

پیشگویی نرم افزاری مناطق مجاز پذیرش پپتیدهای بیگانه در زیر واحد CstH پیلی CS3 اشریشیا کلی

وجیهه اسکندری^۱، باقر یخچالی^{۱*} و زرین مینوچهر^۱

^۱ تهران، پژوهشگاه ملی مهندسی ژنتیک و زیست فناوری، پژوهشکده زیست فناوری صنعت و محیط زیست

^۲ زنجان، دانشگاه زنجان، دانشکده علوم، گروه زیست شناسی

تاریخ دریافت: ۸۸/۱/۲۰ تاریخ پذیرش: ۸۸/۱۰/۷

چکیده

عرضه پپتیدها روی سطح باکتریها کاربردهای فراوان و مهیجی در بیوتکنولوژی و پزشکی دارد. فیمبریه باکتریها از جمله پیلی CS3 اشریشیاکلی کاندیدای مناسبی جهت ارائه اپی توپها در سطح سلول باکتریایی است. مناطق مجاز این پروتئین، با قابلیت پذیرش پپتیدهای خارجی بدون ایجاد تغییرات و اثرات قابل توجه در عملکرد و ساختار پروتئین، می تواند برای وارد کردن پپتید خارجی استفاده شود. در این مطالعه مناطق مجاز زیر واحد اصلی (CstH) پیلی CS3 برای ورود پپتیدهای خارجی به کمک روشهای پیشگویی ساختار دوم پروتئین، ارزیابی وضعیت پروتئین از لحاظ احتمال بیان به صورت اینکلوزن بادی، تعیین منحنیهای مسیر آب دوستی پروتئین، تطابق توالی مورد بررسی با توالیهای موجود در بانکهای اطلاعات زیستی، ارائه مدلهای ساختار سوم از پروتئین و آنالیز مدلهای حاصل از لحاظ مناطق قابل دسترس با نرم افزارهای مربوطه، به ترتیب اسید آمینه های سطحی هگزا هستیدین، پپتید غنی از سیستئین (قابلیت اتصال به فلزات)، سم مقاوم به حرارت اشریشیاکلی مولد انتروتوکسین (heat stable enterotoxin) و شاخص آنتی ژنیک انتروتوکسین NSP4 روتاویروس در سطح *E. coli* مطابقت دارد.

واژه های کلیدی: Bacterial surface display, *E. coli*, CS3 fimbriae, Bioinformatics Tools

* نویسنده مسئول، تلفن: ۰۲۱-۴۴۵۸۰۳۵۳ پست الکترونیکی: bahar@nigeb.ac.ir

مقدمه

ساندویچی به پروتئین سطحی، در سطح سلول باکتری ارائه شوند (۲۵).

نمایش سطحی باکتریایی کاربردهای فراوانی در توسعه واکسنهای زنده، غربال کتابخانه های پپتیدی، ساخت بیوکاتالیستها و جاذبه های بیولوژیکی دارد (۱۰، ۱۶ و ۳۴).

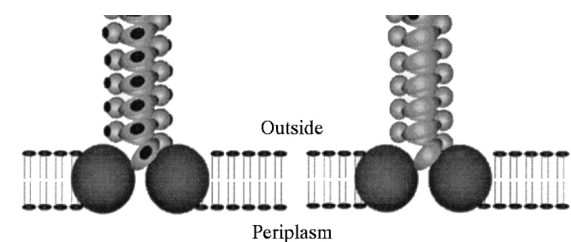
در باکتریهای گرم منفی پروتئینهای سطحی مختلفی شامل پروتئینهای غشای خارجی مانند LamB, PhoE و OmpA (۱۰) لیوپروتئین TraT و لیوپروتئینهای متصل به پپتیدوگلیکانها (۱۰)، پروتئین فلاژلین تازه باکتریایی (۱۹) و (۴۰) و پروتئین فیمبرین تاژک باکتری (۲۱ و ۳۲) برای

مهندسی پروتئین افق روشنی را جهت رسیدن به موفقیتهای بزرگ در علوم پزشکی و صنعت فراهم آورده است که عرضه سطحی باکتریایی (Bacterial Cell Surface Display) یکی از این تکنیکها برای دستکاری پروتئین در محیط *in vitro* محسوب می شود. این تکنیک این امکان را فراهم می نماید که بتوان پپتیدها یا پروتئینها را در سطح سلولهای میکروبی به وسیله اتصال به پروتئینهای سطحی میکروبی ارائه نمود (۹، ۱۲ و ۳۸).

پروتئینهای نمایشی (پروتئین مسافر) می توانند از طریق اتصال به انتهای آمینی، انتهای کربوکسیلی و یا اتصال

سودمند است و امروزه این اطلاعات توسط ابزارها و روشهای بیوانفورماتیکی استخراج می شوند.

در این مطالعه با کمک ابزارها و روشهای بیوانفورماتیکی مناطق مجاز زیرواحد CstH پپلی CS3 باکتری اشریشیاکلی بررسی و پیشگویی شده است.



شکل ۱- فیمبریه *E. coli*، مدلی جهت نمایش پپتیدهای هترولوگ به وسیله زیرواحد اصلی. پروتئین طبیعی (سمت راست) و پروتئین هیبرید حاوی پپتید هترولوگ (سمت چپ)

مواد و روشها

بانکهای اطلاعاتی یا پایگاه های داده (Databases):

توالی اسیدهای آمینه زیرواحد اصلی پپلی CS3 از بانک اطلاعات NCBI با شماره جستجوی X16944 (۱۸) و نیز از بانک اطلاعات توالی پروتئین Uni-Prot KB/Swiss-Prot (۵) با شماره جستجوی P15488 به دست آمد.

ساختار سه بعدی پروتئین در بانک اطلاعاتی طبقه بندی ساختار پروتئینها، PDB (۸) جستجو شد.

تقسیم بندی تاخوردگی یا فولد (Fold) پروتئین در بانک اطلاعاتی SCOP (۲) جستجو شد.

سرورها (Servers): توالیهای همردیف توالی زیرواحد اصلی فیمبریه CS3 با استفاده از سرور و برنامه جستجوی blast (۳ و ۴) در NCBI جستجو شد.

از سرور CPH Models (۱۵ و ۲۳) برای جستجوی پروتئینهای همولوگ CstH استفاده شد.

سرور PHYRE (۷ و ۳۷) به منظور یافتن پروتئینهای با تاخوردگی مشابه با پروتئین مورد بررسی، استفاده شد.

نمایش سطحی پپتیدها در سطح باکتری به کار برده شده است.

انواع مختلفی از فیمبریهها به خصوص در باکتریهای گرم منفی شناسایی شده است، فیمبریهها ارگانلهای اتصال باکتریها هستند که آنها را قادر می سازد به بافتهای خاص میزبان متصل شوند. این ساختارهای نخ مانند طویل به تعداد بیش از ۵۰۰ عدد در سطح سلول وجود دارد (۱۶) و (۳۴). ساختار یک فیمبریه شامل صدها زیرواحد یکسان و اصلی است که از طریق اندرکنش اسیدهای آمینه N و C ترمینال روی هم قرار گرفته و حالت استوانه ای شکل پپلی را بوجود می آورند (۳۶) (شکل ۱).

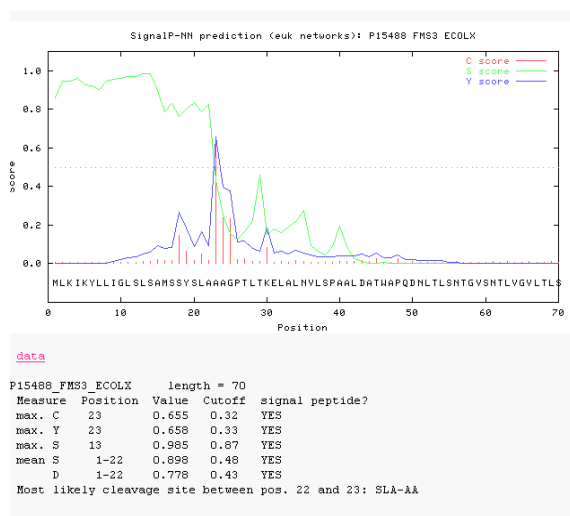
پپلی CS3 که به وسیله بسیاری از سویه های اشریشیاکلی مولد انترتوکسین بیان می شود، به عنوان یک سیستم مناسب عرضه سطحی باکتریایی برای نمایش پپتیدهای هترولوگ مانند اپی توپهای متصل شونده به فلزات استفاده شده است. ساختار این پپلی مانند سایر پپلیها شامل صدها زیر واحد اصلی و یکسان است که CstH نامیده می شود (۴۲) و استراتژی به کارگیری آن مشابه سایر پروتئینهایی که بدین منظور استفاده می شود، شامل وارد کردن توالی هترولوگ درون محلهای مجاز و ارائه آن در سطح می باشد (۱۶ و ۳۲) (مناطق مجاز محلهای هستند که پپتیدهای خارجی را بدون اینکه مداخله ای در ساخت و عملکرد پروتئین داشته باشد، می پذیرند).

همچنان که ذکر شد اسیدهای آمینه ناحیه N و C ترمینال زیرواحد اصلی در ساخت پپلی درگیر هستند ولی بقیه مناطق با توجه به ماهیت اسیدهای آمینه آن می توانند در دسترس محیط و یا مخفی از آن باشد که این مناطق کاندیدایی برای مناطق مجاز به شمار می روند. لذا به نظر می رسد آگاهی از مسیر آب دوستی پروتئین، ساختار دوم و نیز ساختار سوم زیرواحد CstH برای شناسایی و یافتن مناطقی که بتوانند در سطح پپلی و در دسترس محیط قرار گیرند و پذیرای پپتید خارجی باشند بسیار ضروری و

ارزیابی شد.

منحنی آب دوستی پروتئین: با کمک سرور pep window منحنی آب دوستی هریک از توالی‌ها به صورت نمودار نشان داده شدند.

شناسایی و ارائه الگو با کمک سرورهای شناساگر فولد (Fold) پروتئین: با ارسال توالی پروتئین به سرور شناساگر فولد (Fold) پروتئین (PHYRE) مدل سه بعدی برای توالی ارائه شد.



شکل ۲- توالی نشانه زیر واحد اصلی (CstH) فیمبریبه CS3

نتایج

جهت پیشگویی مناطق مجاز، داده‌ها و اطلاعات حاصل از سرورها و نرم افزارها، در سطوح مختلف جمع آوری و با یکدیگر مقایسه شدند که به شرح زیر می‌باشند:

پیشنهاد مناطق مجاز بر اساس ساختار دوم و منحنیهای آب دوستی: عقیده بر این است که احتمال در سطح قرارگیری اسیدهای آمینه‌ای که آرایش پیچیده دارند، بیشتر است (۱۴، ۱۶، ۲۶ و ۲۸) لذا پس از تعیین توالی نشانه پروتئین CstH توسط سرور Signal P 3.0 (شکل ۲) و به دست آوردن توالی پروتئین بالغ، ساختار دوم پروتئین توسط سرورهای مختلف پیشگویی شد. جهت یافتن مناطق با آرایش پیچیده، نتایج پیشگویی ساختار دوم توسط نرم

همردیفی و مقایسه توالیها با کمک سرورهای Fasta3 (۲۹) و BLAST (۳ و ۲۲) انجام گرفت.

توالی نشانه پروتئین با کمک سرور SignalP3.0 (۶) شناسایی شد.

آب دوستی پروتئین به وسیله سرور Pep window (۳۳) تعیین شد.

ساختار دوم پروتئین با سرورهای PHD (۳۱)، SABLE (۱) و GORIV (۱۳) پیشگویی شد.

نکته: برای پارامترهای مختلف به جز مواردی که ذکر شده، از پیش فرض سرورها استفاده شده است.

نرم افزارها: از نمایشگر Swiss PDB viewer (۲۰) و Discovery studio (۲۷) برای مشاهده و تجزیه و تحلیل مدل‌های حاصل استفاده شد.

پیشگویی ساختار دوم: ساختار دوم توالی CstH و مشتقات آن به طور جداگانه با سرورهای ذکر شده در سه سطح مارپیچ (helix)، پیچه یا کویل (coil) و رشته (strand) پیشگویی شدند و با مقایسه نتایج پیشگویی‌ها، ساختار دوم توافقی به دست آمد. در ساختار حاصل اگر تفاوتی در وضعیت ساختار دوم پیشگویی شده به وسیله سرورها برای یک اسید آمینه وجود داشت حالتی که حداقل به وسیله دو سرور، پیشگویی شده بود به عنوان حالت توافقی انتخاب شد و در غیر این صورت آن حالت با حرف U نمایش داده شده است.

پیشگویی توالی نشانه: حضور و موقعیت برش توالی نشانه به کمک سرور Signal p 3.0 پیشگویی شد. این سرور بر مبنای مدل‌های شبکه عصبی مصنوعی و مدل مارکف مخفی (Hidden Markov Model) بنا گردیده است.

احتمال تشکیل اینکلوزن بادی: احتمال تشکیل اینکلوزن بادی پروتئینها با کمک سرور pep state (۳۰) آنالیز و

افزارهای مختلف با یکدیگر مقایسه شدند (شکل ۳).

پیش گویی ساختار دوم زیرواحد اصلی فیمبریه CS3 با سرور PHD

```
AA |AAGPTLTKEALNVLSPAALDATWAPQDNLTLNNTGVSNTLVGVLTLSNTSIDTVSIAST|
PHD | EEEEEEEEE EEE EEEE EEEEEEEEE EEEEEEE |
Rel | 999741356552123542120332476715645675111468999962674146999732 |

AA |NVSDTSKNGTVTFAHETNNSASFATTISTDNANITLDKNAGNTIVKTTNGSQLPTNLPLK|
PHD | EEEEEEE E EEEE EEE EEEEE E |
Rel | 745756686168767336788422114324788335415895288537994442665753 |

AA |FITTEGNEHLVSGNYRANITITSTIK|
PHD |EEEE EEEE EEEEEEEEE |
Rel | 77626722323357634787766779
```

پیش گویی ساختار دوم زیرواحد اصلی فیمبریه CS3 با سرور GOR IV

```
AAGPTLTKEALNVLSPAALDATWAPQDNLTLNNTGVSNTLVGVLTLSNTSIDTVSIAST
Cccccchhhhhhccccchhhccccccccceccccccceeeeeeeccccccceeeeeec

NVSDTSKNGTVTFAHETNNSASFATTISTDNANITLDKNAGNTIVKTTNGSQLPTNLPLK
Cccccchhhhhhccccchhhccccccccceccccccceeeeeeeccccccceeeeeec

FITTEGNEHLVSGNYRANITITSTIK
eeccccccceccccccceeeEeeeeec
```

پیش گویی ساختار دوم CstH با سرور SABLE

```
AAGPTLTKEALNVLSPAALDATWAPQDNLTLNNTGVSNTLVGVLTLSNTSIDTVSIAST 60
CCCCCHHHHHHHHHCCCCCCCCCCCCCCCCCECCCCCCCCCEEEEEEECCCCCEEEEEEC
678757688888765654445555777765446776654446666655666545666555

NVSDTSKNGTVTFAHETNNSASFATTISTDNANITLDKNAGNTIVKTTNGSQLPTNLPLK 120
CCCCCCCCCEEEEEEECCCCCCCCCEEEEEEECCCCCEEEEEEECCCCCEEEEEEECCCCCEEEE
77766788758877547787655566577866787467887567657887567766467

FITTEGNEHLVSGNYRANITITSTIK 146
EECCCCCEEECCCEEEEEEEEEEEEC
77657875565577566778776655
```

پیش گویی در دسترس قرارگیری نسبی، زیرواحد اصلی فیمبریه CS3 با سرور SABLE

```
> 1 60
AAGPTLTKEALNVLSPAALDATWAPQDNLTLNNTGVSNTLVGVLTLSNTSIDTVSIAST
443441233111311232112121134321213333122211100112332132121123
33344653355636531344342555233533222351146436536222333524421

> 61 120
NVSDTSKNGTVTFAHETNNSASFATTISTDNANITLDKNAGNTIVKTTNGSQLPTNLPLK
423433542212112324432311221224313121334233211322443412331213
231134412464644332322253326323243646452314365333214253326334
```

> 121 146
 FITTEGNEHLVSGNYRANITITSTIK
 10224333211323131212023322
 66435244357114634264632255

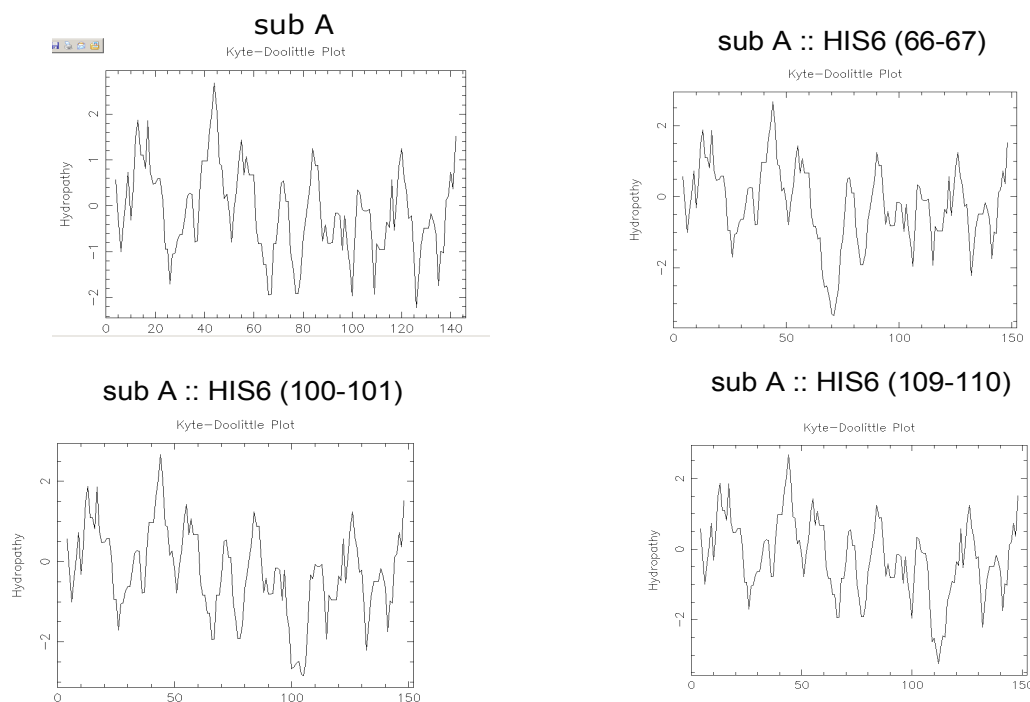
شکل ۳- پیش گویی ساختار دوم زیرواحد اصلی فیمبریه CS3 توسط نرم افزارهای مختلف و پیش گویی در دسترس قرارگیری آن توسط نرم افزار SABLE (خط دوم میزان در دسترس و یا مخفی بودن را نشان می دهد و خط سوم بیانگر تراز اطمینان می باشد.)

آمینو ۶۶-۶۷، ۱۰۰-۱۰۱ و ۱۰۹-۱۱۰ پروتئین CstH بالغ قرار دارند.

سپس لینکر هگزا هیستیدینی در این مناطق وارد شده و ساختار دوم و نیز منحنی آب دوستی (شکل ۴) آنها آنالیز گردید. همچنانکه ملاحظه می شود پس از وارد کردن لینکر هگزا هیستیدینی در مناطق فوق الذکر، میزان آب دوستی پروتئین در آن مناطق اندکی نیز افزایش می یابد. لذا با لحاظ کردن نتایج حاصل از ساختار دوم و منحنیهای آب دوستی، به ترتیب نواحی ۶۶-۶۷، ۱۰۰-۱۰۱ و ۱۰۹-۱۱۰ به عنوان کاندیدایی مناطق مجاز پذیرش پپتید خارجی (لینکر هگزا هیستیدینی) انتخاب شدند.

از طرف دیگر احتمال در سطح قرارگیری اسیدهای آمینه باردار یا قطبی در سطح پروتئین منطقی بنظر می رسد، بنابراین منحنی آب دوستی پروتئین و در دسترس بودن نسبی آبی پروتئین تعیین و با ساختارهای دوم پیشگویی شده، مقایسه شدند. آن دسته مناطقی که به وسیله سرورهای پیشگویی ساختار دوم به عنوان پیچ پیکه پیشگویی شده و در منحنیهای آب دوستی نیز نسبت به مناطق مجاور آب دوست تر بودند به عنوان کاندیدای مناسبی جهت مناطق مجاز و وارد کردن پپتیدهای خارجی انتخاب شدند که این مناطق به ترتیب الویت در ناحیه اسیدهای

Hydropathy plot



شکل ۴- منحنی آب دوستی زیرواحد اصلی و مشتقات آن (اعداد داخل پرانتز موقعیت ورود لینکر را نشان می دهد).

توالی معادل با CstH و شباهت عملکردی با فیمبریه CS3 به عنوان مدل مناسب انتخاب شد.
Lineage الگو به شرح زیر است.

- 1- Class: All beta proteins
- 2- Fold: Common fold of diphtheria toxin/transcription factors/cytochrome f sandwich; 9 strands in 2 sheet; greek-key; subclass of immunoglobulin-like fold
- 3- Super family: Bacterial adhesins
- 4- Family : Pilus subunits
- 5- Protein: F1 capsule antigen CafI
- 6- Species: Yersinia pestis

مقایسه توالی مدل و پروتئین CstH : پروتئینهایی که ساختار و عملکرد مشابه دارند لزوماً دارای توالی مشابه نیستند اما احتمال وجود ساختار مشابه در پروتئینهای دارای توالی مشابه بیشتر است. لذا جهت تعیین شباهت بین توالی مدل (الگو) و توالی هدف (CstH)، توالی اسیدآمینه پروتئین CstH به همراه هریک از مشتقات آن به سرور Fasta ارسال شدند که نتیجه مقایسه، وجود شباهت قابل قبول بین توالی الگو (کپسول F₁ آنتی ژن CafI) و توالیهای هدف مورد بررسی بود.

آنالیز استروشمیمی مدل‌های پیشگویی شده : در مرحله اول خواص استروشمیمیایی مدل‌های ارائه شده بر اساس نقشه راما چاندوران بررسی شد. همچنانکه در شکل ۵ نیز مشاهده می‌شود، تمامی اسیدهای آمینه در مناطق مجاز نقشه قرار دارند این نتایج بیانگر مناسب بودن مدل‌های حاصل می‌باشد.

مرحله بعدی آنالیز، بر اساس محاسبه مینیم انرژی برای اسیدهای آمینه در دسترس، اسیدهای آمینه بوجود آورنده پیچ، اسیدهای آمینه قطبی و محاسبه میدان نیرو (Force-field) انجام گرفت. همچنان که ملاحظه می‌شود امتیازات به دست آمده برای هریک از مدلها متفاوت است، و این

احتمال تشکیل اینکلوزن بادی : بیان پروتئینهای نو ترکیب در باکتری *E. coli* با احتمال تشکیل اینکلوزن بادی همراه است. لذا باید پروتئینهای نو ترکیب از لحاظ احتمال تشکیل اینکلوزن بادی در سلول مورد بررسی قرار گیرند.

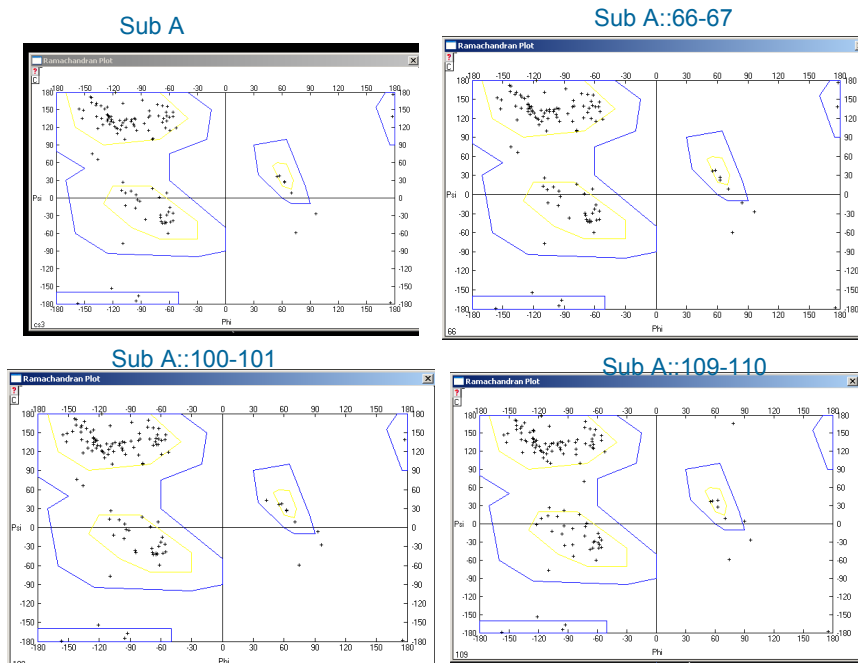
آنالیز پپتیدها با کمک سرور PEPSTATS قبل و بعد از وارد کردن لینکر هگزا هستیدینی نشان داد که عدم احتمال تشکیل اینکلوزن بادی برای توالیهای هیبرید حاوی پپتیدهای خارجی در همان حدود توالیهای دست نخورده است. لذا به نظر می‌رسد که سه موقعیت ۶۶-۶۷، ۱۰۰-۱۰۱ و ۱۰۹-۱۱۰ کاندیدایی مناسبی برای پذیرش لینکر خارجی هستند (جدول ۱).

جدول ۱- احتمال عدم تشکیل اینکلوزن بادی

Sequences	
Intact Sequence , 66-67 , 100-101 , 109-110	
Improbability of expression in inclusion bodies	0.843 , 0.818 , 0.818 , 0.818

ارائه مدل با کمک سرور شناساگر ساختار سوم پروتئین: از آنجایی که ساختار سه بعدی پروتئین CstH تعیین نشده است و آگاهی از ساختار سوم جهت آنالیز و بررسی مناطق در دسترس پروتئین و نیز پردازش آن با نرم افزارهای مشاهده و آنالیز استروشمیمیایی مفید می‌باشد، لذا ساختار سوم CstH پیشگویی شد. جهت پیشگویی ساختار سوم، ابتدا توالی طبیعی و تغییر یافته، زیر واحد اصلی پپلی به سرورهای پیشنهاد الگو بر اساس همولوژی، ارسال شدند ولی الگویی برای توالیهای ارسالی ارائه نگردید. سپس هریک از توالیها به سرور شناساگر ساختار سوم پروتئین (۲۴) ارسال و تعدادی مدل توسط سرور ارائه گردید که از میان مدل‌های ارائه شده، الگوی کپسول F₁ آنتی ژن caf₁ به سبب داشتن بیشترین شباهت توالی اسید آمینه‌ای، طول

اختلاف امتیازات در الویتهای ترتیبی پیشنهادی برای جایگاههای مجاز لحاظ گردید (جدول ۲).



شکل ۵- نقشه راما چاندران زیرواحد اصلی فیمبریه CS3 و مشتقات آن

جدول ۲- وضعیت مدل‌های ارائه شده از لحاظ انرژی

Models	Energy minimization (KJ/mol)				Compute energy (force field) (KJ/mol)
	Accessible aa (Accessible surface = 30%)	Secondary structure Coil	Secondary structure helix	Polar amino acid	Solvent
CS3 fimbrial subunit H	1973613.875	53138.633	2025180.250	239494.633	2025818.125
sub H :: HIS6 (66-67)	1979471.500	52851.223	2024797.250	240097.406	2025435.625
sub H :: HIS6 (100-101)	2355926.750	53198.875	2407823.500	259279.172	2408461.500
sub H :: HIS6 (109-110)	34270272.000	71893.797	34334860.000	54744.703	34335496.000

نشان داد که لینکرهای هگزاهیستیدینی در جایگاههای مجاز در سطح ملکول قرار می‌گیرند که بیانگر مناسب بودن جایگاههای انتخاب شده به عنوان جایگاه مجاز پذیرش پپتیدهای خارجی می‌باشد (شکل ۶).

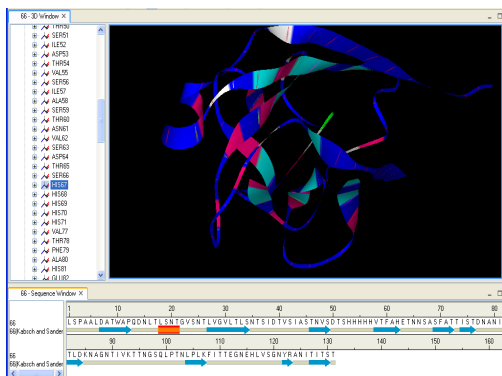
در مرحله سوم، بررسی بر پایه شکل شماتیک ارائه شده ملکول و مناطق در سطح قرار گرفته آن و وضعیت هریک از لینکرهای هگزاهیستیدینی در این مناطق، با کمک نمایشگر Discovery Studio انجام پذیرفت. نتیجه

مکول CstH، مقایسه شدند و Root-Mean-square (RMSD) آنها محاسبه گردید (جدول ۳) که RMSD های کمتر نشان دهنده شباهت بیشتر دو پروتئین مورد مقایسه است.

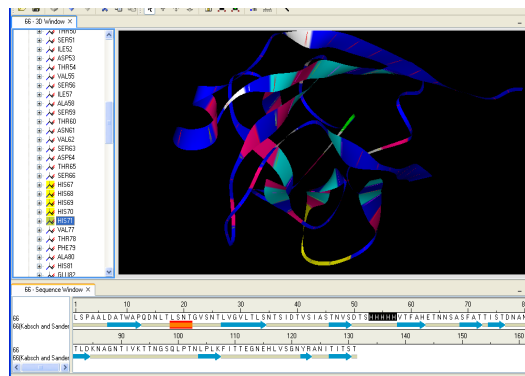
از آنجایی که مقایسه ساختارها همانند مقایسه توالیها بینش عمیقی نسبت به منشاء، مکان، انتراکشن و فعالیت یک پروتئین فراهم می کند لذا در مرحله بعدی پژوهش، ساختار سوم هریک از مدلها با مدل ارائه شده برای

۶۶-۶۷

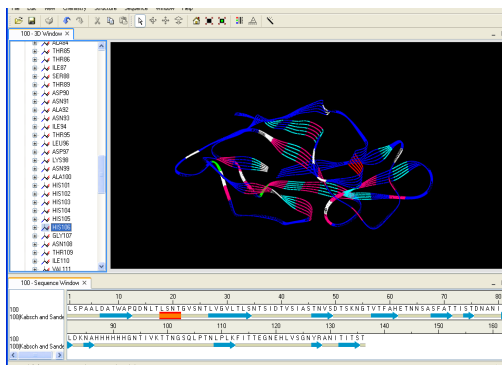
Exposed regions = Blue , buried regions Red



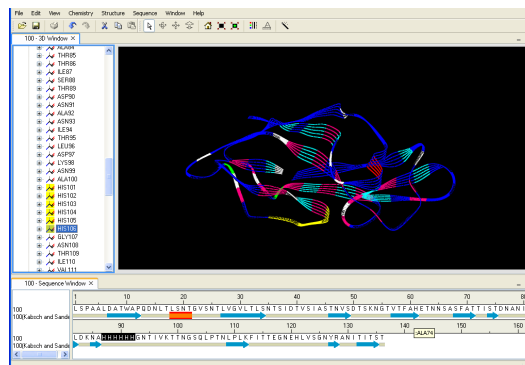
Blue = Exposed Red = Buried yellow= Selected linker



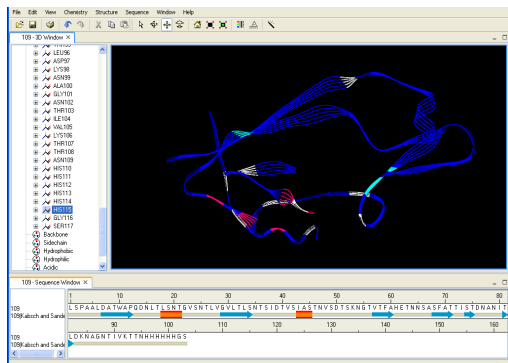
Exposed regions = Blue , buried regions Red



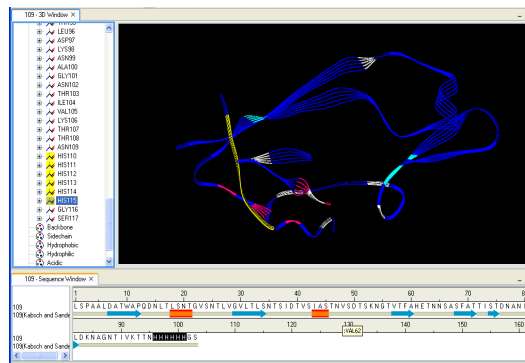
Blue = Exposed Red = Buried yellow= Selected linker



Blue = Exposed Red = Buried



Blue = Exposed Red = Buried yellow= Selected linker



شکل ۶- موقعیت لینکر هگزا هیستیدینی نسبت به مناطق در دسترس پروتئین CstH: مناطق در دسترس در این پروتئین با رنگ آبی و لینکر هگزا هیستیدینی با رنگ زرد مشخص شده است.

جدول ۳- مقایسه RMSD هریک از مدل ها نسبت به مدل زیر واحد CstH

ساختار ۱	ساختار ۲	RMSD(A°)
مدل H subunit CS3 fimbrial subunit	مدل sub H :: HIS6 (66-67)	0.39
مدل H subunit CS3 fimbrial subunit	مدل sub H :: HIS6 (100-101)	1.45
مدل H subunit CS3 fimbrial subunit	مدل sub H :: HIS6 (109-110)	4.41

انجام می‌پذیرفت و اغلب اوقات نیز استفاده از چنین مناطقی برای پپتیدهای خارجی به دلیل مطابقت نداشتن با اپی‌توپهای گزارشگر با شکست مواجه می‌شد (۴ و ۳۵) ولی اکنون با به کارگیری و انتخاب الگوریتمهای دقیق و یاری جستن از نرم افزارهای کامپیوتری، روشهای بیوانفورماتیکی جهت پیشگویی مناطق آب دوست (مناطق هیدروفیل)، ساختار دوم و ساختار سه بعدی پروتئین و پیشگویی مناطق مجاز پروتئین با اطمینان بیشتری انجام می‌شود و بدین ترتیب انجام پروژه های پیچیده مهندسی پروتئین، بر مبنای پیشگویی ساختاری، آنالیز استرئوشیمیایی و انرژی آنها با استفاده از روشهای بیوانفورماتیک و در مدت زمان بسیار کوتاه نسبت به دهه‌های گذشته دور از ذهن نیست (۱۱ و ۳۹).

در این مطالعه به کمک نرم افزارها، سرورها و روشهای پیشگویی ساختار دوم پروتئین، روشهای ارزیابی وضعیت پروتئین از لحاظ احتمال بیان به صورت اینکلوژن بادی، تعیین منحنیهای مسیر آب دوست و آب گریز پروتئین، آنالیز تطابق و هم‌ردیفی توالی مورد بررسی با توالیهای موجود در بانکهای اطلاعات زیستی، ارائه مدل‌های ساختار سوم از طریق متد شناساگر فولد (fold-recognition) و آنالیز مدل‌های حاصل از لحاظ مناطق قابل دسترس، موقعیت های ۶۶-۶۷، ۱۰۰-۱۰۱ و ۱۰۹-۱۱۰ به عنوان مناطق مجاز زیر واحد CstH پپلی Cs3 برای وارد کردن پپتیدهای خارجی پیش‌بینی شد. مقایسه نتایج این تحقیق با تجارب آزمایشگاهی قبلی برای نمایش سطحی هگزا هستیدین، پپتید غنی از سیستمین (قابل اتصال به

در نهایت همه یافته ها با یکدیگر و با نتایج حاصل از انرژی کل مقایسه شدند و بدین ترتیب موقعیتهای ۶۶-۶۷، ۱۰۰-۱۰۱ و ۱۰۹-۱۱۰ به عنوان کاندیدایی مناطق مجاز پذیرش پپتید خارجی (لینکر هگزا هستیدینی) پیشنهاد شدند.

بحث

یکی از پیشرفتهای چشمگیر در سالهای اخیر، بهره‌گیری از ابزارهای بیوانفورماتیکی جهت کسب اطلاعات با ارزش ساختاری و عملکردی پروتئینها و کاربرد آنها در مهندسی پروتئینها می‌باشد، اطلاعاتی که گاه برای یافتن آنها با استفاده از روشهای آزمایشگاهی ماهها و سالها وقت صرف می‌شد.

عرضه سطحی باکتریایی یکی از روشهای مهندسی پروتئین محسوب می‌شود که کاربردهای فراوانی در تهیه واکسنهای زنده، غربال کتابخانه‌های پپتیدی، ساخت بیوکاتالیست‌ها و جاذبه‌های زیستی دارد. جهت عرضه و نمایش پپتیدهای خارجی باید آنها را در مکانهایی از پروتئین به نام مناطق مجاز وارد کرد.

تصور می‌شود نواحی در معرض محیط (Exposed area) و آب دوست که ساختار پیچیده نیز دارند کاندیدای مناسبی، برای مناطق مجاز و وارد کردن پپتیدهای بیگانه جهت عرضه سطحی هستند (۱۶، ۲۶، ۳۵ و ۳۷).

در دهه‌های گذشته پیشگویی مناطق مجاز، با روشهای آزمایشگاهی و با صرف هزینه و زمان زیاد مانند وارد کردن پپتیدها یا اپی‌توپهای کوچک گزارشگر و بیان آنها

دارد که این یافته ها بیانگر قابل اتکا بودن نتایج مطالعات نرم‌افزاری (حداقل در این نوع پروتئینها و کاربرد مورد نظر) می باشد.

فلزات سنگین) (۳۳)، سم مقاوم به حرارت اشرشیاکلی مولد انتروتوکسین (heat stable enterotoxin) و شاخص آنتی ژنیک انتروتوکسین NSP4 روتاویروس (۱۷) نشان داد که نقاط مجاز پیش‌بینی شده با نتایج آزمایشگاهی مطابقت

منابع

- 1- Adamczak, R., Porollo, A. and Meller, J., 2004., "The SABLE server, which was rigorously " Bioinformatics
- 2- Andreeva A, Howorth D, Brenner S.E, Hubbard T.J.P, Chothia C., Murzin A.G., 2004, "SCOP database in 2004: refinements integrate structure and sequence family data." *Nucleic Acids Res*, 32: 226–229
- 3- Altschul SF, Gish W , Miller W, Myers EW, Lipman DJ., 1990, "Basic local alignment search tool". *J Mol Biol*, 215:403–410
- 4- Armstrong, S.K. and M.A. McIntosh., 1995, "Epitope insertions define functional and topological features of the Escherichia coli ferric enterobactin receptor". *J Biol Chem*, 270: 2483-8
- 5- bairoch A ,et al ,2009, " The Universal Protein Resource (UniProt)". *Nucleic Acids Research*, 37: 169-174
- 6- Bendtsen JD, N. H., vonHeijne G, Brunak S., 2004, "Improved prediction of signal peptides: SignalP 3.0". *J. Mol. Biol*, 340: 783–795
- 7- Bennett-Lovsey R.M, Herbert A.D, Sternberg M.J.,2008, " Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre". *Proteins*, 70: 611-25.
- 8- Berman H. M, Westbrook J, Feng Z, Gilliland G, Bhat T.N, Weissig H, and Shindyalov I. N , Bourne P.E., 2000, " The Protein Data Bank". *Nucleic Acids Res*, 28: 235–242
- 9- Chang H.J , Shen S.Y , Lo S.J., 1999, " Expression of foreign antigens on the surface of Escherichia coli by fusion to the outer membrane protein traT". *J Biomed Sci*, 6: 64-70.
- 10- Chen W, Georgiou G., 2002, "Cell-Surface display of heterologous proteins: From high throughput screening to environmental applications ". *Biotechnol Bioeng*, 79: 496-503
- 11- Cheng, H., 2007, "Consensus Data Mining (CDM) Protein Secondary Structure Prediction Server: combining GOR V and Fragment Database Mining (FDM) ". *Bioinformatics*, 23: 2628-30
- 12- Daugherty. PS., 2007. "Protein engineering with bacterial display". *Curr Opin Struct Biol* .17:474-80.
- 13- Garnier J, G. J.-F., Robson B. .,1996. "GOR IV GOR secondary structure prediction method version IV". *Methods in Enzymology* 266: 540-553
- 14- Grater.F, Heider p, Zangi R and Berne B. J.,2008, "Dissecting entropic coiling and poor solvent effects in protein collapse". *J .Am. Chem. Soc*, 130: 11578-9.
- 15- HardestyJ.,2005., "Protein Simulation, Function and Prediction." *Computational Structural Biology*
- 16- Hedegaard L, Klemm P., 1989, "Type 1 fimbriae of Escherichia coli as carriers of heterologous antigenic sequences". *Gene*, 85: 115-24
- 17- Hosseini S.A, Yakhchali B., 2002, "Construction and overexpression of a novel hybride protein consisting of Rotavirus enterotoxin NSP4 and subunit of CS3 pili in E.coli ". *Iranian Journal of Biology* 12:1-12
- 18- Jalajakumari M.B , Thomas CJ and Manning PA., 1989, "Genes for biosynthesis and assembly of CS3 pili of CFA/II enterotoxigenic Escherichia coli: novel regulation of pilus production by bypassing an amber codon". *Mol Microbiol*, 3: 1685-95.
- 19- Kjaergaard K, Schembri M. A and Klemm P., 2001, "Novel Zn²⁺-chelating peptides selected from a fimbria-displayed random peptide library". *Appl Environ Microbiol* , 67: 5467–5473
- 20- Kaplan W and littlejohn T.G ., 2001, "Swiss-PDB Viewer (Deep View) ". *Brief Bioinform* , 2: 195–197
- 21- Klemm. P and Schembri M. A ., 2000, "Fimbrial surface display systems in bacteria: from vaccines to random libraries". *Microbiology*, 146: 3025-3032

- 22- Korf. I., 2003, "Serial BLAST searching ". *Bioinformatics*, 19 : 1492-1496
- 23- Krieger.E., Nabuurs.S.B. and Vriend.G.,2003, " Homology Modeling. " *Methods Biochem Anal*, 44: 509-23.
- 24- Kroemer. R.T, Robinson AJ D. S and Richards W.G .,1996, "Prediction of the three-dimensional structure of human interleukin-7 by homology modeling". *Protein Eng*, 9 : 493-8
- 25- Lee S.Y., Choi J.H. and Xu. Z., 2003, "Microbial cell-surface display ". *Trends Biotechnol*, 21: 45-52
- 26- Macdonald J.R and W.C. Johnson, Jr., 2001, "Environmental features are important in determining protein secondary structure". *Protein Sci*, 10: 1172-7
- 27- Niemeyer E. A.,2000, "Application GuideTools and methods used in Discovery Studio® for the visualization". *Nucleic Acids Research*, 28: 235-242
- 28- Othon C.M., 2009, "Solvation in protein (un)folding of melittin tetramer-monomer transition". *Proc Natl Acad Sci U S A*, 106:12593-8
- 29- Pearson, W. R., 2000, "Flexible sequence similarity searching with the FASTA3 program package ". *Methods Mol Biol*, 132: 185-219
- 30- Roger G.H., 2000, "Expression of soluble heterologous proteins via fusion with NusA protein". in *inNovations*, 11: 4 - 7
- 31- Rost B, Sander C and Schneider R., 1994, "PHD--an automatic mail server for protein secondary structure prediction". *Comput Appl Biosci*, 10: 53-60
- 32- Sifār B, Yakhchali B, Arbabi M., 2005, "Enhanced bioadsorption of cadmium and nickel byE. coli displaying a metal binding motif using CS3 fimbriae". *IRANIAN JOURNAL of BIOTECHNOLOGY*, 3: 180-186
- 33- Sanger.et al., 2001, "pepwindow Displays protein hydropathy Sanger." *Nucleic Acids Research*, 29: 1185-90
- 34- Schembri P. K and Schembri M. A., 2000, "Fimbrial surface display systems in bacteria: from vaccines to random libraries". *Microbiology*, 146: 3025-3032
- 35- Schlehuber L.D and Rose J.K., 2004, "Prediction and identification of a permissive epitope insertion site in the vesicular stomatitis virus glycoprotein ". *J Virol*, 78: 5079-87
- 36- Soto G.E and Hultgren S.J., 1999, "Bacterial adhesins: common themes and variations in architecture and assembly" . *J Bacteriol*, 181: 1059-71
- 37- Sujatha M.S and Balaji P.V., 2006, "Fold-recognition and comparative modeling of human alpha2,3-sialyltransferases reveal their sequence and structural similarities to CstII from *Campylobacter jejuni*". *BMC Structural Biology*, 6
- 38- Stahl S and Uhlen M., 1997, *Bacterial surface display: trends and progress*. *Trends Biotechnol.*, 15: 185-92
- 39- Vivona S , et al., 2008, "Computer-aided biotechnology: from immuno-informatics to reverse vaccinology". *Trends Biotechnol*, 26: 190-200
- 40- Westerlund W. B., 2000, "Peptide display on bacterial flagella: principles and applications". *Int J Med Microbiol*, 290: 223-30
- 41- WR P.,1990, "Rapid and sensitive sequence comparison with FASTP and FASTA". *Methods Enzymol*, 183: 63-98
- 42- Yakhchali B, Manning PA ., 1997, "Epitope analysis of the CS3 fimbrial subunit of human Enterotoxigenic *Escherichia coli* and the construction of novel CS3::ST and CS3::LT-B immunogens". *Behring Inst Mitt*, 98:124-134
- 43- Ye J , McGinnis S, and Madden T.L ., 2006, "BLAST: improvements for better sequence analysis". *Nucleic Acids Res*, 34 : 6-9

اطلاعات تکمیلی:

<http://scop.mrc-lmb.cam.ac.uk/scop/>

<http://www.ncbi.nlm.nih.gov/BLAST>

<http://www.cbs.dtu.dk/services/CPHmodels/>

Fasta3 : <http://www.ebi.ac.uk/fasta3/>

BLAST2 : <http://www.ebi.ac.uk/blastall/>

از سرورها و سایتهای علمی زیر برای پردازش اطلاعات مورد نظر استفاده شد.

<http://www.ncbi.nlm.nih.gov>

<http://www.rcsb.org>

<http://www.uniprot.org>

SABLE (Solvent AccessiBiLitiEs, secondary structures) <http://sable.cchmc.org> http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_gor4.html

<http://www.cbs.dtu.dk/services/SignalP/>

<http://mobylye.pasteur.fr/cgi-bin/MobylyePortal/portal.py?form=pepwindow>

http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_phd.html

In-silico prediction of permissive sites of CstH subunit of *E. coli* CS3 pili for insertion of foreign peptides

Eskandari V.^{1,2}, Yakhchali B.¹ and Minucmehr Z.¹

¹ National institute of Genetic Engineering and Biotechnology, Industrial and Environmental Biotechnology Department, Tehran, I.R. of IRAN

² Biology Dept., Faculty of Science, Zanjan University, Zanjan, I.R. of IRAN

Abstract

Display of peptides on the surface of bacteria have many applications in medicine and biotechnology. Bacterial fimbriae including CS3 pili of *E. coli* are suitable candidates for expression of foreign epitopes on the surface of bacteria. Permissive sites of this protein, with capability to accept foreign peptides without changes and considerable effects on the structure and function of the protein can be used for expression of foreign peptides. In this study, the permissive sites of the CstH subunit of the CS3 pili were predicted at amino acids 66-67, 100-101 and 109-110 of the mature protein using secondary structure prediction procedures, prediction of disulfide bonds, determination of hydrophathy of the protein, alignment of the protein sequence with the sequences in the databases, tertiary structure (modeling) and analysis of the resulted modes for accessible regions. The results of the in-silico studies were in agreement with the previous laboratory works including display of hexahistidine, cystein reach peptide, ST toxin of *enterotoxigenic E. coli* and an epitope of *Rotavirus* NSP4 in this system.

Keywords: Permissive site, CS3 pili, *E. coli*, hybrid protein, bacterial display