manipulation Suite مجموعه ای از برنامه های جاوا اسکریپت برای تولید، قالب بندی و آنالیز توالی های کوتاه DNA و پروتئین است. پایگاه محاسباتی Sequence Manipulation Suite یک توالی DNA را می پذیرد و یک نمودار گرافیکی متشکل از یک نوار افقی برای هر کدون ایجاد می کند. طول نوار متناسب با فرکانس کدون در جدول فرکانس کدونی است که وارد می کنید. از نمودار کدون برای یافتن بخش هایی از توالی DNA که ممکن است ضعیف بیان شوند استفاده می شود. همچنین، پایگاه محاسباتی Sequence Manipulation Suite یک یا چند توالی DNA را می پذیرد و تعداد و فرکانس هر نوع کدون را برمی گرداند. از آنجایی که این برنامه همچنین فرکانس های کدون هایی را که برای همان اسید آمینه (کدون های مترادف) کد می کنند، مقایسه می کند، می توان از آن برای ارزیابی اینکه آیا یک دنباله برای کدون های مترادف خاصی ترجیح می دهد استفاده کرد. در ادامه از وب سرور RACC برای آنالیز کدون های کمیاب استفاده شد. ترجیح استفاده از کدون ژن لوسیفراز در ابزار GCUA مورد مطالعه قرار گرفت (۱۳). تعداد کدون های ترجیحی توسط CBI نرمال شده بین -۱ و ۱ تعیین شد. با نرم افزار ACUA، تنوع سوگیری استفاده از کدون با استفاده از ENC و CBI تجزیه و تحلیل شد. محتویات AT1، GC1، AT2، GC2، AT3 و GC3 نیز محاسبه شد (۳۸). برای تجزیه و تحلیل درخت فیلوژنتیک، توالی های مرجع پروتئین گونه های

G3 مورد مطالعه قرار گرفتند. مقادیر GC1، GC2، AT1، GC2، AT2، GC3 و AT3 برای مطالعه رابطه بین فشارهای ترکیبی و تنوع استفاده از کدون محاسبه شد. چولگی (دم تقارن توزیع احتمالی) و میزان cds کدونهای GC3 به ترتیب ۰،۳۹ و ۱۷،۱۵۳ محاسبه شد (جدول ۱). محتوای GC موقعیت کدون اول و دوم با محتوای GC در موقعیت کدون سوم مقایسه شد (جدول ۱).

شد به گونه ای که تمام باقیمانده‌های اتصال سوبسترا درگیر شوند (۳۵).

## نتایج

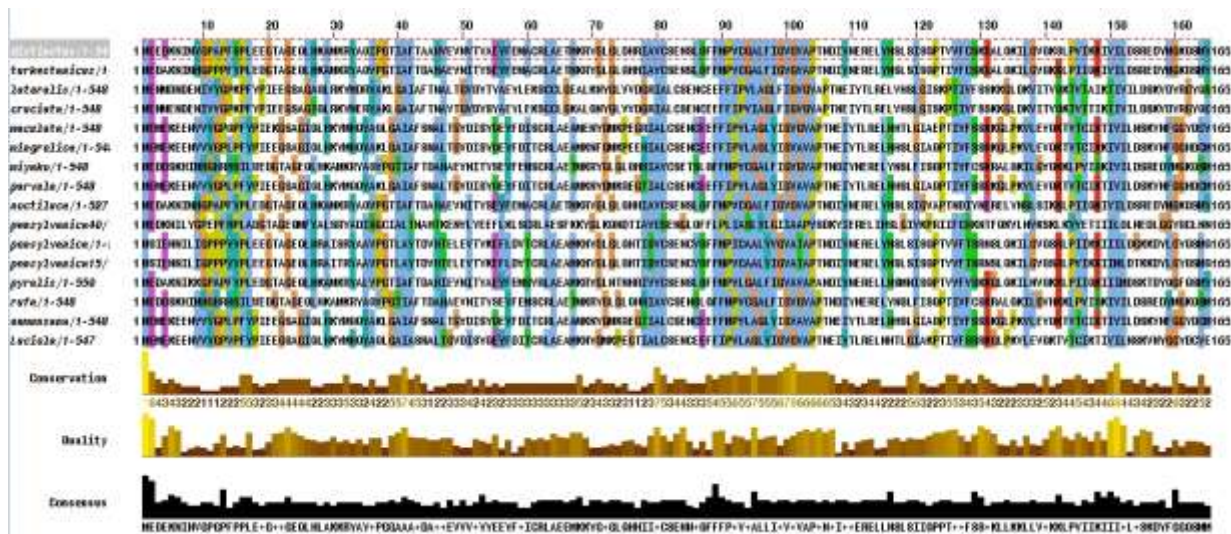
استفاده از کدون و تجزیه و تحلیل ترکیب نوکلئوتیدی: ترکیب نوکلئوتیدی و رابطه بایاس کدون به ترتیب با در نظر گرفتن مقادیر C، C3، GC، GC3، A، A3، T، T3، G و

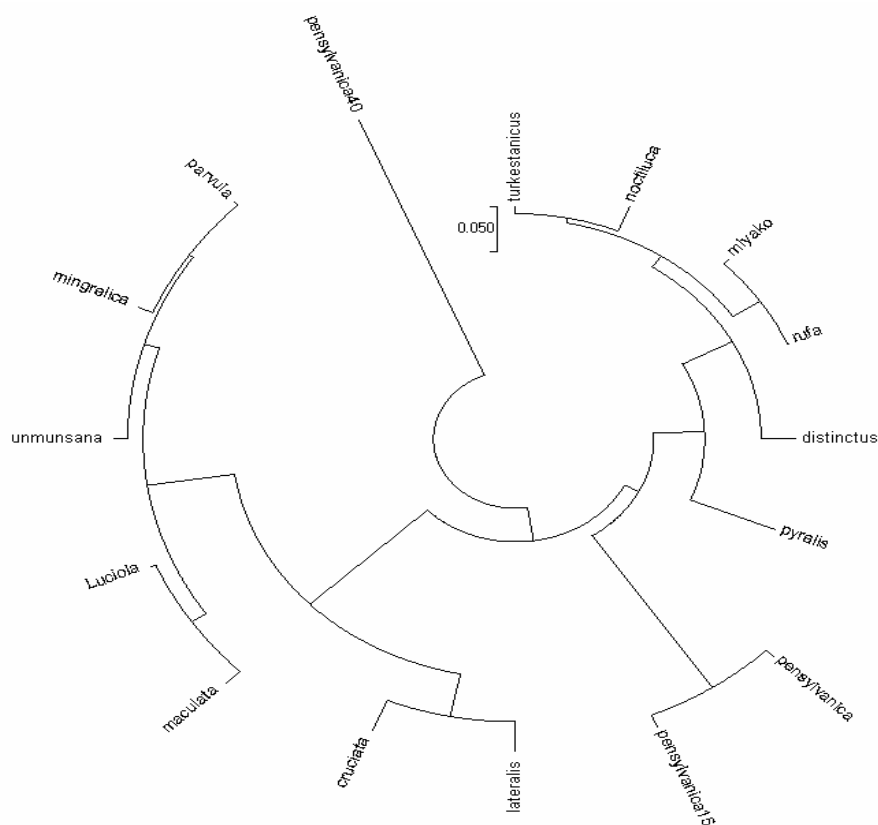
جدول ۱- تجزیه و تحلیل بیوانفورماتیک توالی ژن لوسیفراز

|          |        |          |        |          |        |          |        |
|----------|--------|----------|--------|----------|--------|----------|--------|
| A        | 536    | A1       | 181    | A2       | 186    | A3       | 168    |
| T        | 486    | T1       | 170    | T2       | 219    | T3       | 97     |
| G        | 345    | G1       | 86     | G2       | 62     | G3       | 197    |
| C        | 277    | C1       | 111    | C2       | 81     | C3       | 85     |
| AT%      | 62.165 | AT1%     | 21.35  | AT2%     | 24.635 | AT3%     | 16.119 |
| GC%      | 37.835 | GC1%     | 11.983 | GC2%     | 8.698  | GC3%     | 17.153 |
| AT       | 0.049  | AT1      | 0.031  | AT2      | -0.081 | AT3      | 0.268  |
| Skewness |        | Skewness |        | Skewness |        | Skewness |        |
| GC       | 0.109  | GC1      | -0.127 | GC2      | -0.133 | GC3      | 0.397  |
| Skewness |        | Skewness |        | Skewness |        | Skewness |        |
| Total bp | 1644   | CAI      | 0.595  | ENc      | 44.031 |          |        |

تکاملی با استفاده از روش تصحیح Poisson (۵۳) محاسبه شد و بر حسب واحد تعداد جایگزینی اسید آمینه در هر سایت است. آنالیزهای تکاملی در MEGA7 انجام شد (۲۳). ارزیابی کدونهای نادر: کدون های مشکل ساز در سرور <https://people.mbi.ucla.edu/sumchan/caltor> آنالیز شدند.

روابط تکاملی: شانزده توالی مشابه برای نمودار درخت فیلوژنتیک توسط سرور MEGA7 انتخاب شدند. توالی اسید آمینه این لوسیفرازها بازیابی و از NCBI به دست آمد. تاریخچه تکاملی با استفاده از روش Neighbor-Joining استنباط شد. درخت بهینه با مجموع طول شاخه = ۱،۳۴۳۶۲۲۱۵ نشان داده شده است (شکل ۱). فواصل





شکل ۱. تجزیه و تحلیل تکاملی و ساختاری آنزیم های لوسیفراز. (الف) درخت فیلوژنتیک خانواده لوسیفراز *Lampyridae* از *Luciolineae* (ب) هم ترازوی توالی پروتئین های آنزیم لوسیفراز از خانواده لامپرایدا. دامنه عناصر ساختاری ثانویه در بالا نشان داده شده است.

| Amino Acid | Rare Codon | Frequency of Occurrence |
|------------|------------|-------------------------|
| Arginine   | CGA        | 3                       |
|            | CGG        | 0                       |
|            | AGG        | 0                       |
| Isoleucine | AGA        | 4                       |
| Leucine    | AUA        | 1                       |
| Proline    | CUA        | 4                       |
|            | CCC        | 7                       |

جدول ۲. فراوانی اسیدهای آمینه نادر Ile, Arg, Leu و Pro در

توالی لوسیفراز *Luciola sp.*

در حقیقت، موقعیت باقیمانده ها نقش مهمی در ساختار و فعالیت آنزیم ها دارد. فرکانس و تعداد کدون ها در Sequence Manipulation Suite تعیین شدند. کسر، تعداد و فراوانی کدون ها برای هر باقیمانده در جدول ۳ ارائه شده است. نتایج نشان داد که کدون ها فرکانس های متفاوتی در این ژن دارند (جدول ۳).

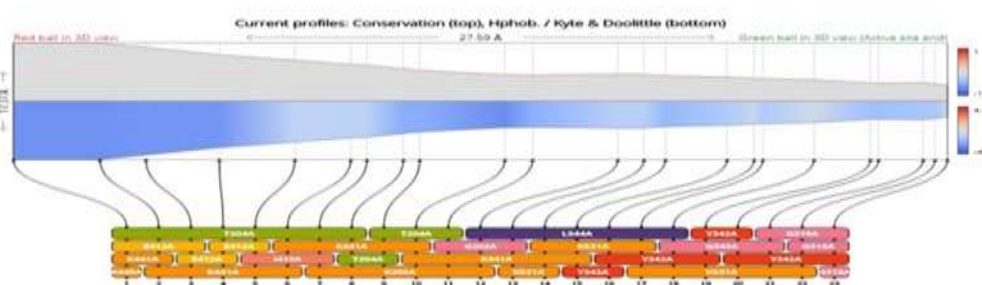
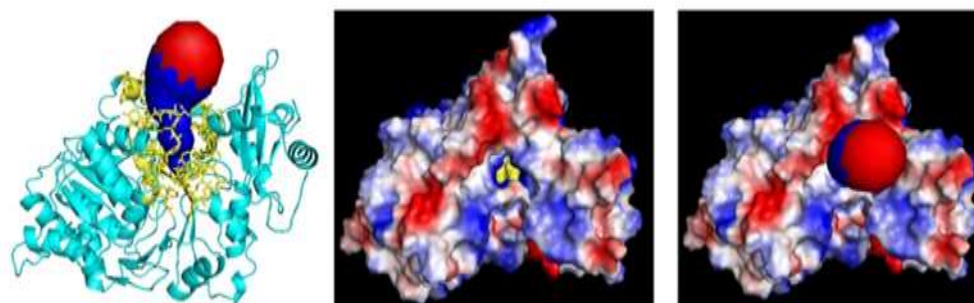
کدون های نادر آن دسته از کدون هایی هستند که با فرکانس های کمتر از ۱۰٪ در ارگانسیم های بیانی خاص مانند اشریشیا کلی (E.coli) در مقایسه با میزان اصلی استفاده می شوند. به این منظور توالی DNA هدف در قالب fasta تهیه می شود. جدول کدونی از پایگاه داده <http://www.kazusa.or.jp/codon> استفاده می شود. با توجه به جدول کدونی، کدون های نادر با رنگ متفاوت در توالی DNA برجسته می شوند. این تجزیه و تحلیل نشان می دهد که هفت کدون نادر Pro 18, 183, 335, 368, 385, 396, 474، هفت کدون نادر Arg 130, 219, 331, 338, 387, 388, 538 یک کدون نادر Ile ۴۰۳، و چهار کدون نادر Leu 273, 310, 334, 527 وجود دارد. این تجزیه و تحلیل همچنین یک تکرار دوگانه پشت سر هم کدون نادر Arg را نشان می دهد (جدول ۲).

جدول ۳. ویژگی‌های نوکلئوتیدی لوسیفراز *Luciola sp.* تعداد، نسبت در ۱۰۰۰ و کسر فراوانی کدون‌ها برای اسیدهای آمینه

| AmAcid | Codon | Number | /1000 | Fraction | AmAcid | Codon | Number | /1000 | Fraction |
|--------|-------|--------|-------|----------|--------|-------|--------|-------|----------|
| Ala    | GCG   | 0.00   | 0.00  | 0.00     | Asn    | AAT   | 12.00  | 21.90 | 0.57     |
| Ala    | GCA   | 10.00  | 18.25 | 0.33     | Asn    | AAC   | 9.00   | 16.42 | 0.43     |
| Ala    | GCT   | 18.00  | 32.85 | 0.60     | Pro    | CCG   | 2.00   | 3.65  | 0.06     |
| Ala    | GCC   | 2.00   | 3.65  | 0.07     | Pro    | CCA   | 10.00  | 18.25 | 0.32     |
| Cys    | TGT   | 6.00   | 10.95 | 0.75     | Pro    | CCT   | 12.00  | 21.90 | 0.39     |
| Cys    | TGC   | 2.00   | 3.65  | 0.25     | Pro    | CCC   | 7.00   | 12.77 | 0.23     |
| Asp    | GAT   | 18.00  | 32.85 | 0.78     | Gln    | CAG   | 3.00   | 5.47  | 0.21     |
| Asp    | GAC   | 5.00   | 9.12  | 0.22     | Gln    | CAA   | 11.00  | 20.07 | 0.79     |
| Glu    | GAG   | 5.00   | 9.12  | 0.12     | Arg    | AGG   | 0.00   | 0.00  | 0.00     |
| Glu    | GAA   | 38.00  | 69.34 | 0.88     | Arg    | AGA   | 4.00   | 7.30  | 0.27     |
| Phe    | TTT   | 14.00  | 25.55 | 0.58     | Arg    | CGG   | 0.00   | 0.00  | 0.00     |
| Phe    | TTC   | 10.00  | 18.25 | 0.42     | Arg    | CGA   | 3.00   | 5.47  | 0.20     |
| Gly    | GGG   | 2.00   | 3.65  | 0.04     | Arg    | CGT   | 8.00   | 14.60 | 0.53     |
| Gly    | GGA   | 21.00  | 38.32 | 0.40     | Arg    | CGC   | 0.00   | 0.00  | 0.00     |
| Gly    | GGT   | 23.00  | 41.97 | 0.44     | Ser    | AGT   | 8.00   | 14.60 | 0.33     |
| Gly    | GGC   | 6.00   | 10.95 | 0.12     | Ser    | AGC   | 2.00   | 3.65  | 0.08     |
| His    | CAT   | 6.00   | 10.95 | 0.50     | Ser    | TCG   | 0.00   | 0.00  | 0.00     |
| His    | CAC   | 6.00   | 10.95 | 0.50     | Ser    | TCA   | 4.00   | 7.30  | 0.17     |
| Ile    | ATA   | 1.00   | 1.82  | 0.03     | Ser    | TCT   | 8.00   | 14.60 | 0.33     |
| Ile    | ATT   | 30.00  | 54.74 | 0.75     | Ser    | TCC   | 2.00   | 3.65  | 0.08     |
| Ile    | ATC   | 9.00   | 16.42 | 0.23     | Thr    | ACG   | 2.00   | 3.65  | 0.06     |
| Lys    | AAG   | 11.00  | 20.07 | 0.24     | Thr    | ACA   | 11.00  | 20.07 | 0.31     |
| Lys    | AAA   | 34.00  | 62.04 | 0.76     | Thr    | ACT   | 20.00  | 36.50 | 0.56     |
| Leu    | TTG   | 13.00  | 23.72 | 0.30     | Thr    | ACC   | 3.00   | 5.47  | 0.08     |
| Leu    | TTA   | 14.00  | 25.55 | 0.32     | Val    | GTG   | 9.00   | 16.42 | 0.18     |
| Leu    | CTG   | 1.00   | 1.82  | 0.02     | Val    | GTA   | 20.00  | 36.50 | 0.41     |
| Leu    | CTA   | 4.00   | 7.30  | 0.09     | Val    | GTT   | 16.00  | 29.20 | 0.33     |
| Leu    | CTT   | 11.00  | 20.07 | 0.25     | Val    | GTC   | 4.00   | 7.30  | 0.08     |
| Leu    | CTC   | 1.00   | 1.82  | 0.02     | Trp    | TGG   | 1.00   | 1.82  | 1.00     |
| Met    | ATG   | 13.00  | 23.72 | 1.00     | Tyr    | TAT   | 9.00   | 16.42 | 0.41     |
|        |       |        |       |          | Tyr    | TAC   | 13.00  | 23.72 | 0.59     |
|        |       |        |       |          | End    | TGA   | 0.00   | 0.00  | 0.00     |
|        |       |        |       |          | End    | TAG   | 0.00   | 0.00  | 0.00     |
|        |       |        |       |          | End    | TAA   | 1.00   | 1.82  | 1.0      |

(۲۹). در این روش، کانال‌ها به صورت هندسی محاسبه شدند و هم‌خط مرکزی و هم‌حجم اشغال شده توسط کانال گزارش شد. نتایج تحلیل کانال‌ها و وضعیت بستر در شکل ۲ نشان داده شده است.

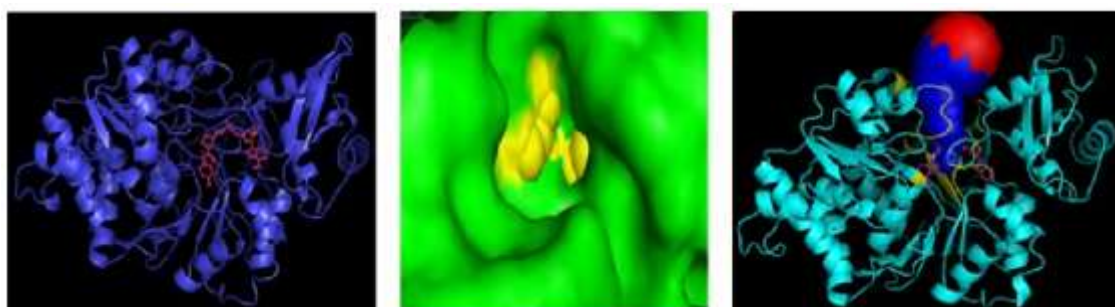
**استخراج کانال مولکولی:** ما قبلاً ساختار این لوسیفراز را مدل‌سازی کردیم و در ادامه برای تحلیل ساختار جامع، از سرور رامپاندران برای تجزیه و تحلیل نمودار رامپاندران این مدل استفاده شد (۲). در ادامه برای استخراج کانال مولکولی از روش ChExVis استفاده شد



شکل ۲. استخراج کانال در لوسیفراز. منافذ با رتبه برتر با استفاده از مدل توپ گرد (آبی) و موقعیت بستر در انتهای کانال نشان داده شده است. محل باقیمانده‌هایی که در کانال دخیل هستند: THR204, GLU412, LYS441, LYS445, LYS441, GLU412, THR204, LYS208, LYS441, ILE413, TYR342, TYR342, GLY343, GLU441, GLU412, LYLE335R44LY, THR204, THR204, LYS208, LYS441, ILE413, MET3, GLY318, GLY318

این فضا، بیشتر باقیمانده‌های متصل به سوسترانت انتخاب و ارزیابی شدند (شکل ۳A). کمپلکس ۵-O-[N-5 (DehydroLuciferyl)-sulfamoyl]-adenosine-luciferase با شبکه‌ای از برهمکنش‌ها در شکل ۳B نشان داده شده است. همانطور که در شکل ۳A نشان داده شده است، محل سوسترانت یک شکاف بزرگ در آنزیم است. نتایج، باقی مانده‌هایی را شناسایی کرد که با باقی مانده‌هایی که قبلاً در لوسیفرازهای دیگر شناسایی شده بودند، مشابه بودند.

**داکینگ مولکولی:** برای ارزیابی اتصال مولکولی، از AutoDock Vina نسخه ۱,۱,۲ (۴۹) استفاده می‌شود. ساختار سه بعدی آنزیم در نرم افزار MGL به فرمت PDBQT تبدیل شد (۳۱). فرمت SDF آدنوزین ۵-O-[N-5 (DehydroLuciferyl)-sulfamoyl] از پایگاه داده PubChem گرفته شد و در Open Babel نسخه ۲,۳,۱ به فرمت PDB تبدیل شد (۳۸). فضای جستجو در موقعیت مشابهی از *Luciola cruciata luciferase* (2d1tA) تعیین شد (۳۵). در

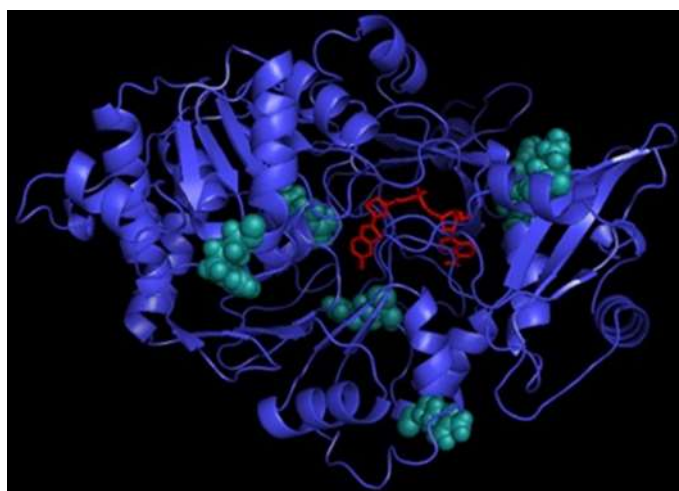




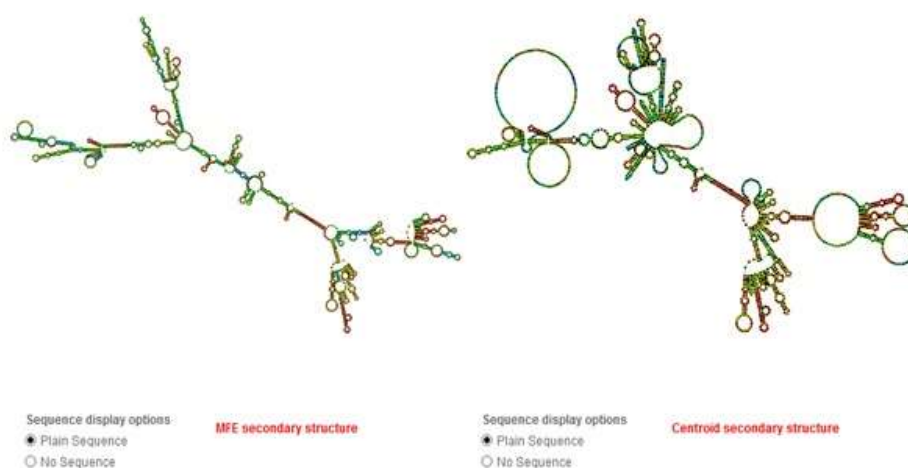
این الگوریتم داده‌هایی را در مورد توزیع و فراوانی کدون‌های رایج و کمیاب ارائه می‌دهد. به طور تجربی، نشان داده شده است که کدون‌های کمیاب می‌توانند نرخ ترجمه محلی را کاهش دهند، اما همبستگی کامل نیست. بزرگترین اندازه در کدون ۲۴۳ با  $\text{Min} = 50\%$  قرار داشت.

یک نمودار دایره‌ای یک ساختاری تولید می‌کند که در آن همپوشانی پایه‌ها معمولاً کوچک است (شکل ۷ب).

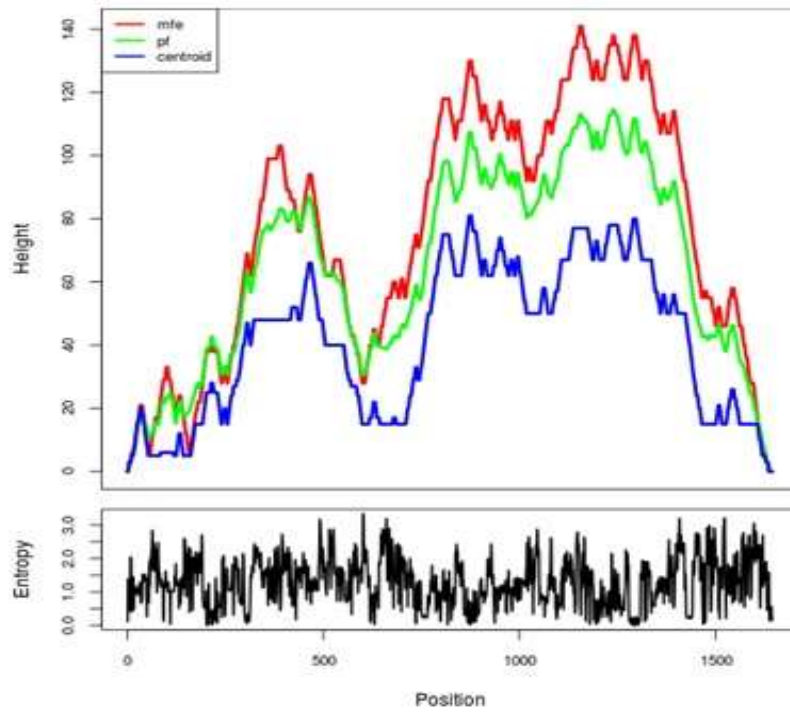
ارزیابی میزان تاخوردگی پروتئین: نادر بودن نسبی کدون‌ها برای ارزیابی سرعت فولدینگ پروتئین در الگوریتم Min-Max، تجزیه و تحلیل شد (۶) (شکل ۸).



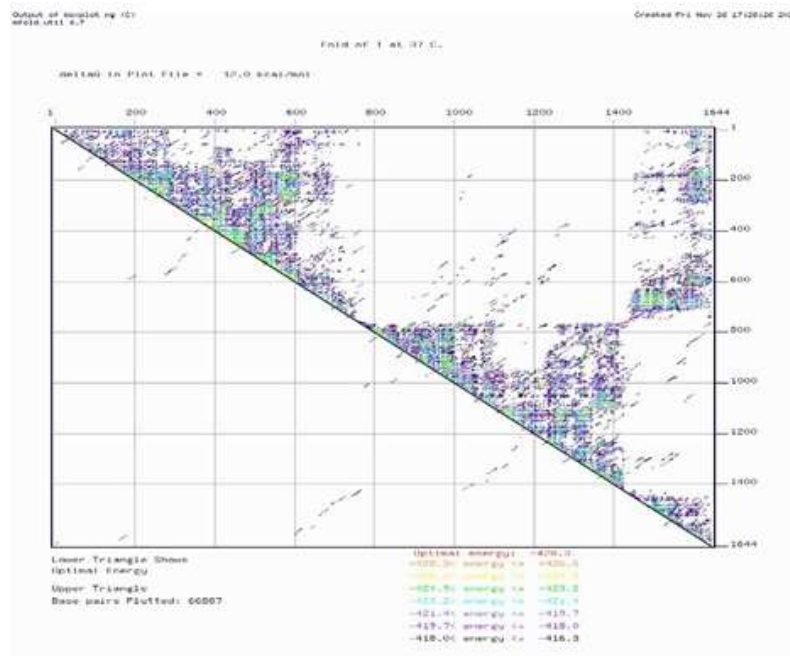
شکل ۴. تصویر روبانی کدون‌های کمیاب Arg (سبز) و DSLA (قرمز) در مدل مولکولی لوسیفراز. همانطور که در شکل مشخص است کدون‌های نادر در بخش‌های مختلف ساختار لوسیفراز قرار گرفته‌اند و این کدون‌ها در فرایند فولدینگ آنزیم لوسیفراز نقش مهمی ایفا می‌کنند.

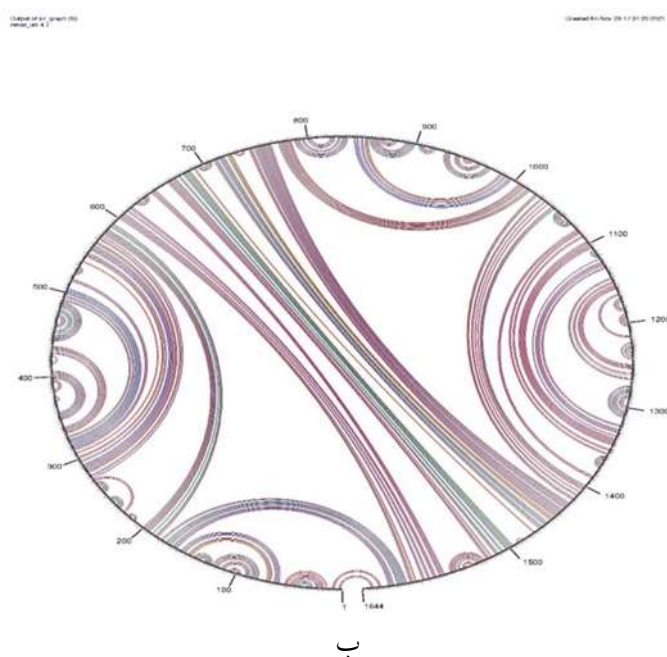


شکل ۵. ساختار دوم MFE و احتمال جفت مبتنی بر Centroid ژن لوسیفراز.

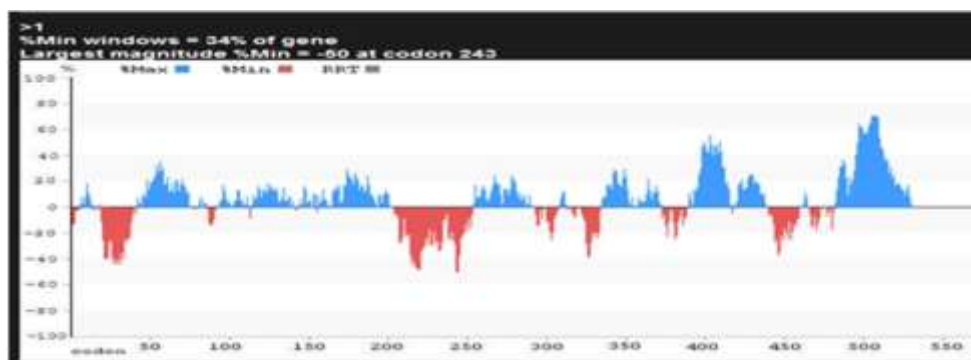


شکل ۶. نمایش گراف ساختار MFE. نتایج با استفاده از RNAfold 2.4.18 محاسبه شده است.





شکل ۷. الف: تا شدن RNA در وب سرور UNAFold نتیجه می‌دهد. کمان‌های G-C به رنگ قرمز کشیده شده‌اند. کمان‌های A-U، A-T به رنگ آبی کشیده شده‌اند. کمان‌های G-T، G-U به رنگ سبز کشیده شده‌اند. سایر کمان‌ها، در صورت وجود، به رنگ زرد هستند. ب: نمایش جفت پایه در نمودار دایره‌ای.



شکل ۸. توزیع و فراوانی کدون‌های رایج و کمیاب در توالی نوکلئوتیدی لوسیفراز *Luciola sp*

نور مناسب تأثیر بگذارد (۱۴). علاوه بر این، موقعیت کانال، فولدینگ صحیح RNA، اتصال سوپسترا و کدون‌های کمیاب تأثیرات زیادی بر ساختار و فعالیت این آنزیم‌ها دارند (۱۷). به منظور حل این مشکل، شبیه‌سازی و شناسایی لوسیفرازهای جدید بسیار حیاتی است. برای این منظور، کلونینگ و تعیین توالی لوسیفراز ایرانی گونه *Luciola sp* قبلاً انجام شد. در ادامه این مطالعه، تجزیه و تحلیل چالش‌های مشکل‌ساز بیان این آنزیم همانند

## بحث

بیولومینسانس ظرفیت موجودات زنده برای انتشار نور مرئی است. با تولید فوتون در واکنش بیوشیمیایی، بیولومینسانس رخ می‌دهد که در زمینه‌های مختلف زیست‌شناسی به عنوان ابزاری در سنجش واکنش‌ها مانند سنجش ATP استفاده می‌شود. اگرچه در این سنجش‌ها، تابش نور مناسب بسیار مهم است، عوامل مختلفی مانند بیان کم، عدم فعالیت پروتئین و pH ممکن است بر تابش

انفعالات غیرکوالانسی ایجاد شده توسط این Arg ممکن است تأثیر مهمی بر ساختار، فولدینگ و تشکیل جایگاه فعال مناسب داشته باشد. در طراحی جهش‌ها نیز باید به این یافته‌ها توجه کرد. علاوه بر این، در یک مطالعه قبلی، داکینگ بین آنزیم لوسیفرز *Luciola Lateralis* و لوسیفرین و سوبستراهای AMP مورد مطالعه قرار گرفت و نتایج نشان دهنده ساختار مشابه محل اتصال سوبسترا است (۴۹). در مطالعه دیگری، الگوی اتصال مشابهی از DSLC و PTC124-AMP نشان داده شد (۳). نتایج یک جایگاه اتصال سوبسترای مشابه را نشان می‌دهد و این مطالعات فرآیند داکینگ را تایید کرده و جایگاه فعالی که شناسایی شده است برای مطالعات بیشتر در مهندسی پروتئین قابل اعتماد است.

ما از روش استخراج کانال مولکولی بر اساس نمایش کمپلکس آلفا استفاده کردیم. در این روش کانال‌های امکان‌پذیر به صورت هندسی محاسبه شدند. نتایج نشان می‌دهد که ChExVis کانال‌هایی را در این آنزیم شناسایی کرده است که مورد مطالعه قرار گرفته و بینشی در مورد عملکرد مولکولی این آنزیم ایجاد خواهند کرد. برای مطالعه فولدینگ mRNA، فولدینگ RNA در فضای *in silico* از سرور RNAfold در پایگاه محاسباتی ViennaRNA انجام شد. ارزیابی الگوریتم Min-Max% نشان می‌دهد که این آنزیم باقیمانده‌های خود را با رایج‌ترین کدون‌ها رمزگذاری می‌کند. پارامترهای سوگیری استفاده از کدون به عنوان ترکیب پایه، CAI، CBI و ENC در این ژن مورد بررسی قرار گرفت. نتایج نشان می‌دهد که این ژن غنی از AT به میزان ۶۲ درصد است. از طرف دیگر، AT2% و GC3 دارای فرکانس بالاتر و AT3 و GC2 دارای فرکانس بسیار کمتری هستند. در بین ژن‌های مختلف، فشار جهش و انتخاب طبیعی دلایلی برای تنوع استفاده از کدون هستند (۵۲). علاوه بر این، یک رابطه معکوس بین بیان ژن و ENC وجود دارد. مقدار ENC پایین‌تر بیان ژن و ترجیح استفاده از کدون را نشان می‌دهد (۴).

کدون‌های نادر، CAI، CBI، ENC، فولدینگ RNA، و داکینگ سوبسترا انجام شد. پیش از این مطالعات جامع متعددی در این زمینه بر روی لوسیفرز و آنزیم‌های دیگر انجام شده است (۳۲-۳۴).

برای تولید نور، برخی از باقیمانده‌ها نقش مهمی دارند و موقعیت ساختاری این بقایا در مدل *Luciola sp.* مورد بررسی قرار گرفت. علاوه بر این، وضعیت کدون‌های کمیاب مورد مطالعه قرار گرفت و برخی از کدون‌های کمیاب Arg در مدل سه‌بعدی ساختار لوسیفرز آنالیز شدند. مطالعات ما نشان داد که این Arg‌ها دارای برخی اینتراکشن‌های غیرکوالانسی هستند که در تشکیل بخش‌های مختلف ساختار آنزیم نقش دارند. سرعت فولدینگ آنزیم در این باقیمانده‌ها کاهش یافته است و برای مطالعه این ادعا به داده‌های تجربی نیاز است. در مرحله بعد، ارزیابی داکینگ مولکولی با استفاده از Vina AutoDock (31) انجام شد. تجزیه و تحلیل دقیق نشان می‌دهد که جایگاه فعال لوسیفرز کرم شب تاب ژاپنی (d1t۲) مشابه مدل ما است (۳۵) و از این رو، فضای جستجو با توجه به محل اتصال سوبسترا در لوسیفرز کرم شب تاب ژاپنی انجام می‌شود. نتایج در نرم‌افزار PyMOL و Ligplot مورد تجزیه و تحلیل قرار گرفت. مدل مناسب همراه با باقیمانده‌های درگیر در محل اتصال انتخاب شد. با این حال، اگرچه محل کلی بسترها شبیه به لوسیفرز ژاپنی *Luciola cruciata* بود، به دلیل حذف یون  $Mg^{2+}$  و چرخش ناحیه C ترمینال، همه باقی مانده‌های قابل پیش‌بینی در نتایج داکینگ مشاهده نشدند.

مطالعه قبلی نشان می‌دهد که کدون‌های کمیاب سرعت ترجمه را کاهش می‌دهند و ترشح پروتئین را در اشریشیا کلی تسهیل می‌کنند (۲۴). در توالی‌های سیگنال‌پپتید در *Streptomyces coelicolor*، برخی کدون‌های کمیاب برای Arg، Ile و Lys بیش از تکرار شده‌اند (۲۴). در این راستا، نتایج مطالعه ما هفت باقیمانده با کدون نادر را شناسایی کرد که دور از مکان اتصال سوبسترا هستند. فعل و

است. این مطالعه به ما در طراحی و کاربردهای بیولوژیکی آینده سنجش مبتنی بر لوسیفراز کمک خواهد کرد.

### نتیجه‌گیری نهایی

مطالعه قبلی نشان می‌دهد که کدون‌های کمیاب سرعت ترجمه را کاهش می‌دهند و تغییر سرعت ترجمه می‌تواند روی ترکیب پروتئین تأثیر بگذارد. این تجزیه و تحلیل نشان می‌دهد که هفت کدون نادر Pro، هفت کدون کمیاب نادر Arg، یک کدون نادر Ile و چهار کدون نادر Leu وجود دارد. به نظر می‌رسد توالی ناحیه ۲۰۰ تا ۲۵۰ در فرایند فولدینگ آنزیم‌نقش مهمی دارد. نتایج ما نشان داد که برخی از کدون‌های کمیاب Arg دارای اینتراکشن‌های غیرکووالانسی هستند و با تغییر سرعت فولدینگ آنزیم در تشکیل بخش‌های مختلف ساختار آنزیم نقش دارند. این داده‌ها نشان می‌دهند که ترکیبات کدونی در محل مناسب تاخوردگی و اتصال بستر نقش دارند. نتایج داکینگ جایگاه اتصال سوبسترا را شناسایی کرده و جایگاه فعالی که شناسایی شده برای مطالعات بیشتر در مهندسی پروتئین قابل اعتماد است. در طراحی جهش‌ها نیز باید به این یافته‌ها توجه کرد. این مطالعه در کاربردهای بیولوژیکی لوسیفراز بسیار مهم است.

قدردانی :

این پژوهش در قالب طرح پژوهشی شماره (۱۷۲۷/۹۹/ص/۷) با استفاده از اعتبار پژوهشی پژوهشگاه علوم و تکنولوژی پیشرفته و علوم محیطی، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، کرمان ایران انجام شده است.

یک مطالعه نشان می‌دهد که تغییر سرعت ترجمه می‌تواند روی ترکیب پروتئین تأثیر بگذارد (۲۰). تغییرات در الگوی استفاده از کدون می‌تواند تأثیرات چشمگیری بر روی مقدار پروتئین سنتز شده و بازده پروتئین به درستی تا شده داشته باشد. تغییرات در فرکانس استفاده از کدون در یک میزبان بیان هترولوگ می‌تواند منجر به تغییراتی در نرخ سنتز پروتئین محلی شود، زیرا کدون‌های مترادف نادر نسبت به هم‌تایان معمول خود آهسته‌تر ترجمه می‌شوند. به طور کلی، کاهش سرعت ترجمه، زمان در دسترس بخش‌های N-ترمینال پروتئین را افزایش می‌دهد تا قبل از ظاهر شدن قسمت‌های C-ترمینال، بخش‌های N-ترمینال به یک ساختار پایدار تا شوند (۴۱). این داده‌ها نشان می‌دهند که ترکیبات کدون ممکن است در محل مناسب تاخوردگی و اتصال بستر نقش داشته باشند. همچنین باید توجه داشت که واگرایی بیشتری بین کدهای ژنتیکی میتوکندری و هسته ای وجود دارد که در سنتز پروتئین موثر است. از سوی دیگر، در ژن IBDV-vVP2، میزان سوگیری استفاده از کدون مترادف بر ساختار ثانویه پروتئین تأثیر می‌گذارد و نتایج، دیدگاه‌های مفیدی را برای درک تکامل و پاتوژنز ارائه می‌کنند (۴۳). در این مطالعه، وضعیت استفاده از کدون، کدون‌های کمیاب و ارتباط بین محل‌های اتصال بستر و کدون‌های کمیاب مشخص شد. این داده‌ها یک نمای کلی از این آنزیم به عنوان تکامل و ساختار پروتئین ارائه می‌دهند. با این حال، برای ارزیابی مکانیسم‌های فعالیت آنزیم، یک رویکرد جهش‌زایی جهت‌دهی مکان است. در این مطالعه برخی از این باقیمانده‌های حیاتی شناسایی شدند. اما توضیح نتایج نیازمند مقایسه این ژن و مدل ساختاری با سایر ساختارها و ژن‌ها و تجزیه و تحلیل آزمایشگاهی

### منابع

1. Agashe D, Martinez-Gomez NC, Drummond DA, et al. 2013. Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. *30(3):549-60*.
2. Anderson RJ, Weng Z, Campbell RK, et al. 2005. Main-chain conformational tendencies of amino acids. *60(4):679-89*.
3. Auld DS, Lovell S, Thorne N, et al. 2010. Molecular basis for the high-affinity binding and stabilization of firefly luciferase by PTC124. *Proc. Natl. Acad. Sci. 107(11):4878-83*.
4. Bird CD, Emery NJ. 2009. Insightful problem solving and creative tool modification by captive nontool-

- using rooks. *Proc. Natl. Acad. Sci. U. S. A.*106(25):10370-5.
5. Chappelle EW, Picciolo GL. 1975. Analytical applications of bioluminescence and chemiluminescence.
  6. Clarke IV TF, Clark PL. 2008. Rare codons cluster. *PloS one.*3(10).
  7. Daniel E, Onwukwe GU, Wierenga RK, et al. 2015. ATGme: Open-source web application for rare codon identification and custom DNA sequence optimization. *BMC bioinformatics.*16(1):303.
  8. DeLano WLJCNopc. 2002. Pymol: An open-source molecular graphics tool.40(1):82-92.
  9. DeLuca M, McElroy WD. 1974. Kinetics of the firefly luciferase catalyzed reactions. *Biochemistry.*13(5):921-5.
  10. Eed HR, Abdel-Kader NS, El Tahan MH, et al. 2016. Bioluminescence-sensing assay for microbial growth recognition.2016.
  11. Emamzadeh AR, Hosseinkhani S, Sadeghizadeh M, et al. 2006. cDNA cloning, expression and homology modeling of a luciferase from the firefly *Lampyroidea maculata*. *BMB Reports.*39(5):578-85.
  12. Fattahi M, Malekpour A, Mortazavi M, et al. 2014. The characteristics of rare codon clusters in the genome and proteins of hepatitis C virus; a bioinformatics look.6(4):214.
  13. Fuhrmann M, Hausherr A, Ferbitz L, et al. 2004. Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene. *Plant molecular biology.*55(6):869-81.
  14. Gabriel GV, Viviani VR. 2014. Novel application of pH-sensitive firefly luciferases as dual reporter genes for simultaneous ratiometric analysis of intracellular pH and gene expression/location. *Photochemical & Photobiological Sciences.*13(12):1661-70.
  15. Hosseinkhani SJP, Pharmacology. 2008. Expression and Purification of the luciferase enzyme and in Vivo ATP Assay.12(2):109-14.
  16. Hunt RC, Simhadri VL, Iandoli M, et al. 2014. Exposing synonymous mutations.30(7):308-21.
  17. Kane JF. 1995. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current opinion in biotechnology.*6(5):494-500.
  18. Kaplan W, Littlejohn TG. 2001. Swiss-PDB viewer (deep view). *Brief. Bioinform.*2(2):195-7.
  19. Kargar F, Mortazavi M, Savardashtaki A, et al. 2019. Genomic and protein structure analysis of the luciferase from the Iranian bioluminescent beetle, *Luciola sp.* *Int. J. Biol. Macromol.*124:689-98.
  20. Kimchi-Sarfaty C, Oh JM, Kim I-W, et al. 2007. A "silent" polymorphism in the MDR1 gene changes substrate specificity.315(5811):525-8.
  21. Komar AAJHmg. 2016. The Yin and Yang of codon usage.25(R2):R77-R85.
  22. Kudla G, Murray AW, Tollervey D, et al. 2009. Coding-sequence determinants of gene expression in *Escherichia coli*. *science.*324(5924):255-8.
  23. Kumar S, Stecher G, Tamura KJMb, et al. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets.33(7):1870-4.
  24. Li Y-D, Li Y-Q, Chen J-s, et al. 2006. Whole genome analysis of non-optimal codon usage in secretory signal sequences of *Streptomyces coelicolor*.85(3):225-30.
  25. Lionetto MG, Caricato R, Calisi A, et al. 2016. Biomonitoring of water and soil quality: a case study of ecotoxicological methodology application to the assessment of reclaimed agroindustrial wastewaters used for irrigation.27(1):105-12.
  26. Lorenz R, Bernhart SH, Zu Siederdisen CH, et al. 2011. ViennaRNA Package 2.0. *Algorithms Mol. Biol.*6(1):1-14.
  27. Lorenz R, Bernhart SH, Zu Siederdisen CH, et al. 2011. ViennaRNA Package 2.0.6(1):1-14.
  28. Lundin A. 2000. Use of firefly luciferase in ATP-related assays of biomass, enzymes, and metabolites. *Methods in enzymology.*305:346-70.
  29. Masood TB, Sandhya S, Chandra N, et al. 2015. CHEXVIS: a tool for molecular channel extraction and visualization. *BMC bioinformatics.*16(1):1.
  30. MiSKA W, GEIGER R. 1988. A new type of ultrasensitive bioluminogenic enzyme substrates. I. Enzyme substrates with D-luciferin as leaving group. *Biological chemistry Hoppe-Seyler.*369(1):407-12.
  31. Morris GM, Huey R, Lindstrom W, et al. 2009. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.*30(16):2785-91.
  32. Mortazavi M, Hosseinkhani SJPE, Design, Selection. 2011. Design of thermostable luciferases through arginine saturation in solvent-exposed loops.24(12):893-903.
  33. Mortazavi M, Nezafat N, Negahdaripour M, et al. 2016. In silico evaluation of rare codons and their positions in the structure of cytosine deaminase and substrate docking studies.2(2).
  34. Mortazavi M, Zarenezhad M, Gholamzadeh S, et al. 2016. Bioinformatic Identification of Rare Codon Clusters (RCCs) in HBV Genome and Evaluation of RCCs in Proteins Structure of Hepatitis B Virus. *Hepat. Mon.*16(10).
  35. Nakatsu T, Ichiyama S, Hiratake J, et al. 2006. Structural basis for the spectral difference in luciferase bioluminescence. *Nature.*440(7082):372.
  36. Niersch J, Vega-Rubín-de-Celis S, Bazarna A, et al. 2021. A BAP1 synonymous mutation results in exon skipping, loss of function and worse patient prognosis.24(3):102173.

37. Novoa EM, de Poupiana LRJTIG. 2012. Speeding with control: codon usage, tRNAs, and ribosomes. *28(11):574-81*.
38. O'Boyle NM, Banck M, James CA, et al. 2011. Open Babel: An open chemical toolbox. *Journal of cheminformatics.3(1):33*.
39. Pawlak K, Wnetrzak M, Mackiewicz D, et al. 2021. Models of genetic code structure evolution with variable number of coded labels. *210:104528*.
40. Plotkin JB, Kudla GJNRG. 2011. Synonymous but not the same: the causes and consequences of codon bias. *12(1):32-42*.
41. Sander IM, Chaney JL, Clark PLJJotACS. 2014. Expanding Anfinsen's principle: contributions of synonymous codon selection to rational protein design. *136(3):858-61*.
42. Sauna ZE, Kimchi-Sarfaty CJNRG. 2011. Understanding the contribution of synonymous mutations to human disease. *12(10):683-91*.
43. Seligmann HJB. 2018. Alignment-based and alignment-free methods converge with experimental data on amino acids coded by stop codons at split between nuclear and mitochondrial genetic codes. *167:33-46*.
44. Sharma Y, Miladi M, Dukare S, et al. 2019. A pan-cancer analysis of synonymous mutations. *10(1):1-14*.
45. Spencer PS, Siller E, Anderson JF, et al. 2012. Silent substitutions predictably alter translation elongation rates and protein folding efficiencies. *422(3):328-35*.
46. Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques.28(6):1102-4*.
47. Theodosiou A, Promponas VJ. 2012. LaTcOm: a web server for visualizing rare codon clusters in coding sequences. *Bioinformatics.28(4):591-2*.
48. Tina K, Bhadra R, Srinivasan N. 2007. PIC: protein interactions calculator. *Nucleic acids research.35(suppl 2):W473-W6*.
49. Trott O, Olson AJ. 2010. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.31(2):455-61*.
50. Vriend G. 1990. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graphics.8(1):52-6*.
51. Walsh IM, Bowman MA, Santarriaga IFS, et al. 2020. Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *117(7):3528-34*.
52. Zhang Z, Dai W, Wang Y, et al. 2013. Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). *Arch. Virol.158(1):145-54*.
53. Zuckerkandl E, Pauling L. Evolutionary divergence and convergence in proteins. *Evolving genes and proteins: Elsevier; 1965. p. 97-166*.
54. Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.31(13):3406-15*.
55. Zuker MJNar. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *31(13):3406-15*.

# Evaluation of RNA Folding, Evolutionary Relationships, and Molecular Modeling of Substrate-Enzyme Interactions in the *Luciola sp.* Luciferase by molecular docking and bioinformatic analysis

Mortazavi M.<sup>1</sup>, Torkzadeh-Mahani M.<sup>1</sup>, Rahimi M.<sup>1</sup> and Riahi-Madvar A.<sup>2</sup>

<sup>1</sup>Dept. of Biotechnology, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, I.R. of Iran.

<sup>2</sup>Dept. of Molecular and Cell Biology, Faculty of Basic Sciences, Kosar University of Bojnord, Bojnord, I.R. of Iran.

## Abstract

Luciferases are enzymes that are involved in the emission of light, are widely used in the field of industrial and medical biotechnology. We previously reported the cloning of the luciferase gene from the Iranian firefly, *Luciola sp.* Using computational methods, we analyzed the hidden layer of genetic and structural data in the *Luciola sp.* For this, with the help of ACUA software, the ENC and CBI were studied and the AT1, GC1, AT2, GC2, AT3, and GC3 contents were calculated. In the following, the UNAFold and Mfold Servers were used to predict RNA folding. The 3D model analysis and ChExVis method were used for extracting, storing, and analyzing luciferase channels. Finally, *in silico* substrate docking was conducted in AutoDock Vina. The GC3 Skewness and GC3% of the cds were calculated as 0.39 and 17.153, respectively. This analysis shows that there are some rare codons of Pro, Arg, Ile, and Leu. The docking of 5'-O-[N-(Dehydroluciferyl)-sulfamoyl]-adenosine with luciferase was conducted, and a network of interactions was identified. In the RNAfold applet, the best secondary structure has a free energy of -397.70 kcal/mol, and the thermodynamic free energy is -427.28 kcal/mol. In the ensemble, the frequency of the MFE structure is 0.00 % and the diversity of the ensemble is 477.94. In this study, some hidden biological information from the genetic code was identified, which may have a function in the proper folding and substrate-binding site, and shows that codon compositions may have roles in the proper folding of this enzyme.