

## تعیین ساختار سوم پروتئین HMGB4 انسانی با استفاده از دو روش مبتنی بر همولوژی -

### مدلینگ

صفا لطفی\*، مجتبی مرتضوی و علی ریاحی مدوار

کرمان، دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته، پژوهشگاه علوم و تکنولوژی پیشرفته و علوم محیطی، گروه بیوتکنولوژی

تاریخ دریافت: ۹۷/۲/۲۹ تاریخ پذیرش: ۹۷/۷/۱۷

### چکیده

پروتئین HMGB4 که در سال ۲۰۰۸ شناخته شد از اعضای خانواده پروتئین‌های غیرهیستونی HMGB پستانداران است. پروتئین‌های HMGB از طریق دو موتیف که بترتیب HMGB-box A و HMGB-box B نامیده می‌شوند به DNA اتصال می‌یابند. پروتئین‌های HMGB به DNA تغییر یافته با سیس‌پلاتین اتصال یافته و مانع دسترسی عوامل ترمیم برشی DNA به این نقاط می‌شوند. HMGB4 در مقایسه با HMGB1 عضو دیگر این خانواده با میل ترکیبی بالاتری به این نوع DNA اتصال یافته و با قدرت بیشتری ترمیم این نقاط آسیب‌دیده را مهار می‌نماید. بنظر می‌رسد که حساسیت بسیار بالای تومورهای سلولهای زایای بیضه نسبت به سیس‌پلاتین ناشی از بیان بسیار بالای HMGB4 در بافت بیضه باشد. در این مطالعه ساختار سوم HMGB4 انسانی با استفاده از نرم‌افزار مدلر و سرور I-TASSER که هردو بر پایه همولوژی مدلینگ عمل می‌نمایند تعیین گردید. نتایج نشان می‌دهد که مدل‌های ایجاد شده توسط هردو روش از کیفیت و پایداری مناسبی برخوردار هستند. هرچند مدل I-TASSER در مقایسه با مدلر تفاوت مشاهده شده در ویژگی‌های اتصال HMGB1 و HMGB4 به DNA پلاتینه را بهتر توجیه می‌نماید. بر اساس نتایج حاصل، والین ۱۷ و فنیل‌آلانین ۳۸ از HMGB-box A و لوسین ۱۰۱ و والین ۱۲۰ از HMGB-box B بنیان‌هایی هستند که با وارد شدن بین جفت بازهای شیار کوچک DNA موجب اتصال پروتئین HMGB4 به DNA می‌شوند. همچنین نتایج نشان می‌دهد که بخش انتهایی کربوکسیل HMGB4 فاقد ساختار سوم منظم می‌باشد. نتایج حاصل می‌تواند به فهم بهتر ارتباط ساختار و عملکرد این پروتئین و همچنین طراحی داروهای جدید ضدسرطان کمک شایانی نماید.

واژه‌های کلیدی: پروتئین HMGB4، سیس‌پلاتین، همولوژی مدلینگ، نرم افزار مدلر، سرور I-TASSER

\* نویسنده مسئول، تلفن: ۰۳۴۳۳۷۶۶۱۱، پست الکترونیکی: safalotfi@ut.ac.ir

### مقدمه

پروتئین‌های HMG (High Mobility Group) (HMG2)، HMGB3 و HMGB4 است. پروتئین‌های HMGB1-3 با وزن مولکولی تقریباً ۲۵ کیلودالتون، از دو دومین اتصال یابنده به DNA به نام HMGB box-A (باکس A) و HMGB box-B (باکس B) و یک دم اسیدی طویل در انتهای کربوکسیل تشکیل شده اند (۲۵، ۲۶ و ۳۹). پروتئین HMGB4 با وزن تقریبی ۲۲ کیلودالتون که اولین بار در سال ۲۰۰۸ شناخته شد تنها عضو این خانواده است که فاقد دم اسیدی می‌باشد (۶). اولین بار مطالعات انجام گرفته در موش نشان داد که بیان

پروتئین‌های HMG (High Mobility Group) (HMG) که اولین بار در سال ۱۹۷۳ شناسایی شدند بزرگترین گروه پروتئین‌های غیرهیستونی کروماتین را تشکیل می‌دهند (۱۲). پروتئین‌های HMGB که فراوان ترین گروه پروتئین‌های HMG در پستانداران می‌باشند در تنظیم فعالیت‌های متعددی نظیر نسخه برداری، همانندسازی، نوترکیبی V(D)J و ترمیم DNA مشارکت می‌نمایند (۳، ۴، ۷ و ۲۹). خانواده پروتئین‌های HMGB متشکل از پروتئین‌های HMGB1 (HMG1، آمفوترین)،

بمنظور تعیین ساختار سوم پروتئین HMGB4 انسانی، ابتدا توالی این پروتئین با فرمت FASTA از پایگاه اطلاعاتی UniProt دانلود گردید. UniProt که یک پایگاه اطلاعاتی با کیفیت بسیار بالا است بطور رایگان قابل دسترس می‌باشد و اطلاعات کاملی را از عملکرد و توالی پروتئین‌ها ارائه می‌دهد. بسیاری از ورودی‌های اطلاعاتی این سایت از پروژه‌های تعیین توالی ژنوم مشتق شده است. این سایت مقادیر بالایی از اطلاعات را در مورد عملکرد زیستی پروتئین‌ها که از نتایج حاصل از تحقیقات علمی بدست آمده‌اند در اختیار محققین قرار می‌دهد (۳۰).

**بررسی کلی ساختار HMGB4 با استفاده از سایت PONDOR:** درگام اول بمنظور پیش‌بینی ساختار کلی پروتئین HMGB4 و مشخص نمودن مناطقی از پروتئین که دارای ساختار سوم منظم و یا نامنظم است توالی پروتئین به سایت PONDOR داده شد و نتایج بدست آمده مورد بررسی قرار گرفت. تنها داده‌ای که PONDOR برای پیش‌بینی مناطق منظم و نامنظم ساختار پروتئین‌ها به آن نیاز دارد توالی پروتئین مورد نظر است (۳۵).

**تعیین ساختار سوم پروتئین HMGB4 با استفاده از دو روش مبتنی بر هومولوژی مدلینگ:** هومولوژی مدلینگ بر این واقعیت تکیه دارد که پروتئین‌هایی که از لحاظ تکاملی به هم مرتبط هستند ساختار مشابهی را نیز دارا می‌باشند. در این روش، ساختار سوم یک توالی پروتئینی خاص (هدف: Target) از طریق همترازسازی آن با توالی یک یا بیش از یک پروتئین با ساختار مشخص (الگو: Template) پیش‌بینی می‌شود. اولین مرحله در هومولوژی مدلینگ تشخیص کلیه ساختارهای پروتئینی مرتبط با توالی هدف می‌باشد. سپس از بین این ساختارها، الگوهای مناسب انتخاب می‌شوند. سپس توالی پروتئین هدف با توالی الگو(ها) همترازسازی شده و در نهایت مدل‌های سه بعدی از پروتئین مورد نظر ساخته می‌شود. در مرحله آخر مدل‌های ساخته شده مورد ارزیابی قرار می‌گیرند (۱۰).

پروتئین HMGB4 در بیضه‌ها بسیار بالا است، این در حالی است که این پروتئین در اغلب بافت‌ها بیان نمی‌شود و یا در معدودی از بافت‌ها نظیر مغز بمیزان بسیار کمی بیان می‌گردد (۵، ۶ و ۲۰).

DNA هدف اصلی داروهای ضدسرطانی حاوی پلاتین (نظیر سیس پلاتین) محسوب می‌شود (۹). این داروها با اتصال به نیتروژن شماره ۷ بازهای پورین موجب ایجاد اتصالات متقاطع در ساختار DNA می‌شوند (۱۵). پروتئین‌های HMGB با میل ترکیبی بالایی به این ساختارهای متشکل از DNA-پلاتین اتصال یافته و مانع دسترسی عوامل ترمیم DNA به این نقاط می‌شوند و بنابراین اثر ضدسرطانی این داروها را تقویت می‌نمایند (۱۴، ۳۷). پروتئین HMGB4 در مقایسه با پروتئین HMGB1 با میل ترکیبی بالاتری به این ساختارها اتصال می‌یابد و همچنین با قدرت بیشتری ترمیم این نقاط آسیب دیده را مهار می‌نماید. بنظر می‌رسد که حساسیت بسیار بالای تومورهای سلول‌های زایای بیضه نسبت به سیس پلاتین ناشی از بیان بسیار بالای این پروتئین در بافت بیضه باشد (۱۹).

بطور کلی پیش‌بینی ساختار سوم یک پروتئین با استفاده از توالی آمینوسیدی آن یک مسئله علمی اساسی محسوب می‌شود. از بین انواع روش‌های موجود برای تعیین ساختار پروتئین‌ها، هومولوژی مدلینگ بعنوان اولین و سریع‌ترین روش در نظر گرفته می‌شود. (۳۱، ۳۴). تا به امروز ساختار سوم HMGB4 در هیچکدام از گونه‌های پستانداران تعیین نشده است. در این مطالعه، ساختار سوم پروتئین HMGB4 انسانی با استفاده از نرم افزار مدلر و سرور I-TASSER تعیین گردیده است. در مجموع نتایج بدست آمده از این تحقیق می‌تواند به فهم بهتر ارتباط بین ساختار و عملکرد این پروتئین و همچنین طراحی داروهای جدید ضدسرطان کمک نماید.

## مواد و روشها

I-TASSER، بمنظور انجام فرایند مدلینگ، از متاسرور LOMETS برای جستجوی الگوهای ساختاری در کتابخانه PDB (Protein Data Bank) استفاده می‌نماید. این متاسرور متشکل از چندین برنامه threading است. این برنامه‌های threading عبارتند از: (۱) MUSTER (۲) FFAS-3D (۳) SPARKS-X (۴) HHSEARCH2 (۵) HHSEARCH I (۶) Neff-PPAS (۷) HHSEARCH (۸) cdPPAS (۱۰) wdPPAS (۹) pGenTHREADER (۳۳). در مجموع پنج مدل برای HMGB4 توسط I-TASSER ساخته شد که بر اساس C-score بهترین مدل انتخاب گردید. درجه قابل اطمینان بودن یک مدل پروتئینی ایجاد شده توسط I-TASSER به صورت کمی توسط C-score اندازه‌گیری می‌شود و بنابراین مدلی با بالاترین میزان C-score بهترین مدل شناخته می‌شود. مقادیر C-score بین ۵- تا ۲ متغیر است.

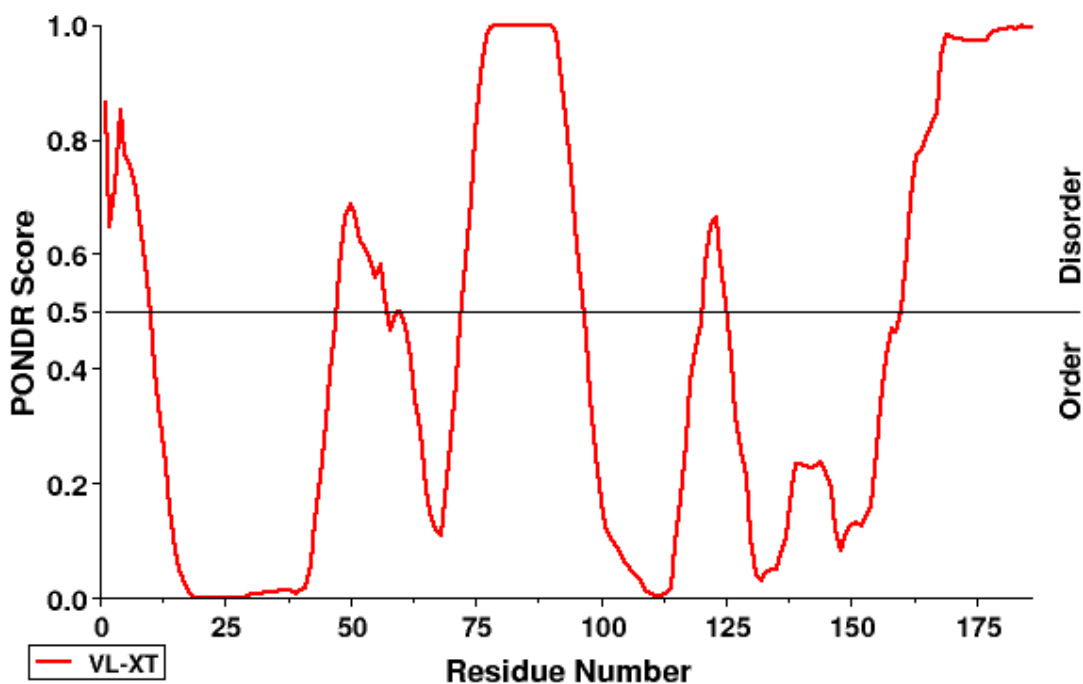
### نتایج

در مرحله اول بمنظور پیش‌بینی ساختار کلی پروتئین HMGB4 و تعیین مناطقی از پروتئین که دارای ساختار سوم منظم و یا نامنظم است توالی پروتئین HMGB4 به سرور PONDR داده شد و نتایج مورد بررسی قرار گرفت. نواحی فاقد ساختار مشخص (دارای ساختار نامنظم) می‌تواند کل یک پروتئین و یا بخش‌هایی از پروتئین را شامل شود. در واقع این نواحی که اساساً یا بطور نسبی و یا بطور کامل فولد نشده‌اند فاقد یک ساختار سوم مشخص می‌باشند. تحقیقات نشان می‌دهد که این نواحی در فعالیت‌های متعددی نظیر تشخیص DNA، تنظیم میل ترکیبی و یا ویژگی اتصال و کنترل طول عمر پروتئین‌ها نقش دارند. اگر چه این نواحی، در وضعیت‌های طبیعی فاقد ساختار سوم مشخص هستند اما ممکن است در نتیجه اتصال به مولکول‌های دیگر انتقال از حالت نامنظم به منظم را طی نمایند (۳۵). نتایج بدست آمده از این بخش در شکل ۱ نمایش داده شده است. همانطور که بخوبی در شکل

تعیین ساختار سوم HMGB4 با استفاده از نرم افزار مدلر: مدلر یک برنامه کامپیوتری است که ساختار سوم پروتئین‌ها را بر پایه همولوژی مدلینگ پیش‌بینی می‌نماید. برای تعیین ساختار سوم پروتئین HMGB4، نسخه 9.15 نرم افزار مدلر مورد استفاده قرار گرفت. تعیین ساختار سوم پروتئین‌ها با نرم‌افزار مدلر که از زبان برنامه‌نویسی پایتون استفاده می‌نماید شامل چندین مرحله می‌شود. همه این مراحل ساخت یک فایل پایتون (با پسوند .py) و اجرای آن با استفاده از نرم‌افزار مدلر را شامل می‌شود. در ذیل این مراحل بطور خلاصه ذکر شده است: ۱- جستجوی پروتئین‌هایی که توالی آمینواسیدی آن‌ها به HMGB4 شباهت دارد و ساختار سوم آن‌ها تعیین شده است. ۲- انتخاب یکی از این پروتئین‌ها بعنوان الگو برای مدل سازی. ۳- همتراز کردن توالی الگو و توالی HMGB4. ۴- مدل سازی و انتخاب مناسب‌ترین مدل از بین مدل‌های ساخته شده. ۵- ارزیابی مدل ساخته شده (۱۰، ۱۱ و ۲۳). بمنظور مشاهده و بررسی مدل‌های ایجاد شده از نرم افزارهای PyMOL و Swiss PDB Viewer استفاده گردید.

**تعیین ساختار سوم HMGB4 با استفاده از سرور I-TASSER**  
**TASSER** : سرور I-TASSER ( Iterative Threading ASSEmblly Refinement ) یک سرویس اینترنتی است که با استفاده از یک روش سلسله‌مراتبی ساختار و عملکرد یک پروتئین را پیش‌بینی می‌نماید. I-TASSER عنوان سرور شماره یک را برای پیش‌بینی ساختار پروتئین در آزمایش‌های CASP7، CASP8، CASP9، CASP10 و CASP11 ( CASP: Critical Assessment of Techniques for Protein Structure Prediction )، به خود اختصاص داده است. این سرور همچنین در آزمایش CASP9 بعنوان بهترین سرور برای پیش‌بینی عملکرد پروتئین شناخته شده است (۲۱، ۳۶ و ۳۸). بمنظور تعیین ساختار سوم پروتئین HMGB4 با استفاده از سرور I-TASSER توالی این پروتئین به سرور مورد نظر داده شد.

مشخص است محور X شماره آمینواسیدها و محور Y، PONDNR Score را مشخص می‌نماید. بخش‌هایی از پروتئین که دارای PONDNR Score بالای ۰/۵ است نشان دهنده مناطقی است که فاقد ساختار منظم است.



شکل ۱- پیش‌بینی بخش‌هایی از پروتئین HMGB4 که فاقد ساختار سوم مشخص است با استفاده از سرور PONDNR

2yrq بهترین الگویی است که می‌توان از آن در مدل‌سازی پروتئین HMGB4 استفاده نمود. زیرا درصد تشابه آن با HMGB4، ۴۴/۱۷٪ و E-value همترازسازی آن برابر صفر است. 2yrq در واقع باکس A و B پروتئین HMGB1 انسانی (بنیان‌های ۱۶۶-۱) است که از طریق روش NMR تعیین ساختار شده است (۲۸) اگرچه 1j3x و 2rtu درصد تشابه توالی بالاتری نسبت به 2yrq دارند و E-value همترازسازی آن‌ها نیز صفر است ولی تعداد بنیان‌های آمینواسیدی آن‌ها نسبت به 2yrq بسیار کمتر است. زیرا 2rtu تنها باکس A پروتئین HMGB1 انسانی (۳۲) و 1j3x دومین N-ترمینال HMGB2 گراز وحشی (۱۶) را شامل می‌شود. پس با توجه به اینکه 2yrq هر دو باکس A و B را در بر می‌گیرد الگوی کاملاً مناسبی برای مدل‌سازی HMGB4 محسوب می‌شود.

همانطور که در بخش مواد و روش‌ها اشاره شد، بمنظور تعیین ساختار سوم پروتئین HMGB4 انسانی، ابتدا توالی این پروتئین از پایگاه اطلاعاتی UniProt دریافت شد و سپس طی چند مرحله با استفاده از نرم‌افزار مدلر (نسخه 9.15) ساختار این پروتئین تعیین گردید. در اولین مرحله پروتئین‌هایی که توالی آمینواسیدی آن‌ها به HMGB4 شباهت دارد و فایل PDB آن‌ها موجود است جستجو شدند. نتایج در شکل ۲ نمایش داده شده است. ستون یک کد PDB و نوع زنجیره (A، B و ...)، ستون پنج تعداد بنیان‌های آمینواسیدی هر PDB، ستون هفت درصد تشابه توالی HMGB4 و توالی PDB مورد نظر که بوسیله طول همترازسازی یک‌دست شده است، ستون هشت مقادیر E-value همترازسازی را نمایش می‌دهد. تمام PDB‌هایی که E-value همترازسازی برابر با صفر دارند در شکل با بیضی‌های قرمز مشخص شده‌اند. با توجه به نتایج بدست آمده

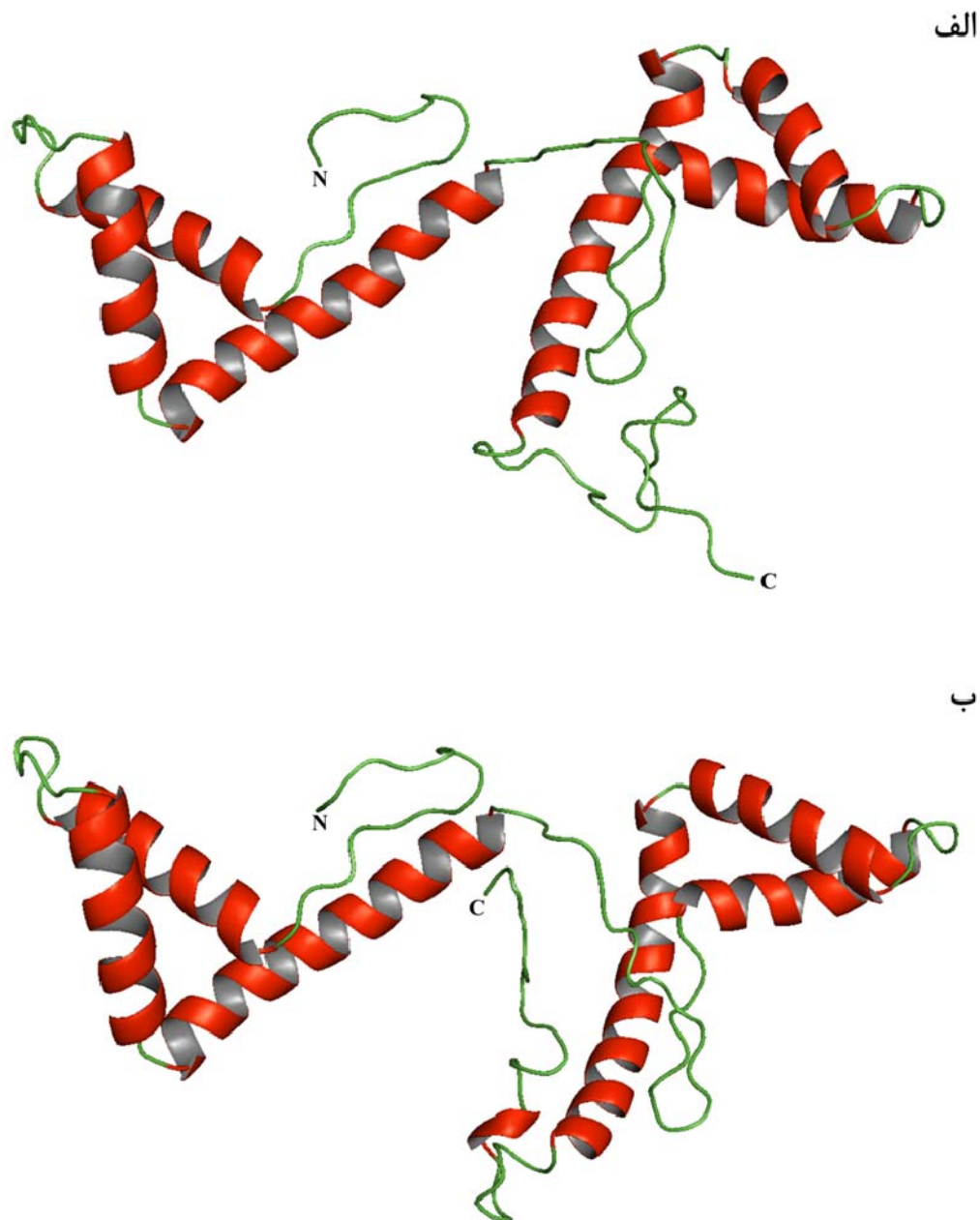
Dynamically allocated memory at		amaxprofile [B, KiB, MiB]:				948797	926.560	0.905					
> 4a3nA	1	298	5050	65	186	27.42	0.56E-04	2	62	95	156	1	62
> 1cg7A	1	4891	7500	93	186	36.25	0.21E-09	3	80	81	160	9	88
> 1ckrA	1	5112	9750	71	186	47.89	0.0	4	71	8	78	1	71
> 2co9A	1	5337	7800	102	186	34.18	0.46E-10	5	79	83	161	8	86
> 2crjA	1	5561	4750	92	186	30.88	0.37E-03	6	68	9	78	8	75
> 2cslA	1	5613	4700	92	186	34.92	0.49E-03	7	63	94	163	9	71
> 2eqzA	1	9816	9350	86	186	43.59	0.0	8	78	1	78	8	85
> 4euwA	1	10070	4250	72	186	23.53	0.41E-02	9	68	94	161	4	71
> 1gt0D	1	14145	4600	80	186	23.68	0.72E-03	10	76	94	169	4	79
> 2gzkA	1	14509	11550	159	186	29.94	0.0	11	155	5	161	1	157
> 1fmeA	1	15888	9150	77	186	41.89	0.0	12	74	88	161	2	75
> 1hrvA	1	16162	5000	73	186	27.94	0.81E-04	13	68	94	161	4	71
> 1hsmA	1	16242	8450	79	186	42.03	0.0	14	69	93	161	3	71
> 1i1iA	1	16796	4750	70	186	26.87	0.29E-03	15	67	94	160	4	70
> 1j3cA	1	18947	9750	79	186	43.42	0.0	16	76	86	161	2	77
> 1j3xA	1	18957	9950	77	186	48.05	0.0	17	77	1	77	1	77
> 1j46A	1	18992	5150	85	186	28.57	0.43E-04	18	77	94	172	6	82
> 2lhjA	1	23833	6200	97	186	30.59	0.20E-06	19	83	81	163	11	95
> 2mrcA	1	26533	6950	92	186	32.50	0.37E-08	20	78	86	163	10	89
> 1qrvA	1	34050	5000	73	186	35.00	0.81E-04	21	60	92	153	3	62
> 2rtuA	1	35881	10300	87	186	46.84	0.0	22	79	1	79	4	82
> 3tq6A	1	37886	6450	194	186	22.67	0.10E-06	23	149	9	163	7	178
> 3u2bC	1	38363	4650	76	186	25.00	0.53E-03	24	68	94	161	4	71
> 1wgfA	1	41888	5150	90	186	28.79	0.45E-04	25	66	87	152	15	80
> 1wx1A	1	42637	5750	73	186	33.33	0.16E-05	26	69	93	163	4	72
> 2yqiA	1	44471	8000	81	186	40.85	0.15E-10	27	71	93	163	10	80
> 2yrqA	1	44511	19150	173	186	44.17	0.0	28	160	1	161	8	170
> 2yulA	1	44632	5400	82	186	27.94	0.11E-04	29	68	94	161	9	76

شکل ۲- نتایج حاصل از جستجوی پروتئین‌هایی با ساختار سوم مشخص که درصد تشابه توالی بالایی با پروتئین HMGB4 دارند با استفاده از نرم افزار مدلر. مواردی که با بیضی‌های قرمز مشخص شده اند همگی دارای E-value همترازسازی برابر با صفر می‌باشند. اما بدایلی که در متن ذکر شده است 2yrq بعنوان الگو برای مدل‌سازی HMGB4 در نظر گرفته شد.

شکل ۳- الف بهترین مدل ایجاد شده توسط نرم افزار مدلر (مدل دو) را که از طریق نرم افزار PyMOL مشاهده شده است نمایش می‌دهد. اگرچه 2yrq الگوی مناسبی برای مدل‌سازی HMGB4 محسوب می‌شود اما بیش از ۲۰ اسیدآمینو انتهایی HMGB4 معادلی در 2yrq ندارند و در نتیجه مدل این بخش انتهایی را بدون استفاده از الگو می‌سازد. البته با توجه به نتایج بدست آمده از سرور PONDR (شکل ۱) بخش انتهایی پروتئین HMGB4 فاقد ساختار سوم مشخص است.

بمنظور تعیین ساختار سوم پروتئین HMGB4 با استفاده از سرور I-TASSER، توالی این پروتئین به فرمت FASTA از پایگاه اطلاعاتی UniProt دریافت شد و در اختیار سرور مورد نظر قرار گرفت. از بین پنج مدل ساخته شده توسط I-TASSER، مدل یک با بیشترین میزان C-score (۰/۷۶-) بعنوان بهترین مدل در نظر گرفته شد. شکل ۳-ب مدل یک را که با استفاده از نرم افزار PyMOL مشاهده شده است نمایش می‌دهد.

در مرحله بعد توالی 2yrq و HMGB4 با استفاده از روش align2d همترازسازی شدند. این روش همترازسازی که توسط نرم‌افزار مدلر استفاده می‌شود بر پایه یک الگوریتم برنامه نویسی پویا بنیان نهاده شده است و با روش‌های همترازسازی توالی-توالی متداول تفاوت دارد زیرا align2d بمنظور ایجاد همترازسازی، اطلاعات ساختاری از الگو را نیز مد نظر قرار می‌دهد (۱۰). پس از اجرای همترازسازی، مدل‌سازی انجام گرفت. در مجموع پنج مدل ساخته شد که بر اساس مقادیر DOPE و GA341 بهترین مدل انتخاب گردید. در واقع مدلی که کمترین میزان DOPE و بیشترین میزان GA341 را داشته باشد بهترین مدل محسوب می‌شود. مقادیر GA341 بین صفر تا یک متغیر است. با توجه به اینکه مقادیر GA341 پنج مدل بسیار به یکدیگر نزدیک بود و DOPE نسبت به GA341 فاکتور قابل اطمینان تری برای تعیین مدل مناسب محسوب می‌شود مدل دو با کمترین میزان DOPE (-13290.73340) بعنوان بهترین مدل در نظر گرفته شد.

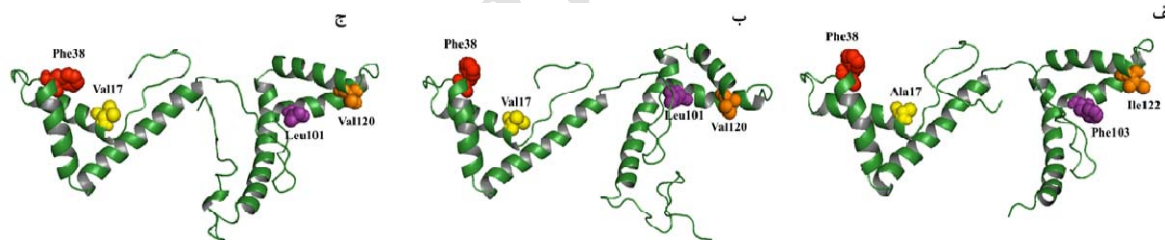


شکل ۳- بهترین مدل ساخته شده از پروتئین HMGB4 توسط نرم افزار مدلر (الف) و سرور I-TASSER (ب) که با استفاده از نرم افزار PyMOL مشاهده شده است.

داده‌ها نمایش داده نشده است). یکی از این ساختارها که توسط هشت برنامه threading سرور LOMETS بعنوان بهترین الگو معرفی شد زنجیره 2yrq A (2yrqA) است که در واقع ساختار محلول باکس A و B پروتئین HMGB1 انسانی (بنیان‌های ۱۶۶-۱) می‌باشد (۲۸) و توسط برنامه

همانطور که در بخش مواد و روش‌ها اشاره شد فرایند مدلینگ توسط I-TASSER با جستجوی الگوهای ساختاری در کتابخانه PDB توسط LOMETS آغاز می‌شود. سرور I-TASSER در مجموع از دو ساختار PDB بمنظور تعیین ساختار سوم پروتئین HMGB4 استفاده نمود

اتصال پروتئین‌های HMGB به DNA به گونه ای است که بنیان‌های آمینواسیدی حجیم هیدروفوب قرار گرفته در سطح مقعر HMG-box های A و B بین جفت بازهای شیار کوچک DNA قرار می‌گیرند. در پروتئین HMGB1، آلانین ۱۷ و فنیل آلانین ۳۸ از باکس A و فنیل آلانین ۱۰۳ و ایزولوسین ۱۲۲ از باکس B این مسئولیت را بر عهده دارند (۲۷). بررسی هم‌ترازسازی‌های انجام شده توسط نرم افزار مدلر و سرور I-TASSER، بنیان‌های معادل این بنیان‌ها را در HMGB4 مشخص می‌نماید (هم‌ترازسازی‌ها نمایش داده نشده است). والین ۱۷ و فنیل آلانین ۳۸ از باکس A و لوسین ۱۰۱ و والین ۱۲۰ از باکس B بنیان‌های حجیم هیدروفوبی هستند که در اتصال پروتئین HMGB4 به DNA نقش دارند. همانطور که مشخص است از این چهار اسیدآمینه فقط یکی از آنها (فنیل آلانین ۳۸) در هر دو پروتئین HMGB4 و HMGB1 حفظ شده است (۱۹). شکل ۴ موقعیت قرارگیری این بنیان‌های حجیم هیدروفوب را در پروتئین‌های HMGB1 (2yrq) و HMGB4 (مدل‌های ایجاد شده توسط مدلر و I-TASSER) نشان می‌دهد.



شکل ۴- مقایسه ساختار پروتئین HMGB1 (2yrqA) (الف) با مدل‌های ایجاد شده از ساختار پروتئین HMGB4 توسط نرم‌افزار مدلر (ب) و سرور I-TASSER (ج). بنیان‌های حجیم هیدروفوب از HMG-box A و HMG-box B که با وارد شدن به درون شیار کوچک DNA موجب اتصال این پروتئین‌ها به DNA می‌شوند به فرم کروی مشخص گردیده‌اند.

امروزه روش‌های مبتنی بر همولوژی مدلینگ بدلیل آنکه سریع‌تر و کم‌هزینه‌تر از روش‌های آزمایشگاهی هستند از اهمیت بسیار بالایی برای تعیین ساختار پروتئین‌ها برخوردار می‌باشند (۳۱). همچنین پیشرفت‌های علمی بدست آمده در این زمینه منجر به افزایش روزافزون قابلیت اطمینان و دقت این روش‌ها شده است. تعیین ساختار

مدلر نیز بعنوان الگو انتخاب گردید. همچنین زنجیره A 3tq6 (3tq6A) که ساختار کریستال فاکتور نسخه برداری A میتوکندریایی (TFAM) انسانی است (۲۲) توسط دو برنامه threading بعنوان الگو انتخاب شد. همانطور که قبلاً اشاره گردید هرچند 2yrq الگوی بسیار خوبی برای مدلسازی HMGB4 محسوب می‌شود اما بیش از ۲۰ اسیدآمینه انتهای HMGB4 معادلی در 2yrq ندارند و بنابراین برای ایجاد یک مدل که تمام طول HMGB4 را پوشش دهد نمیتوان فقط از 2yrq بعنوان الگو استفاده نمود. بنابراین سرور I-TASSER بمنظور ایجاد این نوع مدل، 2yrq و 3tq6 را بطور همزمان بعنوان الگوی مدلسازی مورد استفاده قرار داد.

همانطور که قبلاً در بخش مقدمه اشاره شد پروتئین‌های HMGB از طریق HMG-box های A و B به DNA متصل می‌شوند. هر دومین HMG-box که تقریباً دارای ۷۵ اسید آمینه است از سه آلفا هلیکس که بفرم L و با زاویه تقریبی ۸۰ درجه فولد شده اند تشکیل شده است. در شکل ۳ بخوبی مشخص است که ساختار پروتئین HMGB4 نیز از دو HMG-box A و B تشکیل شده است. در واقع نحوه

## بحث

تعیین ساختار سوم پروتئین‌ها با استفاده از روش‌های آزمایشگاهی همیشه بهترین گزینه بشمار نمی‌رود. زیرا علاوه بر آنکه این روش‌ها نیازمند صرف زمان و هزینه بالا هستند برای برخی از پروتئین‌ها قابل اجرا نمی‌باشند.

سوم پروتئین‌ها از بسیاری از جهات حائز اهمیت فراوان است (۳۴). بسیاری از ترکیباتی که امروزه بعنوان دارو برای درمان بیماری‌ها مورد استفاده قرار می‌گیرند در واقع عملکرد خود را از طریق میانکنش با یک پروتئین خاص انجام می‌دهند. بمنظور طراحی این نوع داروها ابتدا باید ساختار سوم پروتئین مورد نظر وجود داشته باشد. در ضمن با توجه به اینکه بین ساختار و عملکرد یک پروتئین رابطه تنگاتنگی وجود دارد تعیین ساختار سوم یک پروتئین به فهم بهتر عملکرد آن پروتئین کمک شایانی می‌نماید (۲، ۱۰ و ۳۱).

پروتئین HMGB4 که در سال ۲۰۰۸ شناسایی شد از اعضای خانواده پروتئین‌های HMGB است (۶). این خانواده که علاوه بر HMGB4 از سه عضو دیگر به نام های HMGB1، HMGB2 و HMGB3 تشکیل شده است فراوان ترین گروه پروتئین‌های غیرهیستونی HMG را تشکیل می‌دهد (۲۶). پروتئین HMGB4 بر خلاف سایر اعضای این خانواده که همگی بعنوان فعال کننده های نسخه برداری عمل می‌نمایند یک مهارکننده توانمند فرایند نسخه برداری محسوب می‌شود (۶). تا به امروز ساختار سوم پروتئین HMGB4 در هیچکدام از گونه های پستانداران تعیین نشده است. در مطالعه حاضر بمنظور تعیین ساختار سوم پروتئین HMGB4 انسانی از دو روش که هر دو بر پایه هومولوژی مدلینگ بنا شده اند استفاده گردید. در یکی از این روش‌ها نرم‌افزار مدلر برای تعیین ساختار سوم این پروتئین مورد استفاده قرار گرفت و در روش دیگر از سرور I-TASSER برای مدلسازی استفاده شد.

همانطور که بخوبی در مدل‌های ایجاد شده توسط نرم افزار مدلر (شکل ۳-الف) و سرور I-TASSER (شکل ۳-ب) مشخص است پروتئین HMGB4 همانند سایر اعضای خانواده پروتئین‌های HMGB دارای دو دومین HMG-box A و B برای اتصال به DNA می‌باشد. این پروتئین‌ها

بمنظور اتصال به DNA، بنیان‌های حجیم هیدروفوب قرار گرفته در سطح مقعر این دو دومین را بین جفت بازهای شیار کوچک DNA وارد می‌نمایند (۱ و ۲۶). در شکل ۴ موقعیت قرارگیری این بنیان‌های حجیم هیدروفوب در پروتئین‌های HMGB1 و HMGB4 مشخص شده است. از بین این چهار بنیان تنها فنیل‌آلانین ۳۸ در هر دو پروتئین حفظ شده است. همانطور که در بخش مقدمه اشاره شد پروتئین‌های HMGB با اتصال به DNA پلاتینه حاصل از میانکنش داروهای ضدسرطان پلاتینه نظیر سیس‌پلاتین با DNA مانع دسترسی سیستم ترمیم DNA به این نقاط و در نتیجه موجب تقویت اثر ضدسرطانی این نوع داروها می‌شوند. تحقیقات نشان می‌دهد که فنیل‌آلانین ۳۸ در HMGB4 موش نیز وجود دارد نقش بسیار مهمی در اتصال پروتئین‌های HMGB به DNA پلاتینه ایفا می‌نماید (۱۳، ۱۷، ۱۹ و ۲۴). بررسی نحوه قرارگیری این بنیان در مدل‌های ایجاد شده توسط مدلر و I-TASSER و مقایسه آن با فنیل‌آلانین ۳۸ پروتئین HMGB1 (2yrq) اطلاعات جالبی را ارائه می‌دهد (شکل ۵). در این شکل دو پروتئین HMGB4 و HMGB1 با استفاده از گزینه "iterative magic fit" نرم افزار Swiss PDB Viewer بر روی یکدیگر منطبق شده اند و موقعیت فنیل‌آلانین ۳۸ در این دو پروتئین مقایسه شده است. در مدل ایجاد شده توسط I-TASSER زنجیره جانبی این بنیان نسبت به بنیان معادلش در HMGB1 بنحو متفاوتی جهت گیری می‌نماید در حالی که در مدل مربوط به مدلر این جهت گیری متفاوت مشاهده نمی‌شود. نتایج حاصل از تحقیقات *In vitro* نشان می‌دهد که میل ترکیبی HMGB4 و HMGB1 برای اتصال به DNA پلاتینه و همچنین توانایی آنها برای ممانعت از ترمیم این نقاط توسط سیستم ترمیم DNA تفاوت معنی‌داری را نشان می‌دهد (۱۹). با توجه به اینکه فنیل‌آلانین ۳۸ نقش بسیار مهمی را در فرایند اتصال این دو پروتئین به DAN پلاتینه ایفا می‌نماید نظر می‌رسد که مدل ساخته شده توسط I-TASSER در مقایسه با مدلر این

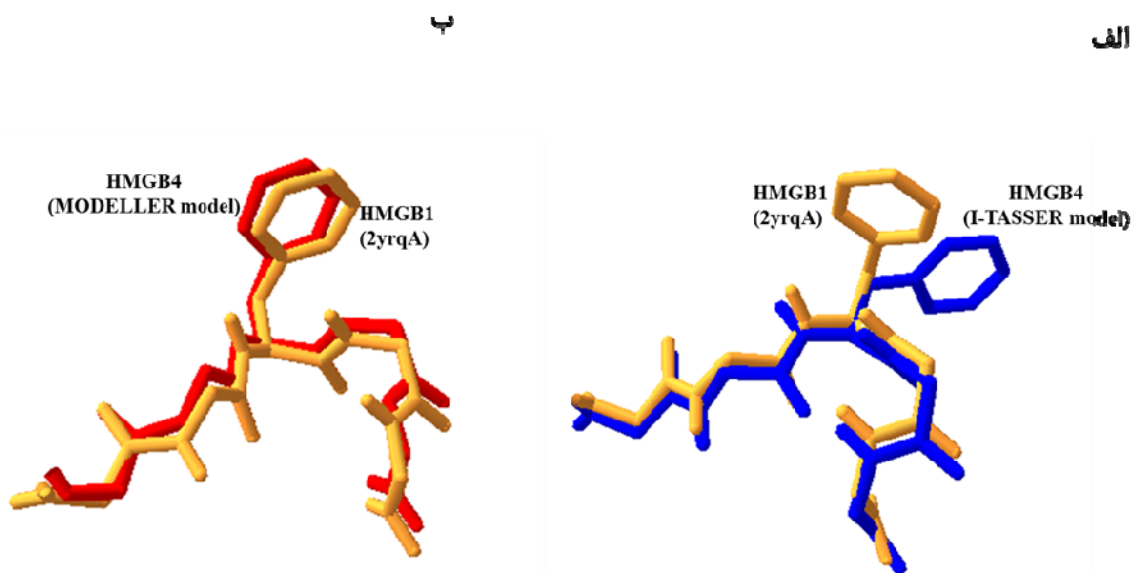
پروتئین HMGB4 که در سال ۲۰۰۸ شناسایی شد از اعضای خانواده پروتئین‌های HMGB است (۶). این خانواده که علاوه بر HMGB4 از سه عضو دیگر به نام های HMGB1، HMGB2 و HMGB3 تشکیل شده است فراوان ترین گروه پروتئین‌های غیرهیستونی HMG را تشکیل می‌دهد (۲۶). پروتئین HMGB4 بر خلاف سایر اعضای این خانواده که همگی بعنوان فعال کننده های نسخه برداری عمل می‌نمایند یک مهارکننده توانمند فرایند نسخه برداری محسوب می‌شود (۶). تا به امروز ساختار سوم پروتئین HMGB4 در هیچکدام از گونه های پستانداران تعیین نشده است. در مطالعه حاضر بمنظور تعیین ساختار سوم پروتئین HMGB4 انسانی از دو روش که هر دو بر پایه هومولوژی مدلینگ بنا شده اند استفاده گردید. در یکی از این روش‌ها نرم‌افزار مدلر برای تعیین ساختار سوم این پروتئین مورد استفاده قرار گرفت و در روش دیگر از سرور I-TASSER برای مدلسازی استفاده شد.

همانطور که بخوبی در مدل‌های ایجاد شده توسط نرم افزار مدلر (شکل ۳-الف) و سرور I-TASSER (شکل ۳-ب) مشخص است پروتئین HMGB4 همانند سایر اعضای خانواده پروتئین‌های HMGB دارای دو دومین HMG-box A و B برای اتصال به DNA می‌باشد. این پروتئین‌ها



در بین این روش‌ها بهترین عملکرد را دارد که نتایج بدست آمده در این مطالعه با نتایج مربوط به این تحقیق مطابقت دارد (۸).

اختلاف را بنحو بسیار خوبی توجیه می‌نماید. همچنین مطالعات انجام گرفته در زمینه مقایسه کیفیت مدل‌های پروتئینی ساخته شده توسط چندین روش متفاوت مبتنی بر همولوژی مدلینگ نشان می‌دهد که سرور I-TASSER



شکل ۵- مقایسه نحوه جهت‌گیری زنجیره جانبی بنیان فیل‌آلین ۳۸ در پروتئین HMGB1 (2yrqA) و مدل‌های ایجاد شده از پروتئین HMGB4 توسط سرور I-TASSER (الف) و نرم‌افزار مدل‌ر (ب).

سرور I-TASSER بمنظور ایجاد مدلی که تمام طول پروتئین HMGB4 را پوشش دهد علاوه بر 2yrq از 3tq6 نیز بعنوان الگو استفاده نمود. در واقع ساختار کریستال فاکتور نسخه برداری A میتوکندریایی (TFAM) انسانی است (۲۲). TFAM حاوی دو HMG-box است و نقش بسیار مهمی در بسته بندی، بقا و همچنین نسخه برداری DNA میتوکندریایی دارد (۱۸). بررسی بهترین مدل ایجاد شده توسط I-TASSER (شکل ۳-ب) نیز نشان می‌دهد که این بخش از پروتئین HMGB4 در کل فاقد ساختار منظم سوم می‌باشد، هرچند بخش بسیار کوچکی از آن دارای ساختار مارپیچ آلفا است. در مجموع اطلاعات بدست آمده از این مطالعه می‌تواند به طراحی داروهای ضدسرطان جدید و فهم ارتباط بین عملکرد و ساختار پروتئین HMGB4 کمک فراوانی نماید.

نرم‌افزار مدل‌ر 2yrq را بعنوان الگو برای مدلسازی مورد استفاده قرار داد. همانطور که در بخش نتایج اشاره شد بیش از ۲۰ اسیدآمینو انتهایی پروتئین HMGB4 معادلی در 2yrq ندارند. این بدین معناست که برای ساخت مدلی سه بعدی از این بخش از پروتئین HMGB4 الگویی وجود ندارد. بررسی کلی ساختار سوم پروتئین HMGB4 با استفاده از نرم‌افزار PONDR (شکل ۱) نشان می‌دهد که این بخش از پروتئین HMGB4 به احتمال زیاد فاقد ساختار سوم مشخص و منظم می‌باشد. البته این احتمال وجود دارد که بهنگام اتصال این پروتئین به DNA و یا پروتئین‌های دیگر، این بخش ساختار سوم منظمی را ایجاد نماید. بمنظور ایجاد ساختار سوم پروتئین HMGB4 با استفاده از نرم‌افزار مدل‌ر بخش انتهایی HMGB4 در همتراسازی لحاظ گردید، در نتیجه نرم‌افزار مدل‌ر این بخش از پروتئین را بدون الگو ایجاد نمود.

## تقدیر و تشکر

می‌باشد که بدینوسیله نویسندگان لازم می‌دانند از دانشگاه مذکور تقدیر و تشکر نمایند.

این مقاله مستخرج از طرح شماره ۹۴/۲۴۲۱/ص/۷ دانشگاه تحصیلات تکمیلی صنعتی و فناوری پیشرفته کرمان

## منابع

۱. امان زاده، ا. و محبت کار، ح. ۱۳۹۴. بررسی نقش HMGB1 در بیماری Cerebral Ischemia و مقایسه داروهای طراحی شده برای مسیر سیگنالی آن. مجله پژوهش‌های سلولی و مولکولی (مجله زیست‌شناسی ایران). ۲۸(۱)، ۲۲-۹.
۲. کیان مهر، ا.س. و مهدی زاده دهوسطی، ر. ۱۳۹۳. مطالعه فیلوژنتیکی باکتری سودوموناس پوتیدا تولید کننده پرولین دهیدروژناز و آنالیز بیوانفورماتیکی آنزیم جداسازی شده. مجله پژوهش‌های سلولی و مولکولی (مجله زیست‌شناسی ایران). ۲۷(۲)، ۲۸۵-۲۹۵.
3. Agresti A, Bianchi ME. 2003. HMGB proteins and gene expression. *Curr Opin Genet Dev*;13(2):170-8.
4. Bustin M. 1999. Regulation of DNA-dependent activities by the functional motifs of the high-mobility-group chromosomal proteins. *Mol Cell Biol*;19(8):5237-46.
5. Card CJ, Anderson EJ, Zamberlan S, Krieger KE, Kaproth M, Sartini BL. 2013. Cryopreserved bovine spermatozoal transcript profile as revealed by high-throughput ribonucleic acid sequencing. *Biol Reprod*;88(2):49.
6. Catena R, Escoffier E, Caron C, Khochbin S, Martianov I, Davidson I. 2009. HMGB4, a novel member of the HMGB family, is preferentially expressed in the mouse testis and localizes to the basal pole of elongating spermatids. *Biol Reprod*;80(2):358-66.
7. Das D, Scovell WM. 2001. The binding interaction of HMG-1 with the TATA-binding protein/TATA complex. *J Biol Chem*;276(35):32597-605.
8. Dolan MA, Noah JW, Hurt D. 2012. Comparison of common homology modeling algorithms: application of user-defined alignments. *Methods Mol Biol*;857:399-414.
9. Eastman A. 1987. The formation, isolation and characterization of DNA adducts produced by anticancer platinum complexes. *Pharmacol Ther*;34(2):155-66.
10. Eswar N, Webb B, Marti-Renom MA, et al. 2006. Comparative protein structure modeling using Modeller. *Curr Protoc Bioinformatics*;Chapter 5:Unit 5 6.
11. Fiser A, Do RK, Sali A. 2000. Modeling of loops in protein structures. *Protein science : a publication of the Protein Society*;9(9):1753-73.
12. Goodwin GH, Sanders C, Johns EW. 1973. A new group of chromatin-associated proteins with a high content of acidic and basic amino acids. *Eur J Biochem*;38(1):14-9.
13. He Q, Ohndorf UM, Lippard SJ. 2000. Intercalating residues determine the mode of HMG1 domains A and B binding to cisplatin-modified DNA. *Biochemistry*;39(47):14426-35.
14. Huang JC, Zamble DB, Reardon JT, Lippard SJ, Sancar A. 1994. HMG-domain proteins specifically inhibit the repair of the major DNA adduct of the anticancer drug cisplatin by human excision nuclease. *Proc Natl Acad Sci U S A*;91(22):10394-8.
15. Jamieson ER, Lippard SJ. 1999. Structure, Recognition, and Processing of Cisplatin-DNA Adducts. *Chem Rev*;99(9):2467-98.
16. Kurita J, Shimahara, H., Yoshida, M., Tate, S. 2004. Solution structure of the N-terminal domain of the HMGB2. <http://www.rcsb.org/pdb/explore/explore.do?structureId=1j3x>.
17. Ohndorf UM, Rould MA, He Q, Pabo CO, Lippard SJ. 1999. Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins. *Nature*;399(6737):708-12.
18. Parisi MA, Clayton DA. 1991. Similarity of human mitochondrial transcription factor 1 to high mobility group proteins. *Science*;252(5008):965-9.
19. Park S, Lippard SJ. 2012. Binding interaction of HMGB4 with cisplatin-modified DNA. *Biochemistry*;51(34):6728-37.

20. Petit FG, Kervarrec C, Jamin SP, et al. 2015. Combining RNA and protein profiling data with network interactions identifies genes associated with spermatogenesis in mouse and human. *Biol Reprod*;92(3):71.
21. Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*;5(4):725-38.
22. Rubio-Cosials A, Sidow JF, Jimenez-Menendez N, et al. 2011. Human mitochondrial transcription factor A induces a U-turn structure in the light strand promoter. *Nat Struct Mol Biol*;18(11):1281-9.
23. Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol*;234(3):779-815.
24. Sanchez-Giraldo R, Acosta-Reyes FJ, Malarkey CS, Saperas N, Churchill ME, Campos JL. 2015. Two high-mobility group box domains act together to underwind and kink DNA. *Acta Crystallogr D Biol Crystallogr*;71(Pt 7):1423-32.
25. Stros M, Launholt D, Grasser KD. 2007. The HMG-box: a versatile protein domain occurring in a wide variety of DNA-binding proteins. *Cell Mol Life Sci*;64(19-20):2590-606.
26. Stros M. 2010. HMGB proteins: interactions with DNA and chromatin. *Biochim Biophys Acta*;1799(1-2):101-13.
27. Thomas JO, Travers AA. 2001. HMG1 and 2, and related 'architectural' DNA-binding proteins. *Trends Biochem Sci*;26(3):167-74.
28. Tomizawa T, Koshihara S., Watanabe S., Harada T., Kigawa T., Yokoyama S. 2008. Solution structure of the tandem HMG box domain from Human High mobility group protein B1. <http://www.rcsb.org/pdb/explore.do?structureId=2yrq>.
29. Travers AA. 2003. Priming the nucleosome: a role for HMGB proteins? *EMBO Rep*;4(2):131-6.
30. UniProt C. 2011. Ongoing and future developments at the Universal Protein Resource. *Nucleic acids research*;39(Database issue):D214-9.
31. Vyas VK, Ukawala RD, Ghate M, Chintha C. 2012. Homology modeling a fast tool for drug discovery: current perspectives. *Indian J Pharm Sci*;74(1):1-17.
32. Wang J, Tochio N, Takeuchi A, Uewaki JI, Kobayashi N, Tate SI. 2013. Redox-sensitive structural change in the A-domain of HMGB1 and its implication for the binding to cisplatin modified DNA. *Biochem Biophys Res Commun*;441:701-6.
33. Wu S, Zhang Y. 2007. LOMETS: a local meta-threading-server for protein structure prediction. *Nucleic acids research*;35(10):3375-82.
34. Xiang Z. 2006. Advances in homology protein structure modeling. *Curr Protein Pept Sci*;7(3):217-27.
35. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN. 2010. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta*;1804(4):996-1010.
36. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. *Nat Methods*;12(1):7-8.
37. Zamble DB, Mikata Y, Eng CH, Sandman KE, Lippard SJ. 2002. Testis-specific HMG-domain protein alters the responses of cells to cisplatin. *J Inorg Biochem*;91(3):451-62.
38. Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*;9:40.
39. Zhang Q, Wang Y. 2010. HMG modifications and nuclear function. *Biochim Biophys Acta*;1799(1-2):28-3.

## Determination of tertiary structure of human HMGB4 protein by two homology modeling-based methods

Lotfi S., Mortazavi M. and Riahi-Madvar A.

Biotechnology Dept., Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman, I.R. of Iran

### Abstract

HMGB4 protein which was identified in 2008 is a member of mammalian non-histone HMGB protein family. All HMGB proteins bind to DNA through two tandem DNA-binding motifs named HMG-box A and HMG-box B. HMGB proteins are known to improve the anticancer properties of platinum-based drugs such as cisplatin by high affinity binding to the platinum-DNA adducts and retarding the process of DNA excision repair. The results from previous studies have revealed in comparison with full-length HMGB1, the full-length HMGB4 binds with higher affinity to platinum-DNA lesions and inhibits the excision repair of the lesions much more efficient. It seems the hypersensitivity of testicular germ cell tumors to cisplatin is originated from strong expression of HMGB4 in this tissue. In the present work, the tertiary structure of human HMGB4 was determined using two methods based on homology modeling (MODELLER software and I-TASSER server). The results show that models created by both methods have good quality and stability. Although I-TASSER model in comparison to the modeller better explain the differences between HMGB1 and HMGB4 binding to the platinated DNA. The results obtained from the work demonstrate that HMGB4 binds to DNA by intercalating of Valine17 and Phenylalanine38 from box A and Leucine101 and Valine120 from box B between base pairs of DNA minor groove. The results also show the C-terminal part of HMGB4 possesses disordered tertiary structure. In general, the results will be helpful to understand the structure-function relationship of the protein and to design new anticancer drugs.

**Key words:** HMGB4 protein, Cisplatin, Homology modeling, MODELLER software, I-TASSER server