Contents lists available at UGC-CARE

International Journal of Pharmaceutical Sciences and Drug Research

[ISSN: 0975-248X; CODEN (USA): IJPSPP]

Available online at www.ijpsdronline.com

Research Article

Microarray Data Analysis, Structure Prediction, and *In silico* Docking of Drugs for Inhibiting Overexpression of High Mobility Group A1 in Human Malignant Neoplasias

Manish Devgun*, Sukhbir Lal

Institute of Pharmaceutical Sciences, Kurukshetra University, Kurukshetra-136119, Haryana, India

ARTICLE INFO

Article history:

Received: 27 March, 2020

Revised: 28 August, 2020

Accepted: 02 September, 2020

Published: 30 September, 2020

Keywords:

High Mobility Group A proteins, Homology modeling, Microarray data analysis, Molecular docking.

DOI:

10.25004/IJPSDR.2020.120504

ABSTRACT

The High Mobility Group A1 (*HMGA1*) gene overexpression has been widely observed in various types of cancers. The raw data for microarray data analysis was obtained from the dataset record GDS3525. The SOM and K-means of the genesis led to the identification of two clusters (each consisting of 30 genes) bearing the *HMGA1* gene. This on further analysis resulted in the identification of 14 similar genes by Easy M-A. The evolutionary similarity of *HMGA1* and *GORASP2* is clearly observed in the phylogenetic tree. Due to the absence of precise structures, the homology modeling was done by using EasyModeller, and the resulting models of proteins *HMGA1* and *GORASP2* were validated by the Ramachandran plot. These models were further put to loop optimization by Modloop, and the output models were assessed by Ramachandran plot (Rampage) and through SAVS (Procheck). The molecular docking was done by using Autodock. This resulted in two ligands, DB11641 (vinflunine) and DB12674 (lurbinectedin), showing potential for the effective treatment of various types of cancers characterized by the overexpression of *HMGA1* and *GORASP2*.

Introduction

Microarray technology has equipped scientists with the capability to explore the expression levels of numerous genes in one single experiment. Microarray data analysis has been employed to identify any significant biomarkers for diseases. Differentially expressed genes have been recognized by exploiting the technique of significance analysis of microarray (SAM).[1-3] The protein microarray database (PMD) has been specially constructed to archive and evaluate protein microarray data.[4]

The High Mobility Group A (HMGA) is a family of non-histone chromatin binding small nuclear proteins, which, as the name suggests, possesses high electrophoretic mobility in polyacrylamide gels. The HMGA family is comprised of *HMGA1* and *HMGA2*. Both of these are encoded by different genes present on chromosomes 6p21 and 12q14-15, respectively. There are three isoforms of *HMGA1*, i.e., *HMGA1a*, *HMGA1b*, and *HMGA1c*. The HMGA proteins have acidic carboxyl-terminal, and they all are linked with chromatin. These proteins have N-terminal DNA binding domains, known as "AT-hook" that interpose in the binding to AT-rich regions of chromatin. The normal human adult cells contain a very low amount of HMGA protein or maybe even absent.[5-13]

One study suggested that *HMGA1* and matrix metalloproteinase-11 have a crucial function in the generation and advancement of human skin cancer.[14] Another study analyzed the *HMGA1* expression in human epidermal squamous carcinoma SCC-13 cells and HeLa

*Corresponding Author: Manish Devgun

Address: Institute of Pharmaceutical Sciences, Kurukshetra University, Kurukshetra-136119, Haryana, India

Email ✉: manishdev19@kuk.ac.in

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

cells. They suggested that *HMGA1* has a significant role in managing autophagy and insinuated a unique way of contributing to *HMGA1* towards cancer progression.[5] Yet another study emphasized the role of both *HMGA1* and 2 in the formation of gastric cancer.[8]

A study evaluated the expression of *HMGA1* proteins in a group of ovarian carcinoma cell lines and tissues and recommended that ovarian cancer cell growth can be repressed by arresting *HMGA1* proteins.[15] Another study concluded that the cancer cells' sensitivity to antineoplastic agents could be increased by blocking *HMGA1* proteins.[6] Yet another study emphasized the importance of overexpression of *HMGA1* in human uterine tumors and suggested *HMGA1* to be a rational therapeutic target.[7]

The three-dimensional protein structure helps to understand not only its functions but also its dynamics, interactions with ligands, and other proteins. The *in vitro* methods of determination of protein structure, i.e., nuclear magnetic resonance (NMR) and X-ray crystallography, are expensive, time-consuming, and complex processes. Moreover, the large size of some proteins also plays a role in the failure of NMR to determine the structure. In the absence of an experimentally determined structure, the *in silico* method of comparative or homology modeling has proved to be an asset. This method is established on the basis of the relationship between the structure of the protein and its amino acid sequence. Human evolution changes the structure slowly *vis-à-vis* the associated sequence. The homology modeling method has been utilized to derive the structure of proteins. The 3D model structures of the proteins have been utilized to identify their probable interactions with other proteins and small ligand molecules, thereby identifying the potential inhibitors by exploiting the technique of *in silico* docking.[16-24] In this study, the vast amount of raw gene expression data for *HMGA1* was analyzed, and differentially co-expressed genes were identified. The phylogenetic analysis was performed to establish closely related genes. The corresponding proteins were modeled by utilizing homology modeling, and potential molecules were identified using docking.

The current study recognizes the crucial role of *HMGA1* in various cancers. Certainly, it creates a groundwork for discovering and developing novel therapeutics in human malignant neoplasias by reporting certain potential drug molecules for the first time, which may act as inhibitors in *HMGA1* and *GORASP2* overexpressed cancers.

Methodology

The hardware comprised of a personal computer (Lenovo IdeaPad 330S) having Intel (R) Core (TM) i3-8130U central processing unit, Windows 10 Home Premium 64-bit operating system having random access memory of 4 GB.

Microarray Data Analysis

Data Retrieval

The Gene Expression Omnibus (GEO) is a public functional genomics data repository supporting Minimum Information About a Microarray Experiment (MIAME)-compliant data submissions.[25,26] The raw data for microarray data analysis was obtained from the dataset record GDS3525. The title is "Ovarian cancer and depression," and the reference series is GSE9116.[27,28] A total of 22,284 gene entries were taken into consideration. The average values from the 10 sample count for the respective genes were calculated.

Easy M-A

The Easy M-A program, developed by Mr. Deepankar Chakraborty, is distributed as a part of Bio-En-Gene-Ier. It has pre-clustering tools, like DelBGene, responsible for removing genes without any defined name. Out of the total values, only 22,216 genes with names were taken into consideration. Another tool, CpyAvg, fills the vacant cells with the average value.

The *HMGA1* average value was identified, and out of the two values (6.038 and 9.041), the greater value of 9.041 was selected for further procedure. Yet another tool, FilterArray, selects only those rows whose average value falls within the specified range. The range of the value specified was determined to be 8.9 to 9.1 (9.041 ± 0.6%), and within this range, 159 rows (159 genes) were selected. The post-clustering tool, SimGene, identifies the common genes from two cluster's text files with significantly similar images.

Genesis

Genesis assimilates numerous microarray data analysis tools, such as, filters, normalization and visualization tools, distance measures, and common clustering algorithms, including hierarchical clustering, self-organizing maps, k-means, principal component analysis, and support vector machines.[29]

The self-organizing maps (SOM) have been used to clarify and streamline the gene-expression data.[30] SOM was availed by adopting various parameters, like dimensions X and Y: 3; iterations: 2,000; alpha: 0.05; radius: 3; initialization: random genes; neighborhood: Gaussian; and topology to be hexagonal. This resulted in the formation of nine clusters of genes.

The K-means clustering has led to the identification of the genes with comparable expression patterns.[30] The K-means clustering was applied and defined by a few parameters: the number of clusters: 9; maximum iterations: 2,000; and runs to be 1. These eventuated nine clusters of genes with similar expressions.

The clusters bearing the *HMGA1* gene were identified, and the same was subjected to the post-clustering tool, SimGene of the Easy M-A.

Phylogenetic Tree

Molecular Evolutionary Genetics Analysis (MEGA version 5) is software for exploring online databases, constructing sequence alignment and phylogenetic trees, and keeping an evolutionary perspective in focus.[31] The multiple sequence alignment was created by using ClustalW. The evolutionary analogy between sequences was derived by calculating the proportion of amino acids/ nucleotide differences between sequences.

Homology Modeling

Both *HMGA1* and Golgi Reassembly Stacking Protein 2 (*GORASP2*, 55kDa) proteins arise from the same node, and thus, homology modeling was done to predict the structure of these two evolutionary proximate proteins. *GORASP2* gene encodes a member of the Golgi Reassembly Stacking Protein family.[32]

Sequence Alignment

Fast alignment (FASTA): The nucleotide sequences or peptide sequences are expressed by the FASTA format, using single-letter codes.[33] The FASTA sequence of *HMGA1* and *GORASP2* were obtained from the website of the National Centre for Biotechnology Information (NCBI).[34,35]

Basic local alignment search tool (*BLAST):* The amino acid sequence of different proteins was investigated using the BLAST algorithm.[36] The BLAST-P was undertaken by using the protein data bank (PDB) proteins database.

Protein homology/ analogy recognition engine (Phyre): To anticipate and interpret protein structure, function, and mutations, Phyre2 was utilized.[37] The FASTA sequences were subjected to Phyre for further analysis.[38]

Templates preparation: The data attained from BLAST and Phyre was subjected to analysis at the Research Collaboratory for Structural Bioinformatics (RCSB) protein data bank, which has facts and figures about the 3D shapes of proteins, nucleic acids, and complex assemblies.[39] The resolution (Å), R value, and method of X-ray diffraction formed the template selection basis.

Molecular Modeling

EasyModeller was employed for performing the homology modeling of *HMGA1* and *GORASP2*. It is a front end graphical interface to MODELLER. EasyModeller has been applied to proteins to attain three-dimensional structure models. This software utilizes and analyzes the sequence and template data.[40] For the purpose of visualization, the Swiss-Pdb viewer was introduced.[41]

Structure prediction: The selected templates were proposed to the EasyModeller. The five models were designed, which were scored on the basis of discreet optimized protein energy (DOPE), GA341, and MODELLER objective function (molpdf). The DOPE score is a statistical tool used by various homology modeling programs and is probably the most reliable at separating native-like models from decoys. In the molpdf, the program minimizes the objective function *F* in relation to Cartesian coordinates of ~10,000 atoms (3D points) that form a system, i.e., one or more molecules. The fundamental principle of GA341 lies in the similarity of the sequence of the template and the model. The successful protein model is required to have a minimum DOPE score and molpdf values. The GA341 values should lie between 0 and 1 for the models to be more fruitful.[42]

Predicted models validation: The predicted models of *HMGA1* and *GORASP2* were validated by the Ramachandran plot accomplished by Rampage and Procheck in structural analysis and verification server (SAVS).[43,44]

Loop Modeling

The shape and the physicochemical properties of the proteins are very important for its function; thus, a precise model is essential to understand the protein/ligand interaction studies. For more clarity on the structure of proteins, Modloop, a web server, was utilized for loop optimization. The output models were submitted to Rampage for Ramachandran plot assessment and Procheck in SAVS. The procedure of loop modeling and consecutive validation was executed till an optimized model is obtained.[45]

Ligand Generation

The potential agents active against *HMGA1* and *GORASP2* proteins were searched using various databases, like PubMed, DrugBank, and ZINC database. Trabectedin was used as a lead compound in the similar chemical structure search with the similarity threshold of 0.6 in DrugBank.[46-49]

Molecular Docking

Molecular docking is an indispensable tool for computer-assisted drug designing. The feasible and reasonable binding mode(s) of the target protein and its ligands are predicted. These modes are then scored by utilizing various scoring functions.[50] The molecular docking was done against *HMGA1* protein and *GORASP2* using AutoDock, which is a suite of automated docking tools.[51]

Results and Discussion

Prediction of Co-Expressed Genes

The SOM and K-means were manually evaluated to identify the clusters bearing *HMGA1* gene. This resulted in cluster number 9 from the SOM and cluster number 5 from K-means. These clusters consisted of 30 genes (Figs 1 and 2).

Identification of Common Genes

These two clusters were subjected to Easy M-A to identify the common genes, which resulted in 14 common genes with similar expressions (Table 1).

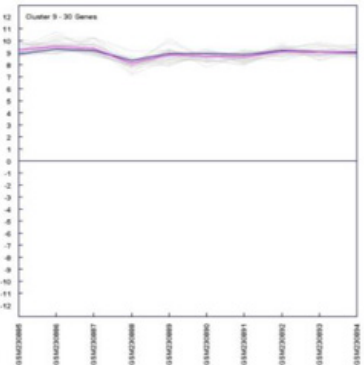

Fig. 1: SOM result cluster no. 9

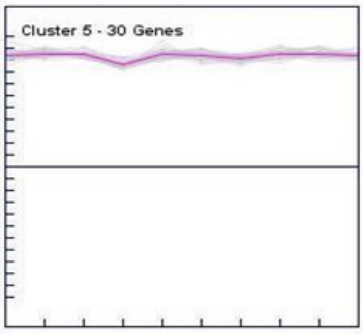

Fig. 2: K-means result cluster no. 5

Fig. 3: Phylogenetic tree: MEGA-estimation of evolutionary distance by using p-distance

Table 1: Common genes found using SimGene from Bio-En-Gene-Ier

S. No.	*Genes*	*S. No.*	*Genes*
1	*HNRNPA3*	8	*IDH1*
2	*VAMP3*	9	*ZC3H15*
3	*SNRPD3*	10	*GORASP2*
4	*HMGA1*	11	PCNP
5	*ST13*	12	*PRPF40A*
6	*SET*	13	*VPS26A*
7	*PSMC2*	14	SPG21

Phylogenetic Tree

The neighbor-joining method in MEGA5 was used to determine the evolutionary history. The displayed optimal tree has a sum of branch length equal to 6.15181327. The evolutionary distances (number of amino acid differences per site) were calculated by using the p-distance method. In this amino acid sequence analysis, all gaps and missing data positions were removed. The concluding data set consisted of 81 such positions. The evolutionary similarity of *HMGA1* and *GORASP2* can be clearly observed in Fig. 3.

Template Generation

The FASTA sequences for both *HMGA1* and *GORASP2* proteins were extracted from NCBI. The GenBank no. of *HMGA1* protein is AAH71863.1. It is a 107 amino acid protein. The GenBank no. of *GORASP2*, 55kDa protein is AAH07770.1. It consists of 452 amino acids. The NCBI was utilized to execute BLAST, using the program BLASTP2.9.0+, 02, and 14 BLAST hits were recorded for *HMGA1* and *GORASP2*, as shown in Figs 4a and b, respectively. The Phyre was also exploited for protein structure prediction. The consolidated data of both BLAST and Phyre was subjected to RCSB protein data bank analysis. The attained results were organized in a sequence of decreasing % ID and increasing resolution (Table 2). The five templates (4j2l, 3gn6, 4zqy, 4do8, and 2h8u) in case of *HMGA1* and five templates (4kfw, 4edj, 5h3j, 5gmi, and

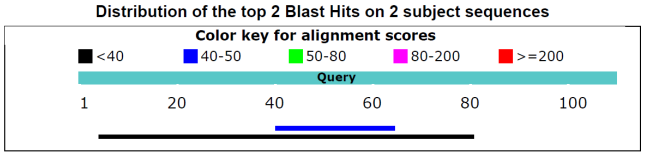

Fig. 4a: BLAST hits in case of *HMGA1* protein

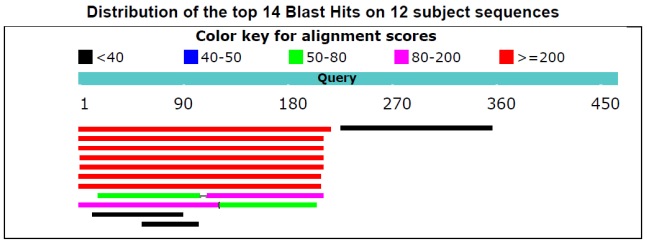

Fig. 4b: BLAST hits in case of *GORASP2*, 55kDa protein

Table 2: Templates using Blast, Phyre, and RCSB protein data bank

S. No.	*Template/ Accession No.*	*ID %*	*Resolution (Å)*	*R-value (free/work)*	*S. No.*	*Template/ Accession No.*	*ID %*	*Resolution (Å)*	*R-value (free/ work)*
HMGA1					*GORASP2*				
1	4j2l	100	3.15	0.294/0.236	1	4kfw	99.53	2.7	0.29/0.202
2	3gn6	8.9	1.8	0.192/0.163	2	4edj	99.52	1.901	0.271/0.209
3	4zqy	44	7.2	2.951/0.268	3	5h3j	99.03	1.33	0.162/0.137
4	4do8	50	5.5	1.802/0.234	4	5gmi	99.52	2.71	0.291/0.256
5	2h8u	56	5.5	2.1/0.266	5	3rle	99.03	1.649	0.222/0.176
					6	4rey	67.96	1.96	0.190/0.152

3rle) in case of *GORASP2* were selected on the basis of their chains, ID%, resolution (≤ 3 Å), and the R value (≤ 0.5).

Molecular Modeling

The molecular modeling was initiated by presenting the selected five templates to the EasyModeller. The analysis of all of the five models was done on the basis of GA341, molpdf, and DOPE scoring functions. This led to the suggestion of model no. 5 for *HMGA1* and model no. 3 in case of *GORASP2* based on lowest DOPE score and minimum value of molpdf, along with GA341 score lying in between 0 to 1.

Validation of Predicted Model

HMGA1 and *GORASP2* models were evaluated by putting forward the PDB files to Rampage and SAVS (Procheck). The Ramachandran plot validated the result. The most stable model (minimum energy) is revealed by its lowest DOPE score and molpdf or with the highest GA341 assessment score. Model number 5 for *HMGA1* and model number 3 for *GORASP2* were selected on these bases for further analysis (Table 3).

Loop Modeling

The selected model of *HMGA1* (no. 5) and that of *GORASP2* (no. 3) were further subjected to loop modeling (ModLoop), and the output models were assessed by Ramachandran plot (Rampage) and through SAVS (Procheck). The protein model loop *HMGA1* (Figs 5a and b) having maximum percentage (92.2%) of residues in

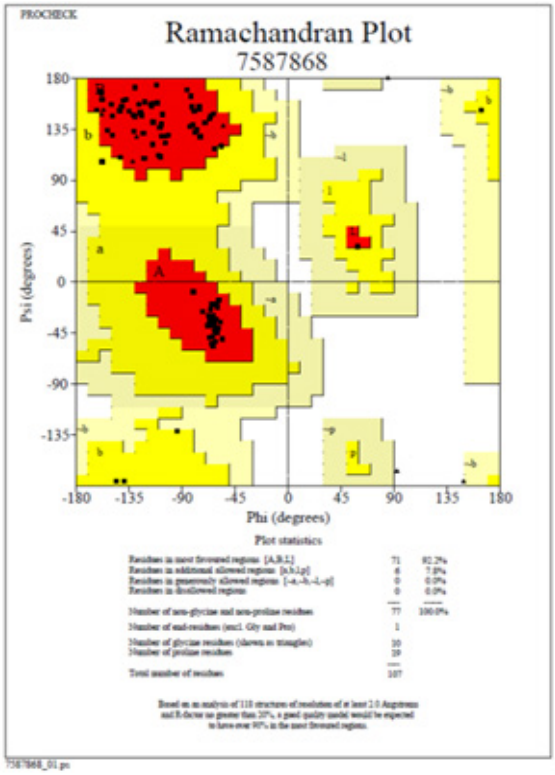

Fig. 5a: Ramachandran plot analysis of *HMGA1* protein model

Table 3: DOPE score and Ramachandran plot analysis of five possible models of *HMGA1* and *GORASP2* proteins

Model No.	Proteins	Molpdf	DOPE	GA341	Validation by Rampage			Validation by SAVS			
					Residues in favored region (Number/%)	Residues in allowed regions (Number/%)	Residues in outlier regions (Number/%)	Residues in most favored regions (Number/%)	Residues in additional allowed regions (Number/%)	Residues in generously allowed regions (Number/%)	Residues in disallowed regions (Number/%)
B99990001	*HMGA1*	3220.81177	-4323.33643	0.00256	90/85.7	11/10.5	4/3.8	59/76.6	13/16.9	3/3.9	2/2.6
	GORASP2	12468.17676	-29597.37109	1	397/88.2	38/8.4	15/3.3	274/77.4	62/17.5	10/2.8	8/2.3
B99990002	*HMGA1*	3272.39478	-4450.28271	0.00191	91/86.7	9/8.6	5/4.8	57/74.0	11/14.3	6/7.8	3/3.9
	GORASP2	10635.45703	-29600.09375	0.99985	404/89.8	36/8	10/2.2	290/81.9	57/16.1	4/1.1	3/0.8
B99990003	*HMGA1*	3393.95264	-4296.90186	0.00354	90/85.7	9/8.6	6/5.7	60/77.9	10/13	2/2.6	5/6.5
	GORASP2	10459.53223	-30248.62109	1	412/91.6	28/6.2	10/2.2	291/82.2	52/14.7	5/1.4	6/1.7
B99990004	*HMGA1*	3126.68262	-4660.67920	0.00679	97/92.4	5/4.8	3/2.9	67/87	6/7.8	3/3.9	1/1.3
	GORASP2	11776.05566	-30003.40039	1	410/91.1	26/5.8	14/3.1	284/80.2	54/15.3	11/3.1	5/1.4
B99990005	*HMGA1*	3087.25586	-4774.40918	0.00236	99/94.3	6/5.7	0/0	68/88.3	7/9.1	1/1.3	1/1.3
	GORASP2	12819.43652	-28423.72266	1	412/91.6	27/6.0	11/2.4	296/83.6	46/13	6/1.7	6/1.7

most favored region and 7.8% residues in additional allowed regions with no residues in generously allowed, as well as, disallowed regions was selected for *HMGA1* and in the case of *GORASP2*, the protein model loop *GORASP2* (Figs 6a and b) having maximum percentage (90.4%) of residues in most favored region and 9.3% residues in the additional allowed region along with 0.3% residues in generously allowed region and no residues in disallowed regions was selected for *GORASP2* (Table 4).

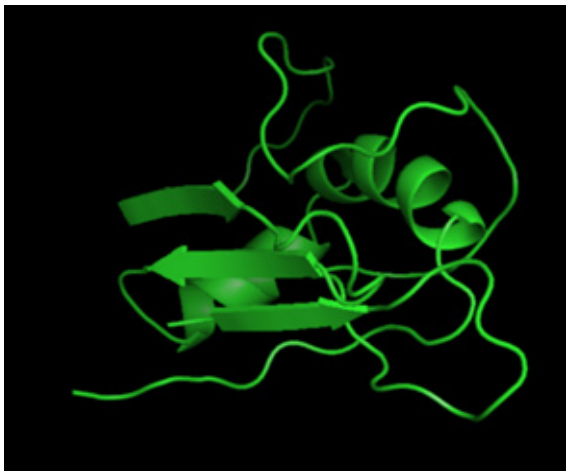

Fig. 5b: *HMGA1* protein model

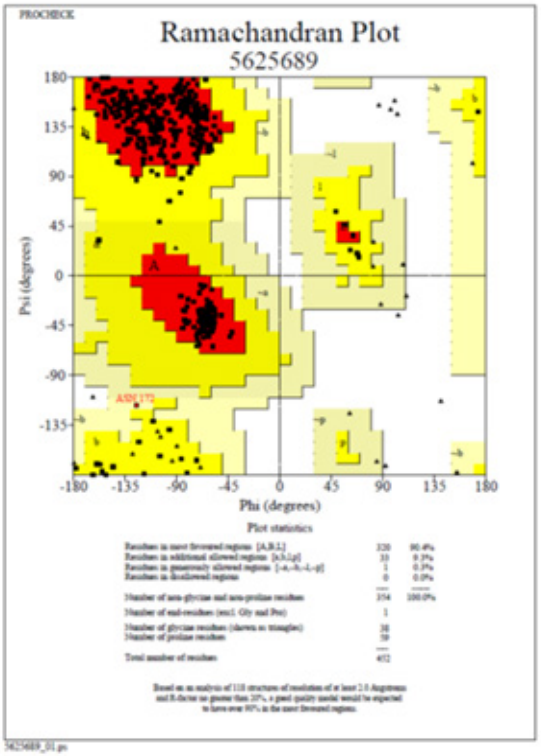

Fig. 6a: Ramachandran plot analysis of *GORASP2* protein model

Ligand Generation and Docking

D'Angelo *et al.* [46] suggested the applicability of trabectedin in the treatment of neoplasias, in which *HMGA1* levels are overexpressed. The Zinc database and the Drugbank were used to accomplish the PDB files of trabectedin (DB05109) and other chemical structures with a similarity threshold of 0.6 to trabectedin. This led to the creation of a total of 20 ligands for the purpose of docking to the protein models of *HMGA1* and *GORASP2* using Autodock. The number of points in the x-, y-, and z- dimensions were fixed at 40, and the center grid was X: 6.445; Y: 28.36; Z: 3.676, along with the spacing (angstrom) at one, in the case of *HMGA1* and for *GORASP2* the number of points in the x-, y-, and z- dimensions were fixed at 78, 56, and 90, respectively, and the center grid was X: -30.453; Y: 13.652; Z: -12.501, along with spacing (angstrom) at one. The docking results (Table 5) displayed the affinity (Kcal/mol), and the most stable docked ligand (having highest affinity or least energy) was found to be Vinflunine (DB11641) (Figs 7a and b) and Lurbinectedin (DB12674) (Figs 8a and b) when docked against *HMGA1* and *GORASP2* proteins, respectively.

In the current study, the clustering result from SOM and K-means led to the identification of 14 genes having a similar expression. Furthermore, the evolutionary similarity of *HMGA1* and *GORASP2* is clearly observed from the phylogenetic tree. The technique of comparative

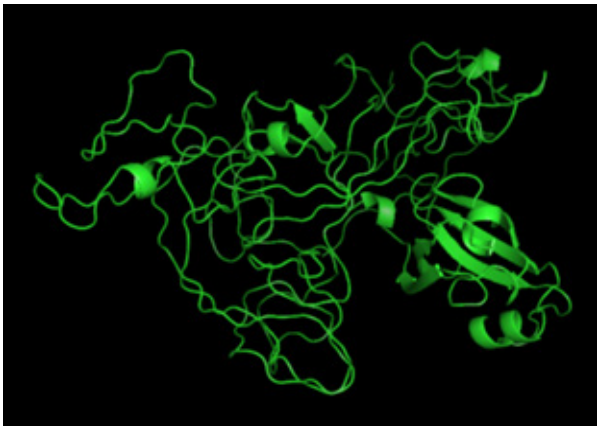

Fig. 6b: *GORASP2* protein model

Table 4: Validation of loop models of *HMGA1* and *GORASP2* proteins

	Validation by Rampage			*Validation by SAVS*			
Loop models of proteins	*Residues in favored regions (%)*	*Residues in allowed regions (%)*	*Residues in outlier regions (%)*	*Residues in most favored regions (%)*	*Residues in additional allowed regions (%)*	*Residues in generously allowed regions (%)*	*Residues in disallowed regions (%)*
Loop model *HMGA1*	96.2	2.9	1	92.2	7.8	0	0
Loop model *GORASP2*	95.6	4.2	0.2	90.4	9.3	0.3	0

Table 5: Docking of ligands against *HMGA1* and *GORASP2* proteins using Autodock 4

	Ligands		*Docking affinity (kcal/mol)*	
S. No.	*Name*	*DrugBank code*	*HMGA1*	*GORASP2*
1	Trabectedin	DB05109	-9.5	-12.4
2	Lurbinectedin	DB12674	-11	-14.8
3	Zalypsis	DB12454	-10.1	-12
4	Penimepicycline	DB13264	-8.4	-10.8
5	Rifabutin	DB00615	-9.8	-12.8
6	Vincristine	DB00541	-10.4	-12.9
7	Vinorelbine	DB00361	-10.1	-12.9
8	Naldemedine	DB11691	-9.1	-11.3
9	Anhydrovinblastine	DB12586	-10.3	-12.5
10	Vinflunine	DB11641	-11.6	-13.6
11	Chlorophyll A	DB02133	-10.3	-12.4
12	Vindesine	DB00309	-10.4	-12.4
13	CGP 4832	DB04220	-10.7	-13.9
14	Siroheme	DB02832	-10.3	-11.7
15	Cefiderocol	DB14879	-8.6	-10.6
16	Bietaserpine	DB13575	-8.9	-10.5
17	2-Phenylheme	DB03906	-8.8	-11.6
18	Ceftolozane	DB09050	-8.5	-8.9
19	Ceftobiprole	DB14733	-8.2	-10.6
20	Bromocriptine	DB01200	-8.9	-11.2

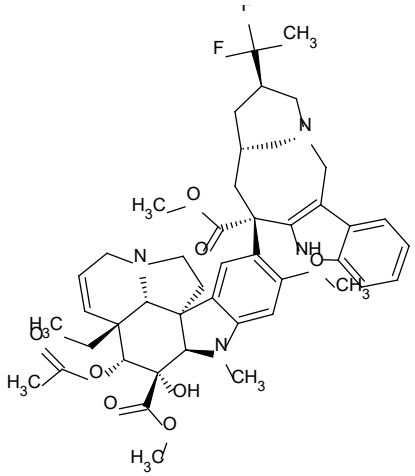

Fig. 7a: Chemical structure of ligand DB11641 (vinflunine)

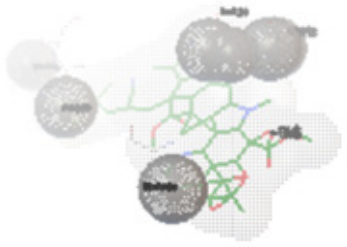

Fig. 7b: Docking of ligand DB11641 (vinflunine) against *HMGA1* protein

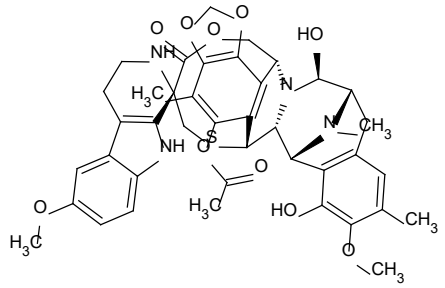

Fig. 8a: Chemical structure of ligand DB12674

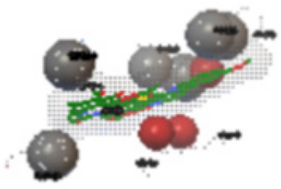

Fig. 8b: Docking of ligand DB12674 (lurbinectedin) against *GORASP2* protein

modeling was used to design models of *HMGA1* and *GORASP2* proteins. These models were evaluated by the Ramachandran plot. The Zinc database and the Drugbank were used to extract the PDB files of trabectedin (DB05109) and other similar chemical structures. Molecular docking done against these proteins led to the fact that the ligand

DB11641 (vinflunine) shows favorable binding interaction with *HMGA1* protein, and ligand DB12674 (lurbinectedin) shows strong binding with *GORASP2* protein. The study suggests that these compounds have potential. Thus, it necessitates further research in developing novel inhibitors for the effective treatment of various types of *HMGA1* and *GORASP2* overexpressed cancers.

Acknowledgement

The author acknowledges the technical support of Mr. Ankur Mohan, M.D., Bio-EGICORE, Lucknow.

References

1. Shahjaman M, Kumar N, mollah mmH, Ahmed MS, Ara Begum A, Shahinul Islam SM, mollah MNH. Robust significance analysis of microarrays by minimum β-divergence method. Biomed Res Int. 2017; 2017:5310198.
2. Fujita A, Sato JR, Rodrigues LdeO, Ferreira CE, Sogayar MC. Evaluating different methods of microarray data normalization. BMC Bioinformatics. 2006; 7:469.
3. Shen P, hour A, Liu LD. Microarray meta-analysis to explore abiotic stress-specific gene expression patterns in Arabidopsis. Bot Stud. 2017; 58(1):22.
4. Xu Z, Huang L, Zhang H, Li Y, Guo S, Wang N, Wang SH, Chen Z, Weng J, Tao SC. PMD: A resource for archiving and analyzing protein microarray data. Sci Rep. 2016; 6:19956.
5. Conte A, Paladino S, Bianco G, Fasano D, Gerlini R, Tornincasa M, Renna M, Fusco A, Tramontano D, Pierantoni gm. High mobility group A1 protein modulates autophagy in cancer cells. Cell Death Differ. 2017; 24(11):1948-1962.
6. D'Angelo D, Mussnich P, Rosa R, Bianco R, Tortora G, Fusco A. High mobility group A1 protein expression reduces the sensitivity of colon and thyroid cancer cells to antineoplastic drugs. BMC Cancer. 2014; 14:851.
7. Hillion J, Roy S, Heydarian M, Cope L, Xian L, Koo M, Luo LZ, Kellyn K, Ronnett BM, Huso T, Armstrong D, Reddy K, Huso DL, Resar LMS. The high mobility group A1 (*HMGA1*) gene is highly overexpressed in human uterine serous carcinomas and carcinosarcomas and drives matrix metalloproteinase-2 (MMP-2) in a subset of tumors. Gynecol Oncol. 2016; 141(3):580-587.
8. Jun KH, Jung JH, Choi HJ, Shin EY, Chin HM. *HMGA1*/HMGA2 protein expression and prognostic implications in gastric cancer. Int J Surg. 2015; 24(PtA):39-44.
9. Palumbo Junior A, de Sousa VPL, Esposito F, De Martino M, Forzati F, Moreira FCB, Simao TA, Nasciutti LE, Fusco A, Ribeiro Pinto LF, Bessa Pereira Chaves C, Meireles Da Costa N. Overexpression of *HMGA1* figures as a potential prognostic factor in endometrioid endometrial carcinoma (EEC). Genes (Basel). 2019; 10(5). pii: E372.
10. Liu J, Mi B, Wang Y, Shi C, Mi X, Lu Y, Yu P. miR-26a suppresses osteosarcoma migration and invasion by directly targeting *HMGA1*. Oncol Lett. 2018; 15(6):8303-8310.
11. Resar LM. The high mobility group A1 gene: transforming inflammatory signals into cancer. Cancer Res. 2010; 70(2):436-439.
12. Sumter TF, Xian L, Huso T, Koo M, Chang YT, Almasri TN, Chia L, Inglis C, Reid D, Resar LM. The high mobility group A1 (*HMGA1*) transcriptome in cancer and development. Curr mol Med. 2016; 16(4):353-393.
13. Wang Y, Hu L, Zheng Y, Guo L. *HMGA1* in cancer: cancer classification by location. J Cell mol Med. 2019; 23(4):2293-2302.
14. Greco M, Arcidiacono B, Chiefari E, Vitagliano T, Ciriaco AG, Brunetti FS, Cuda G, Brunetti A. *HMGA1* and *mmP-11* are overexpressed in human non-melanoma skin cancer. Anticancer Res. 2018; 38(2):771-778.
15. Masciullo V, Baldassarre G, Pentimalli F, Berlingieri MT, Boccia A, Chiappetta G, *et al. HMGA1* protein over-expression is a frequent feature of epithelial ovarian carcinomas. Carcinogenesis. 2003; 24(7):1191-1198.
16. Devgan M. Structure prediction and *in silico* designing of drugs against fork head-box R2 protein. Res J Pharm Technol. 2018; 11(3):913-920.
17. Devgan M. Structure prediction and *in silico* designing of drugs against kallikrein protein 12. Int J Curr Pharm Res. 2017; 9(2):64-69.
18. Devgan M. Structure prediction and *in silico* designing of drugs against homeobox C8 protein. J Chem Pharm Res. 2016; 8(3):891-899.
19. Devgan M, Karar PK, Agarwal G, Mohan A, Gangwar P. *In silico* designing of drugs for the inhibition of AMF-HER2 complex in trastuzumab resistant breast cancer. Indian J Biotechnol. 2016; 15(3):292-298.
20. Devgan M. Structure prediction and insilico designing of drugs for the inhibition of EPH A10 tyrosine kinase receptor protein. Acta Biomedica Scientia. 2016; 3(2):86-94.
21. Devgan M. Homology modeling and molecular docking studies of DNA replication licensing factor minichromosome maintenance protein 5 (MCM5). Asian J Pharm Tech. 2015; 5(1):17-22.
22. Shatabda S, Newton MA, Rashid MA, Pham DN, Sattar A. The road not taken: retreat and diverge in local search for simplified protein structure prediction. BMC Bioinformatics. 2013; 14(suppl2): S19.
23. Petrey D, Honig B. Protein structure prediction: inroads to biology. mol Cell. 2005; 20(6):811-819.
24. Khor BY, Tye GJ, Lim TS, Choong YS. General overview on structure prediction of twilight-zone proteins. Theor Biol Med Model. 2015; 12:15.
25. Barrett T, Edgar R. Gene expression omnibus (GEO): microarray data storage, submission, retrieval, and analysis. Methods Enzymol. 2006; 411:352-369.
26. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, Rudnev D, Lash AE, Fujibuchi W, Edgar R. NCBI GEO: mining millions of expression profiles-database and tools. Nucleic Acids Res. 2005; 33(Database issue):D562-566.
27. Lutgendorf SK, DeGeest K, Sung CY, Arevalo JM, Penedo F, Lucci J 3rd, Goodheart M, Lubaroff D, Farley DM, Sood AK, Cole SW. Depression, social support, and beta-adrenergic transcription control in human ovarian cancer. Brain Behav Immun. 2009; 23(2):176-183.
28. Cole SW, Arevalo JM, Takahashi R, Sloan EK, Lutgendorf SK, Sood AK, Sheridan JF, Seeman TE. Computational identification of gene-social environment interaction at the human IL6 locus.
29. Sturn A, Quackenbush J, Trajanoski Z. Genesis: cluster analysis of microarray data. Bioinformatics. 2002; 18(1):207-208.
30. Wang J, Delabie J, Aasheim H, Smeland E, Myklebost O. Clustering of the SOM easily reveals distinct gene expression patterns: results of a reanalysis of lymphoma study. BMC Bioinformatics.2002; 3:36.
31. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. mol Biol Evol. 2011; 28(10):2731-2739.
32. National Center for Biotechnology Information[Internet]. Bethesda (MD): National Library of Medicine (US); *GORASP2* golgi reassembly stacking protein 2[*Homo sapiens* (human)][updated 2019 Oct 12; cited 2019 Nov 17];[About 10 p.].
33. Wikipedia contributors. FASTA format[Internet]. Wikipedia, The Free Encyclopedia; 2019 Dec 13, 08:26 UTC[cited 2020 Mar 14].
34. National Center for Biotechnology Information[Internet]. Bethesda (MD): National Library of Medicine (US); *HMGA1* protein[Homo sapiens][cited 2019 June 19];[About 2 p.].
35. National Center for Biotechnology Information[Internet]. Bethesda (MD): National Library of Medicine (US); Golgi reassembly stacking protein 2, 55kDa[*Homo sapiens*][cited 2019 June 19];[About 2 p.].
36. Wikipedia contributors. BLAST (biotechnology)[Internet]. Wikipedia, The Free Encyclopedia; 2020 Mar 8, 18:23 UTC[cited 2020 Mar 14].

37. Kelley LA, Mezulis S, Yates cm, Wass MN, Sternberg MJ. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc. 2015; 10(6):845-858.
38. Kelly L, Sternberg M. Phyre2[Internet]. South Kensington Campus (London): Structural Bioinformatics Group, Imperial College (UK);[cited 2019 June 20].
39. Research Collaboratory for Structural Bioinformatics. RCSB Protein Data Bank[Internet]. New Jersey, California: Rutgers, UCSanDiego (US);[cited 2019 June 20].
40. Kuntal BK. EasyModeller (Version 2.0)[Internet]. Hyderabad: Kuntal Kumar Bhusan, University of Hyderabad (India);[cited 2019 June 21].
41. Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis. 1997,* 18(15):2714-2723.
42. Procedure[Internet]. Powai (Mumbai): Department of Biosciences and Bioengineering, Indian Institute of Technology;[cited 2019 June 21].
43. Bakker PD, Lovell S. Rampage[Internet]. Cambridge (UK): Department of Biochemistry, University of Cambridge.[Cited 2019 June 21].
44. NIH MBI Laboratory for Structural Genomics and Proteomics. Structural Analysis and Verification Server (Version 5). UCLA MBI-SAVES web server[cited 2019 June 21].
45. Fiser A. ModLoop[Internet]. San Franscisco (CA): University of California (USA);[cited 2019 June 22].
46. D'Angelo D, Borbone E, Palmieri D, Uboldi S, Esposito F, Frapolli R, Pacelli R, D'Incalci M, Fusco A. The impairment of the high mobility group A (*HMGA1*) protein function contributes to the anticancer activity of trabectedin. Eur J Cancer. 2013; 49(5):1142-1151.
47. Wishart DS. DrugBank[Internet]. Canada: Canadian Institutes of Health Research, Alberta Innovates-Health Solutions, The Metabolomics Innovation Centre;[cited 2019 June 22].
48. Zinc12[Internet]. San Francisco (California): Shoichet Laboratory, Department of Pharmaceutical Chemistry, University of California, San Francisco.[cited 2019 June 22].
49. National Center for Biotechnology Information[Internet]. Bethesda (MD): National Library of Medicine (US); PubMed[cited 2019 June 22].
50. Morris gm, Lim-Wilby M. molecular docking. Methods mol Biol. 2008; 443:365-382.
51. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem. 2010; 31(2):455-461.

HOW TO CITE THIS ARTICLE: Devgun M, Lal S. Microarray data analysis, structure prediction, and *in silico* docking of drugs for inhibiting overexpression of High Mobility Group A1 in human malignant neoplasias. Int. J. Pharm. Sci. Drug Res. 2020;12(5):448-456. **DOI:** 10.25004/IJPSDR.2020.120504