



www.bioinformation.net **Volume 16(6)**

Research Article

From sequence analysis of DPP-4 to molecular docking based searching of its inhibitors

Awad Saeed Alsamghan¹, Afaf S. Alwabli^{2*}, Mohammed Abadi³, Safar A. Alsaleem⁴ Dalia Mohammed Anbari⁵, Amani Saleh Alomari⁶, Othman Alzahrani^{7&8} Qamre Alam⁹ & Mohammed Tarique^{10*}

¹Department of Family and Community Medicine, College of Medicine, King Khalid University, Abha, KSA-61421; ²Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah 21589, Kingdom of Saudi Arabia; ^{3&4}Department of Family and Community Medicine, College of Medicine, King Khalid University, Abha, KSA-61421.^{5&6}Department of Biochemistry, Faculty of Science, King Abdulaziz University, Jeddah 21589, Kingdom of Saudi Arabia; ⁷Department of Biology, Faculty of Sciences, University of Tabuk, Tabuk 71491, Kingdom of Saudi Arabia; ⁸Genome and Biotechnology Unit, University of Tabuk, Tabuk 71491, Kingdom of Saudi Arabia; ⁹Medical Genomics Research Department, King Abdullah International Medical Research Center, King Saud bin Abdulaziz University for Health Sciences, Jamia Millia Islamia, Jamia Nagar, New Delhi-110025, India. Dr. Afaf S. Alwabli- Email: afaalwabli@yahoo.com; Dr. Mohammed Tarique - E-mail id: tariqueaiims@gmail.com; *Corresponding author with equal contribution

Author contacts:

Awad Saeed Alsamghan - E-mail: dr.awadsaeed1@gmail.com; Afaf S. Alwabli - E-mail: afafalwabli@yahoo.com; Mohammed Abadi - E-mail:mabade@kku.edu.sa; Safar A. Alsaleem - Email:safar10abadi@hotmail.com; Dalia Mohammed Anbari - E-mail: danbari@stu.kau.edu.sa; Amani Saleh Alomari-E-mail: aalomari0027@stu.kau.edu.sa; Othman Alzahrani-E-mail: o-alzahrani@ut.edu.sa; Qamre Alam-E-mail: alamqa@ngha.med.sa; Mohammed Tarique - E-mail: tariqueaiims@gmail.com Submitted by Dr. Mohammad Tarique, Center for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, Jamia Nagar, New Delhi-110025, India; E-mail id: tariqueaiims@gmail.com

Received March 3, 2020; Revised April 19, 2020; Accepted May 7, 2020; Published June 30, 2020

Declaration on official E-mail:

The corresponding author declares that official e-mail from their institution is not available for all authors

Declaration on Publication Ethics:

The authors state that they adhere with COPE guidelines on publishing ethics as described elsewhere at https://publicationethics.org/. The authors also undertake that they are not associated with any other third party (governmental or non-governmental agencies) linking with any form of unethical issues connecting to this publication. The authors also declare that they are not withholding any information that is misleading to the publisher in regard to this article.

DOI: 10.6026/97320630016444



Abstract:

Literature data suggests that Dipeptidyl peptidase-4 (DPP-4) is a potential target for type 2 Diabetes Mellitus. Therefore, it is of interest to identify new DPP-4 inhibitors using molecular docking analysis. We document compounds such as STOCK1N-98884, STOCK1N-98881, and STOCK1N-98866 with optimal binding features with DPP-4 from the ligand database at https://www.ibscreen.com/ for further consideration.

Keyword: DPP-4, GLP-1, diabetes, docking analysis, inhibitor

Background:

Insulin resistance in type 2 diabetes and related issues are known [1]. Symptoms associated with the disease include retinopathy, edema, micro aneurysms, nephropathy outlines, symmetrical fringe neuropathy influencing engine and tactile nerves of the smaller attachments [2-4]. Several models of treatments using insulin, secretagogues (sulfonylureas and incretins) and hypoglycemias (biguanides, thiazolidinediones and a-glucosidase inhibitors) are currently available [5-10]. Inhibitors of the dipeptidyl peptidase-4 (DPP-4) are linked with the activities of GLP-1 and gastric inhibitory polypeptide (GIP) [7, 8]. Description of the structural

models for DPP-4 is known [15-17]. Therefore, it is of interest to identify molecules to inhibit DPP-4 using molecular docking analysis.

Methodology:

Sequence to structure modeling and docking analysis of DPP-4:

The DPP-4 protein sequence downloaded from GenBank was analyzed in a comprehensive using tools such as Clustal Omega, Pfam, Prosite, SMART, PANTHER, PHYLIP, STRING and InterProScan, molecular docking and ligand-protein analysis tools to glean valuable insights [11-22].

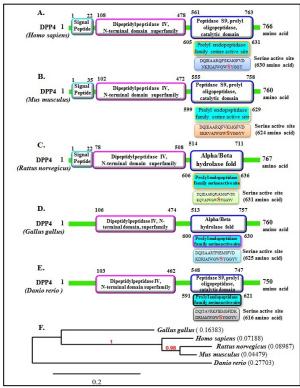


Figure 2: Domain and phylogeny analysis of DDP-4 in different organisms.



А.	
Horo MKTPWKVLLGLLGAAALVTIITVPWLLNKGTDDATADSRKTYTLTDYLKNTYRLKLYSL 60	Homo HFDKSKKYPLLLDVYAGPCSQKADTVFRLMMATYLASTENIIVASFDGRGSGYQGDKINH 592
Mus MKTPWKWLLGLLGVAALVTIITVPTVLLSKDEAAADSRRTYSLADVLKSTFRVKSYSL 58	Mus HFDKSKKYPLLLDVYAGPCSQKADASFRLMWATYLASTENIIVASFDGRGSGYQGDKINH 586
Rattus MKTPWKVLLGLLGVAALVTIITVPVVLLNKDEAAADSRRTYTLADVLKNTFRVKSYSL 58	Rattus HFDKSKKYPLLIDVYAGPCSQKADAAFRLMWATYLASTENIIVASFDGRGSGYQGDKINH 593
Gallus MKTLLKNLLGLVGVAVVITVIAVPLALLT-GESIPESDSRSTYTLENVLMDVVYKTHNL 59	Gallus HLDSSKKYPLLLEVYAGPCSQKVDHVFRINWATYLASTEQIIVASFDGRGSGYQGDEIMH 587
Danio -NGCNKVCVALVGAVVVITLIAIPTAIYVNRDDSALKRTYSFDDFYNDTIRYKTYNL 56	Danio NFDSSKKYPLLIDVYGGPASQNIDYVFRLEWATYLCSTERIIVASFDGRGSGFQGDEINH 578
t stitustitatus is site as in tat	Hammentation a neurophic monometeria
Homo RWISDHEYLYKQE-MNILVFNAEYGNSSVFLENSTFDEFGHSINDYSISPDGQFILLEYN 119	Homo AINRRLGTFEVEDQIEAARQFSKNGFVDNKRIAINGNSYGGYVTSNVLGSGSGVFKCGIA 652
Mus WWVSDFEYLYKQE-MNILLLNAEHGNSSIFLENSTFESFGYHSVSPDRLFVLLEYN 113	Mus AINRRLGTLEVEDQIEAARQFVKNGFVDSKRVAINGNSVGGVVTSNVLGSGSGVFKCGIA 646
Rattus RWSDSEYLYKQE-MNILLFNAEHGNSSIFLENSTFEIFGDSISDYSVSPDRLFVLLEYN 117	Rattus AINKRLGTLEVEDQIEAARQFLKNGFVDSKQVAINGNSYGGYVTSNVLGSGSGVFKCGIA 653
Gallus QWISGNQYLHETSNGNILRFDAETGTSSVVLLNTTISIHEATTAILSPDQRFALLQYK 117	Gallus AINRRLGTYEVEDQISAARTFSEMSFVDKDRIAIWGWSYGGYVTSMVLGSGSGVFKCGIA 647
Danio RWISDNEYLHKINEGHIYLHNAETKESSVYLSNSTFAQVDATDYILSADRKFAAFESN 114	Danio AIYERLGTYEVEDQITAWRKFIENGFIDKDRIAMWGNSYGGYVTSMALGSGSGLFKCGIA 638
Hono YVKQWRHSYTASYDIYDLNKRQLITEERIPIWTQWVTWSPVGHKLAYWWMDIYVKIEPN 179	Hono VAPVSRNEYYDSVYTERYNGLPTPEDNLDHYRNSTWISRAENFKQVEYLLIHGTADDIWH 712 Nus VAPVSRNEYYDSVYTERYNGLPTPEDNLDHYRNSTWISRAENFKQVEYLLTHGTADDIWH 786
Mus YVKQNRHSYTASYNIYDVNKRQLITEEKIPINITQNITNSPEGHKLAYWKNDIYVKVEPH 173	
Rattus YVKQNRHSYTASYSIYDLNKRQLITEEKIPINTQNITNSQEGHKLAYVMKNDIYVKIEPH 177	
Gallus YEKLNRHSYTASYHIYDFNTSSILODALLPNDTQYISNSPVGHKLAYVMMNIYIKASPT 177	
Danio YSKONRHSFTASYSIYNVESGEFLSKVQIPHVTQLLTWAPVGNKLAYVWNFNIYLKASAT 174	Danio VAPVAKNEYYDAVYTERYMHRPQDNFESYKNSTVTDRAKNFKSVQYLLVHGTADDNVH 696
Hono LPSYRITWIGKEDIIYWGITDWYEEEVFSAYSALWISPWGTFLAYAOFNDTEVPLIEY5 239	Homo VAPVSRNEYYDSVYTERYNGLPTPEDNILDHYRNSTVNSRAENFKOVEYLLIHGTADDIWH 712
Mus LPSHRITSTGEENVIYIGITDWYEEEVFGAYSALWISPINITFLAYAQFNDTGVPLIEYS 233	Mus VAPVSRWEYYDSVYTERYMGLPIPEDNLDHYRNSTVMSRAEHFKOVEYLLIHGTADDNVH 706
Rattus LPSHRITSIGEENVITHGITONVICEELIFGAYSALIMISPHRITPGATAGPHDIGVPLIETS 255	Rattus VAPVSRNEYYDSVYTERYNGLPTPEDNLDHYRNSTWNSRAENFKOVEYLLIHGTADWHG 713
Gallus AAPVOITSNGEENKIFNGIPDNVYEEENFGSHSALNNSPNGNFVAYAAFNDTEVPVIEYS 237	Gallus VAPVSRW0YYDSIYTERYMGLPTESDNLRNYNSSTVMARAEKFKEVEYLLIHGTADDNVH 707
Danio AEAVOVTHNGKGNEILNGVPDNVYEEEVFASNEAINNSPOGYLAYLOVNDTGVHSIEYS 234	Danio VAPVAKNEYYDAVYTERYNHR PQONFESYKNSTVTDRAKNEKSVQYLLVHGTADDIWH 696
	mananananananananananananananananananan
Hono FYSDESLOYPKTVRVPYPKAGAWIPTVKFFWNITDSLSSVTNATSIOITAPASNLIGDHY 299	Homo FOOSAOISKALVDVGVDFOAMWYTDEDHGIASSTAHOHIYTHMSHFIKOC 762
Mus FYSDESLOVPKTWIPYPKAGAWIPTVKFFIVIIDSLSSSSSAAPIQIPAPASVARGOHY 293	Mus FQQ5AQISKALVDAGVDFQAIWYTDEDHGIASSTAHQHIYSHISHFLQQC 756
Rattus FYSDESLQYPKTVWIPYPKAGAWNPTVKFFIVNTDSLSSTTTTIPMQITAPASVTTGDHY 297	Rattus RRPNDRODHSSPAHLFPHEPFPPAVLLLTLANDGSPOLTOEHTCPHYLKTA 764
Gallus FYSEDTLOYPKTIRIPYPKAGAKNPTVKFFIVDIONLPDFNSTEISPPAEIKSGOHY 294	Gallus FOOAAOISKALVDAEVDFOANNYTDKDHGISG-OAHKHIYTMSHFIKOC 756
Danio LYGIDOYPVTVFVPYPKAGSVIPRARLFVIDVENPSROSEVWPKSVGSGDHY 287	Danio FOQAAOISKALVENOVDFEAMVYTDKDHSLSG-KARYHLYTHLNHFLKNC 745
	Conto regione regioner control control control and the second sec
Hono LCDVTWATQERISLQWLRRIQNYSVMDICDYDESSGRWHCLVARQHIEMSTTGWVGRFRP 359	Horo FSLP- 766
Mus LCDWWATEERISLQWLRRIQNYSVMAICDVDKINLTWICPSEQQHVEMSTTGWGRFRP 353	Mus FSLH- 760
Rattus LCDWAWVSEDRISLQWLRRIQNYSVMAICDYDKTNLWMICPTTQEHIETSATGWVGRFRP 357	Rattus LLR 767
Gallus LSWTWYTDERICLQWLRRIQNYSVLTICDFESATGNWTCPQEKQLLEESTTGWIGRFQP 354	Gallus FSLP- 760
Danio LSTVTWVTDDRLAVQWMPRRQDSVLLQIYDYDGTKWKESTKFEQKSKTGWVGRYFP 343	Danio FAEKK 750
*, *,*,**.; * *; ;; * *;; , *, ;; ; * *i;; , *,	
Homo SEPHFTLDGNSFYKIISNEEGYRHICYFQIDKKDCTFITKGTWEVIGIEALTSDYLY 416	aligned termini profile
Mus AEPHFTSDGSSFYKTISDKDGYKHICHFPKDKKDCTFTTKGAWEVISTEALTSDYLY 418	D. helix-residues
Rattus AEPHFTSDGSSFYKIVSDKDGYKHICQFQKDRKPEQDCTFITKGAWEVISIEALTSDYLY 417	strand-residues coil-residues
Gallus SVPYFAPDNTTYYKVFSNTEGYKHIHYINGTEAP·····VPITEGKNEVISIAAVTKYFLY 410	
Danio SAPYFAAMMISFYKWMSNDNGYKHLHYVNAGKATPITSGKNEVIYISKVTKDSIY 398	
Horo YISNEYKGNPGGRILLYKIQLSDYTKVTCLSCELNPERCQYYSVSFSKEAKYYQLRCSG 474	- / -
Mus YISNOYKENPGGNILYKIQLIDHTNVKCLSCELIPENCOYAVSFSKEAKYYOLGCNG 468	
Rattus YISNEYKENPGGRNLYKIQLTDHTNKKCLSCDLNPERCQYYSVSLSKEAKYYQLGCRG 475 Gallus YISNQNGENPGGRNLYKHLLESSP-KSTQCVSCDLNQERCQYYSASFSKDAQYYQLNCLG 469	
Danio YVSNEHMARPGORNLYKISISSSGHSAPKCLTCALYEDRCOVISAYFSLNASYFRMDCYG 458	
1.11. 11 11.11.11.11.11.11.11.11.11.11.1	
Hono PGLPLYTLHSS-VI.DKGLRVLEDIJSALDK/RLQ-INV(MPSKKLDFIILNETKFWV(MILPP 532	
Rattus PGLPLYTLHRS-TDQKELRVLEDNSALDKMLQ-DVQMPSKKLDFIVLNETRFWYQMILPP 533	
Gallus PGLPHSTLHRS-SDDQVLRYLENNTELENSLK-DIQHPSKKLGSITVGGYNLWYQHILPP 527	1 Man marked 1
Danio PGLPLFTMDNRGPAKETQVLEDNKKLENILTTELLMPTKKRGTLKIAGFDLWVQMMFPP 518	0 100 200 300 400 500 600 700 80
igure 1: (A) MSA of DDP-4 from different or	ganisms (Homo sapiens, Rattus norvegicus, Mus

Figure 1: (A) MSA of DDP-4 from different organisms (*Homo sapiens, Rattus norvegicus, Mus musculus, Danio rerio* and *Gallus gallus*) (B) Secondary structure information on DDP-4



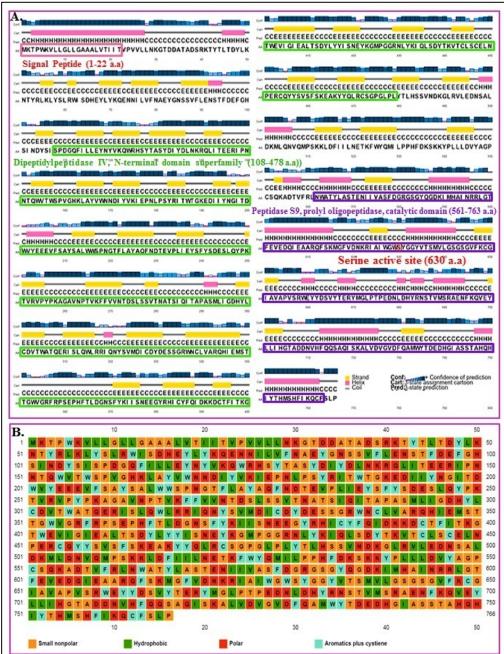
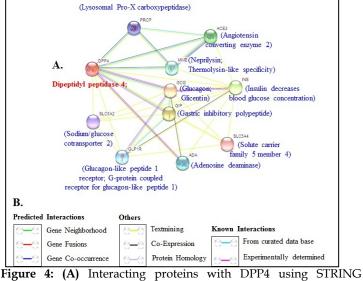


Figure 3: (A) Secondary structure analysis of human DDP-4 (*Homo sapiens*). **(B)** Small non-polar, hydrophobic, polar and aromatic plus cysteine residues in human DDP-4.





v10.database. (B) Explanation of interactions shown.

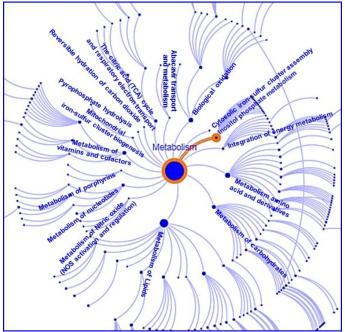


Figure 5: DDP-4 linked pathways.

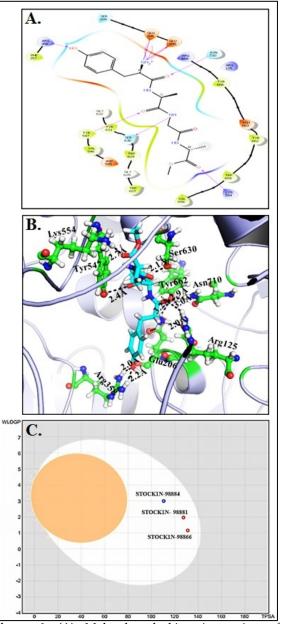


Figure 6: (A) Molecular docking interaction of DPP4 with STOCK1N-98884. (B) Cartoon interpretation of DPP4 with compound STOCK1N-98884. (C) Boiled-egg plot.

Table 1: Lowest binding energy for the Ligands-Protein interaction, along with scores for various interaction types, as detected by GLIDE



GScore; Glide extra precision scores (kcal/mol) Lipophilic E Vdw; Chemscore lipophilic pair term and fraction of the total protein-ligand vdw energy

Compounds ID	Binding Energy	GScore	Lipophilic E vdw	H-bond	Electro	Protein ligands interaction
	MM-GBSA (kcal/mol)					
STOCK1N-98884	-72.7837	-11.56	-2.91	-6.87	-2.01	Glu:205, Glu:206, Try :547, Ser:630 and Asn710
STOCK1N-98881	-61.2792	-10.2	-3.37	-4.44	-2.41	Arg:125, Glu:205, Glu:206, Lys:554, Trp:629 and Ser:630
STOCK1N-98866	-59.2571	-9.58	-2.46	-3.65	-3.19	Arg:125, Try :547, Lys:554 and Trp:629
Known Inhibitor						
Linagliptin	-44.1282	-6.79	-2.22	-2.61	-0.34	Try :547, Ser:630 and Asn710

Electro; Electrostatic rewards Protein ligands interaction; p-p stacking, p-cat interaction and hydrogen bond between the ligands and protein

Table 2: Evaluation of drug-like properties of the lead molecules by Qikprop Maestro 10.5 molecular docking suite

	0 1 1		/~			
Molecule	QPlog Po/w (-2.0 to 6.5)	Q P log HERG (acceptable ange: above -5.0)	QPP Caco (nm/s) <25 – poor >500 – great	Q P log, BB (-3 to 1.2)	QPP MDCK (nm/s)	Q Plog Kp (-8.0 to - 0.1)
STOCK1N-98884	-0.30	-1.056	131.328	-0.94	70.119	-2.798
STOCK1N-98881	3.376	-0.015	283.926	-0.628	485.3	-2.406
STOCK1N-98866	2.219	-3.804	143.431	-1.641	60.643	-3.179

Predicted IC50 value for blockage of HERG K+ channels; (acceptable range above -5.0) Molecule STOCK, InterBioScreen's library (IBS), Q P log Poct; was predicted partition coefficient of octanol/gas, (8.0 to 35.0); QPP Caco, predicted apparent Caco-2 cell permeability in nm/s. Caco-2 cells is a model for the gut blood barrier (nm/s) <25 – poor, >500 – great. Q P log BB, predicted brain/blood partition coefficient; QPP MDCK, predicted apparent MDCK cell permeability in nm/s. MDCK cells are considered to be a good mimic for the blood-brain barrier; (nm/s) <25 – poor, >500 – great; Q P log KP, Predicted skin permeability; Q P log Khsa Prediction of binding to human serum albumin; (acceptable range -1.5 to 1.5)

Table 3: Boiled egg parameters

Molecule	MW	TPSA	XLOGP3	MLOGP	GI absorption	BBB permeant
STOCK1N-98884	430.88	159.85	-0.30	-0.66	High	No
STOCK1N-98881	624.04	158.30	2.99	0.23	Low	No
STOCK1N-98866	421.40	127.08	2.81	0.93	High	No

Table 4: Biological activity spectrum of compounds (Pa - Active; Pi - Inactive)

Molecule	Pa	Pi	Activity
STOCK1N-98884	1.219	0.449	Anti-diabetic
STOCK1N-98881	1.812	0.642	Anti-diabetic
STOCK1N-98866	1.121	0.318	Anti-diabetic

Results & Discussion:

A comprehensive analysis of DDP-4 using sequence and structure information is highly relevant in the fight against T2DM with reference to known data in the literature. The Multiple Sequence Analysis (MSA) of DDP-4 from different organisms such as Homo sapiens (DPP4, 766 amino acid), Rattus norvegicus (DPP4, 767 amino acid), Mus musculus (DPP4, 760 amino acid), Danio rerio (DPP4, 750 amino acid) and Gallus gallus (DPP4, 760 amino acid) is given in Figure 1. Secondary structure information of DDP-4 is also shown in Figure 1. Domain and phylogeny analysis of DDP-4 in different organisms is given in Figure 2. The Secondary structure analysis of human DDP-4 along with small nonpolar, hydrophobic, polar, and aromatic plus cysteine residues in human DDP-4 is shown in Figure 3. Protein-protein interaction network linked to DDP-4 is shown in Figure 4. We further show the DDP-4 associated pathways in Figure 5. The molecular docking interaction of DPP4 with STOCK1N-98884 is given in Figure 6 and Tables 1 to 4. This

information gleaned from the analysis of DDP-4 is relevant in the design and development of novel compounds in combating the disease.

Conclusion:

We document compounds STOCK1N-98884, STOCK1N-98881, and STOCK1N-98866 from the IBS ligand database with optimal binding features with DPP-4 towards combating T2DM.

Conflict of interest:

There are no conflicts of interest.

Acknowledgments:

The authors sincerely thank Almanac Life Science India Pvt. Ltd. for analysis and support.



References:

- [1] Care D Diabetes Care 2019. 42: S81
- [2] Nilsson E et al. Diabetes 2014 63: 2962 [PMID: 24812430]
- [3] Monami M *et al. Obesity and Metabolism* 2013 **15**: 112 [PMID: 22925682]
- [4] Dunkley AJ et al. Diabetes Care 2014 37: 922 [PMID: 24652723]
- [5] Guido RV et al. estudos avançados 2010 24: 81
- [6] Gautier J *et al. Diabetes and metabolism* 2005 **31**: 233 [PMID: 16142014]
- [7] Elrishi M et al. Practical Diabetes International 2007 24: 474
- [8] Bjelke JR *et al. Journal of Biological Chemistry*, 2004 **279**: 34691[PMID: 15175333]
- [9] Gross JL et al. Arquivos Brasileiros de Endocrinologia and Metabologia 2002 **46**: 16
- [10] Quevillon E *et al. Nucleic acids research* 2005 **33**: W116 [PMID: 15980438]
- [11] Tarique M et al. Frontiers in microbiology 2017 8: 130 [PMID: 28232818]

- [12] Castresana J Molecular biology and evolution 2000 17: 540 [PMID: 10742046]
- [13] Malik A et al. Bioinformation 2020 16: 223
- **[14]** Afaq S et al. Bioinformation 2020 **16**: 229
- [15] Nugent T and Jones DT BMC bioinformatics 2013 14: 276 [PMID: 24047460]
- [16] Anisimova M and Gascuel O Systematic biology 2006 55: 539 [PMID: 16785212]
- [17] Mishra J and Bhalla US Biophysical Journal 2002 83: 1298 [PMID: 12202356]
- [18] Bhardwaj SK *et al. Brain research bulletin* 1999 **49**: 155 [PMID: 10435778]
- [19] Johnson TO et al.Nature Reviews Drug Discovery 2002 1: 696 [PMID: 12209150]
- [20] Friesner RA et al. Journal of medicinal chemistry 2004 47: 1739 [PMID: 15027865]
- [21] Tripathi SK et al. Journal of theoretical biology 2013 334: 87 [PMID: 23727278]

Edited by P Kangueane

Citation: Alsamphan *et al.* Bioinformation 16(6): 444-451 (2020)

License statement: This is an Open Access article which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. This is distributed under the terms of the Creative Commons Attribution License

Articles published in BIOINFORMATION are open for relevant post publication comments and criticisms, which will be published immediately linking to the original article for FREE of cost without open access charges. Comments should be concise, coherent and critical in less than 1000 words.



