BIOINFORMATION
Discovery at the interface of physical and biological sciences

open access www.bioinformation.net Hypothesis

Volume 9(16)

Homology modeling and structural validation of tissue factor pathway inhibitor

Piyush Agrawal, Zoozeal Thakur & Mahesh Kulharia*

Centre for Bioinformatics, MD University, Rohtak 124001, India; Mahesh Kulharia – Email: kulharia@gmail.com; *Corresponding author

Received August 16, 2013; Accepted August 17, 2013; Published September 23, 2013

Abstract:
Blood coagulation is a cascade of complex enzymatic reactions which involves specific proteins and cellular components to interact and prevent blood loss. The coagulation process begins by either "Tissue Dependent Pathway" (also known as extrinsic pathway) or by "contact activation pathway" (also known as intrinsic pathway). TFPI is an endogenous multivalent Kunitz type protease inhibitor which inhibits Tissue factor dependent pathway by inhibiting Tissue Factor:Factor VIIa (TF:FVIIa) complex and Factor Xa. TFPI is one of the most studied coagulation pathway inhibitor which has various clinical and potential therapeutic applications, however, its exact mechanism of inhibition is still unknown. Structure based mechanism elucidation is commonly employed technique in such cases. Therefore, in the current study the generated a complete TFPI structural model so as to understand the mechanistic details of it's functioning. The model was checked for stereochemical quality by PROCHECK-NMR, WHATIF, ProSA, and QMEAN servers. The model was selected, energy minimized and simulated for 1.5ns. The result of the study may be a guiding point for further investigations on TFPI and its role in coagulation mechanism.

Background:
Blood coagulation pathway is a complex biological mechanism where specific proteins and cellular components interact to prevent blood loss **[1].** Coagulation is an important part of haemostasis. Haemostasis system allows blood to remain in fluid form in plasma and prevents excessive bleeding during vascular injury. The normal coagulation process begins with the "Tissue Dependent Pathway", initiated by the formation of complex between Factor VIIa and Tissue Factor (TF). Blood coagulation is well regulated patho-physiologically by an important Kunitz type serine protease inhibitor known as TFPI (Tissue Factor Pathway Inhibitor). TFPI is an endogenous anticoagulant with an acidic amino-terminal and basic carboxy-terminal, synthesized by endothelial cells and most of them (approx. 80%) interact with the wall and rest of them circulates in plasma. TFPI are present in different concentration. 50-60% of the circulating TFPI are bound to lipoprotein, 20% of total TFPI are carrier free and approximately 10% of TFPI are confined to platelets **[2].** TFPI is a potent inhibitor of Factor VII::TF complex and its action is regulated by the presence of Factor Xa **[3].** TFPI consists of 3 Kunitz domains each having specific functions. Domain1 binds to Factor VII::TF complex active site whereas domain2 binds to Factor Xa active site thus inhibiting them and regulating coagulation initiation **[4,5,6,7].** Function of domain3 has yet to be determined but it may be involved in lipoprotein-TFPI association [**8**]. The basic and positively charged C-terminus of TFPI is required to bind cell surfaces and cell bound TFPI mediates the internalization and degradation of FX and down regulation of surface TF/FVIIa activity **[6, 8].**

The residues of FVIIa that interacts with Kunitz domain 1 and that of FXa that interacts with Kunitz domain 2 have been reported earlier **[9-11].** Among others, D11, R20 and E46 are the residues of Kunitz domain 1 which are important for interaction with FVIIa. Similarly among others, Y17, R32 and E46 are the residues of Kunitz domain 2 which are important for interactions with FXa **[12].** Actual interaction of how Kunitz domain 1 interacts with TF/FVIIa complex and Kunitz domain 2 interacts with FXa is still unknown. Detailed study of TFPI still needs to be done in order to understand the mechanism of TFPI interaction with TF/FVIIa complex and FXa and how it inhibits the tissue factor dependent pathway. Hence the present

 © 2013 Biomedical Informatics

paper enlists some of the physiochemical and functional properties of TFPI and provides information into its three dimensional structure.

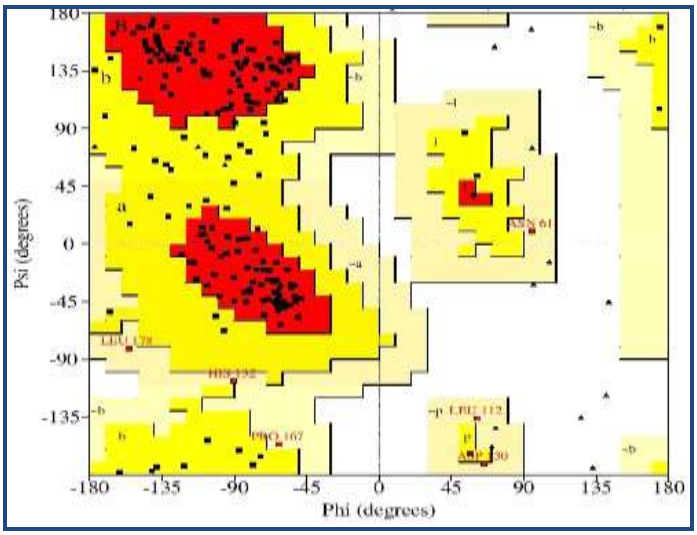

Figure 1: Ramachandran Plot analysis of TFPI. The plot statistics are: total number of residues-220 with 76% in most favored regions [A, B, L], 21.4% additional allowed regions [a,b,l,p], 2.6% in generously allowed regions and 0% in disallowed regions. Number of glycine residues (labelled as G) are 19 and proline residues (labelled as P) are 8.

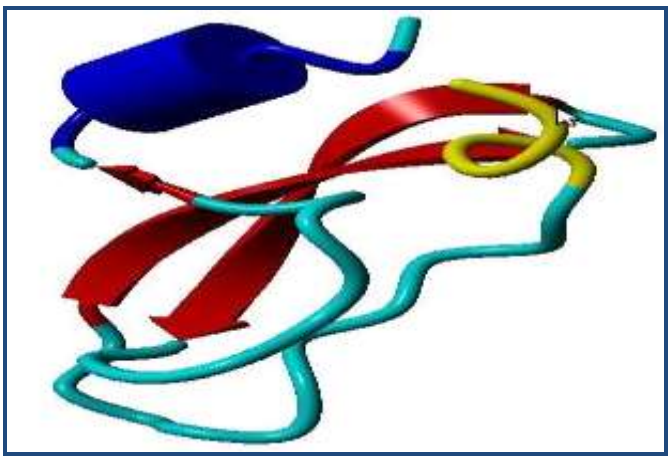

Figure 2: Kunitz domain 1 model generated using YASARA

Methodology:

The study was conducted using Intel(R) Core (TM) i3-2310 M CPU @2.10Ghz 4 core processor and 64 bit operating system.

Multiple Sequence Alignment and Homology Modeling

PDB file of Factor VII::TF complex (PDB ID: 2A2Q, Resolution: 1.80), TFPI sequence (Domain1), Domain2 sequence (PDB ID: 4DTG, Resolution: 1.80), Domain3 sequence (PDB ID: 1IRH, Resolution: Not Applicable) was downloaded which was used as template for building model of Domain1. In order to build model of Kunitz1 domain, Multiple Sequence Alignment was done between full length TFPI sequence and Domain 2 and domain 3 sequence. High resolution (1.80 A) structure of Kunitz2 domain (PDB ID: 4DTG) was selected as template to build the model of K1 domain because of more homology. Model construction and regularization (including geometry optimization) of model was done by optimization protocol in YASARA. The energy of model was minimized using the standard protocols of combined application of simulated annealing, conjugate gradient and steepest descent.

Loop construction was done to join all the 3 Kunitz domain of TFPI with each other. For loop construction, Loopy Software was used which was downloaded from the site "bhapp.c2b2.columbia.edu/software/cgi-bin/software.pl?input=Loopy". Model of Kunitz domain1 and Kunitz domain 2 structures were joined in a single coordinate file using inhouse perl script. Output file was then utilized for the loop construction and the sequence given for loop construction was "RDNANRIIKTTLQQ". The loop was made for the missing atom number from 59-72 in the jointed file. Kunitz 3 domain pdb file was then joined further in the output file obtained after loop construction between Kunitz 1 domain and Kunitz 2 domain. Similarly, loop construction was done in between Kunitz 2 domain and Kunitz 3 domain for the missing atom number from 133-162 and the sequence given for loop construction was "NGFQVDWYGTQLNAVWNSLTPQSTKVPSLF" hence finally leading to generation of complete model of TFPI protein having all the 3 Kunitz domains joined to each other.

Model Refinement

The newly constructed model was solvated and subjected to energy minimization using the steepest descent and conjugate gradient technique to eliminate unwanted contacts between structural water molecules and protein atoms. In this study, MD simulation study was undertaken by using YAMBER3 **[14]** package for the model refinement, which was used to reduce the steric clashes between residues. The constructed TFPI model had to be refined in order to stabilize the backbone. The data obtained after simulation was analyzed for trajectory projection.

Model Validation

Accuracy of predicted model and its stereo chemical properties was evaluated by PROCHECK-NMR **[15].** The model was selected on the basis of various factors such as overall G-factor, no. of residues in core allowed, generously allowed and disallowed regions in Ramachandran plot (**Figure 1**) The model was further analyzed by WHATIF **[16],** QMEAN **[17, 18]** and ProSA **[19].** ProSA was used for the display of Z-score and energy plots.

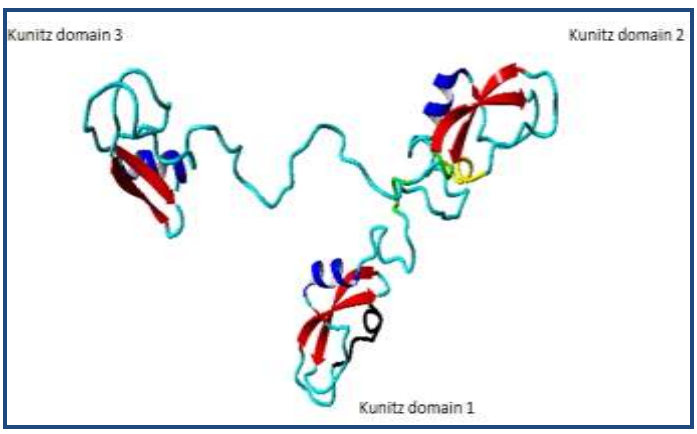

Figure 3: Complete model of TFPI molecule consisting 3 Kunitz domains connected to each other via loops.

© 2013 Biomedical Informatics

Results & Discussion:

Model Building

Sequence alignment of TFPI Kunitz domain 1 with sequences of Kunitz domain 2 and Kunitz domain 3, revealed more sequence homology with Domain 2 (ID= 47%) which was selected as template for the model building of Kunitz domain 1. To build the model 6 times PSI-BLAST was done with the maximum E-value allowed for template being 0.005. Maximum number of templates considered for model building was 6 along with maximum of 5 ambiguous alignment, 4 oligomerization state and number of unaligned loop residue to add to termini being 10. Using domain 2 sequence modeling of Kunitz domain 1 was done with the help of YASARA (**Figure 2**). After model construction of Kunitz domain1 loop construction was done to join domains. Two loops were constructed using Loopy software. The first loop joins the domain 1 and 2 and similarly, second loop join the domain 2 and 3. After joining domains together, complete model of TFPI was generated **(Figure 3).**

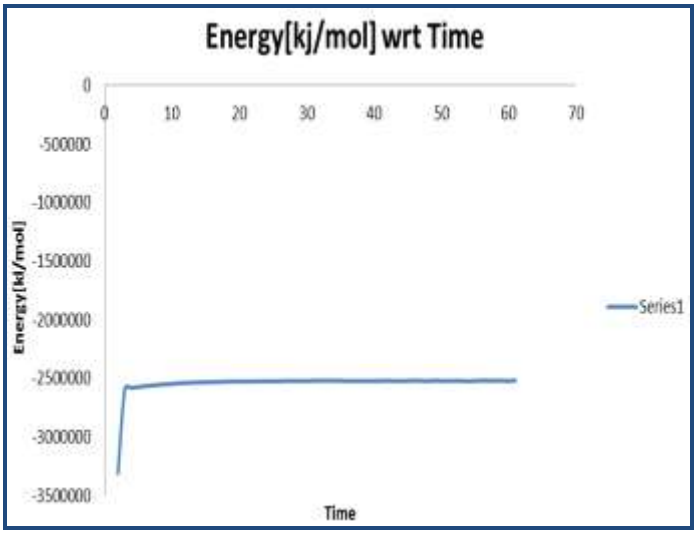

Figure 4: Trajectory Data/plot of Energy Simulation vs Time of modeled TFPI

Model Refinement

Model refinement was carried out to improve accuracy of TFPI model. The newly constructed model was solvated in a box of dimension 106.529 x 77.061 x 73.446 $Å^3$ with 3424 number of water molecules and subjected to sequential application of energy minimization techniques. In the initial phase the energy was minimized by Steepest Descent followed by Conjugate Gradient. Finally the global minima of TFPI model was obtained by Simulated Annealing. This was performed to minimize strain energies and eliminate unfavorable contacts between water molecules and protein atoms. YAMBER3 force field in YASARA dynamics was used for the model refinement, which was used to reduce the steric clashes between residues. The constructed TFPI model had to be refined in order to stabilize the backbone. After back bone refinement the energy was again minimized by the application of above mentioned protocol. The structure was then subjected to *nvt* ensemble (constant number of entities, isochoric and isothermal) based dynamic simulation for 1.5 ns. The temperature was 298K, density was 0.997 and the pH was 7.4 while carrying out refinement under physiological salt concentration of 150mM NaCl.

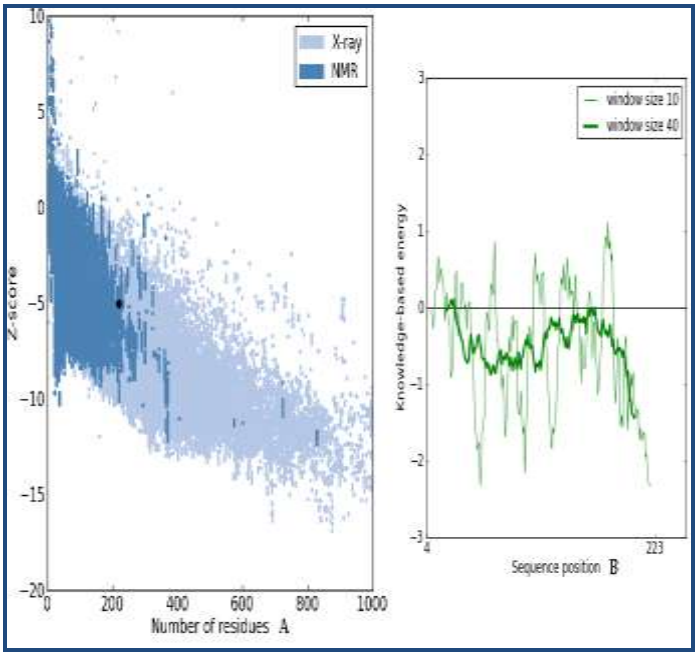

Figure 5: ProSA web service analysis of TFPI. ProSA-web Z-score of all protein chains in PDB determined by X- ray crystallography (light blue) or NMR-spectroscopy (dark blue) with respect to their length. The z-score of modeled TFPI is highlighted as large dots and the right graph is showing energy plot of modeled TFPI.

The trajectory was obtained for overall energy simulation of the modeled TFPI for 1500 picoseconds (ps) and it revealed that overall energy stabilized after a peak of -2589436.038 kJ/mol at 25 ps and tended to remain in plateau phase further for rest of the period **(Figure 4).** This reflected that simulation was achieved with stable energy for rest of the period (50-1475ps) for the TFPI. Almost the similar trajectory was obtained for the plots of different energy contributions against simulation run time. The contribution due to steric parameters like bond strain, dihedral angle, bond coloumb, Van der Waal was found maximum at 25 ps with the values of 329479.911 kJ/mol, 50047.973 kJ/mol, -3610916.379 kJ/mol and 512036.939 kJ/mol respectively, which stabilize further to a stationary phase for the rest of the period (50-1475ps), except dihedral angle which shows variations in the value **(Available with authors).** The contribution due to angle and planarity was slight different which shows maximum energy at the peak of 75 ps and 1475ps with the value of 130803.859 kJ/mol and 453.925 kJ/mol respectively, and then stabilize further for the rest of period (100-1475ps) **(Available with authors)**. These trajectory patterns support and validates the simulation profile of modeled TFPI. The trajectory pattern of energy due to RMSDs [A] :CA, Backbone and Heavy atoms differed from the trajectories of other parameters contributing to the overall energy of interactions of modeled TFPI **(Avauilable with authors).** The trajectory plots of energy due to RMSDs [A]: CA and Backbone showed a continuous increase with respect to time even after 1475ps. The deviation of trajectory plot of energy due to angle, planarity, RMSDs [A]:CA, and Backbone from other contributing parameters may be due to slow computational speed and performance available and lack of time to carry out further simulations.

© 2013 Biomedical Informatics

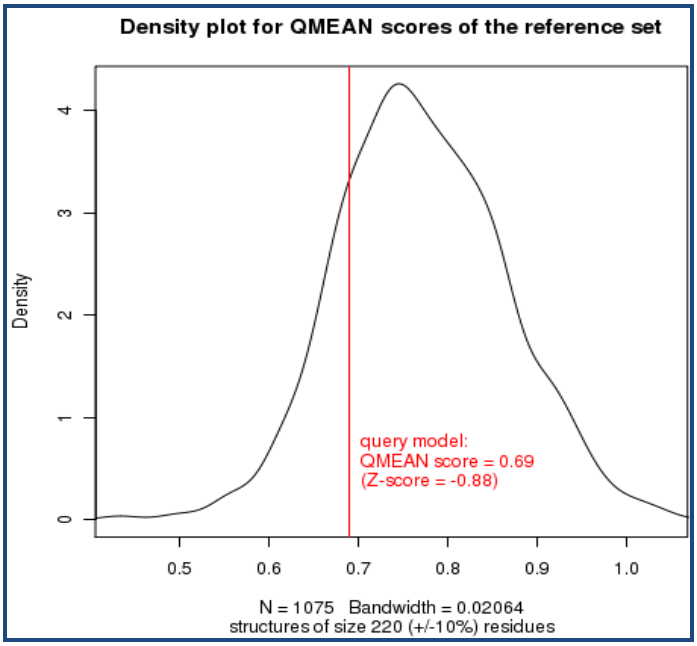

Figure 6: Density plot for QMEAN showing the value of Z-score and QMEAN score.

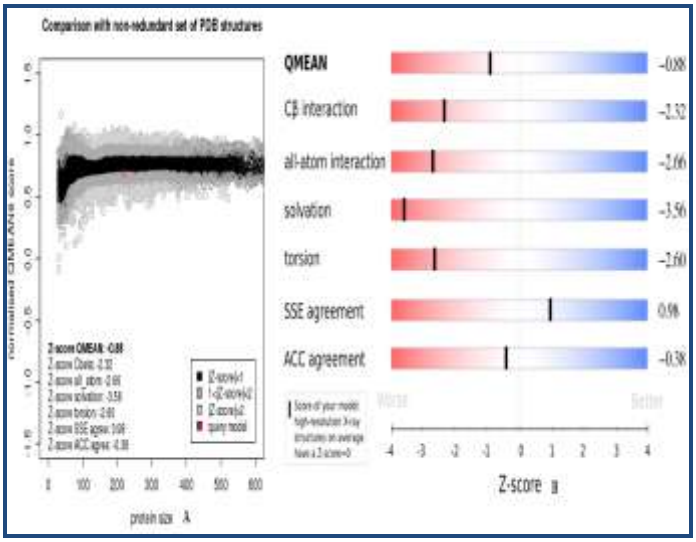

Figure 7: Plot showing the QMEAN value as well as Z-score.

The trajectory was also obtained for overall energy simulation of modeled TFPI for 1500 picoseconds (ps) with respect to residues present in modeled TFPI. The trajectory reflected that the highest value of different parameters such as RMSDs [A]: CA, Backbone, Heavy atoms and RMSA[A] was achieved by the GLY residue number 223 and the values were 15.611 kJ/mol, 15.824 kJ/mol, 15.874 kJ/mol and 7.934 kJ/mol, respectively. The lowest was achieved by the ASN residue number 45 and the values were 1.356 kJ/mol, 1.397 kJ/mol, 1.528 kJ/mol and 1.414 respectively (**Available with authors)** The result obtained in the present study has provided a good picture of molecular dynamics of modeled TFPI.

Model Validation

Model generated was energy minimized. Ramachandran analysis of the model was done via PROCHECK-NMR server **[2].** The model showed good stereochemical property in terms of overall G-factor value of -0.68 indicating that geometry of model corresponds to the probability conformation with 97.4% residue in the core region of Ramachandran plot showing high accuracy of model prediction. The number of residues in allowed and generously allowed region was 76% and 2.6% respectively and none of the residue was present in the disallowed region of the plot **(Figure 1)**. Plot between phi and psi angle for all amino acid residues of our TFPI protein was also obtained via PROCHECK-NMR showing their possible conformational state in Ramachandran map **(Available with authors)**. In order to get a good structure plot was made between Chi-1 and Chi-2 value for all the amino acid residues **(Available with authors).** Circular variance and average G-factor obtained for all the 220 amino acids reveals the accessibility of the protein residues and their favorable conformations. RMS Z-score for anomalous bond length and bond angle as determined by WHAT-IF 1.142 and 1.264 respectively, which is very close to 1.0 suggesting very high model quality.

ProSA was used to check the three dimensional model of TFPI for potential errors. The program displays 2 characteristics of the input structure: its Z-score and a plot of its residue energies. The Z-score of -5.02 indicates the overall model quality of TFPI **(Figure 5)**. Z-score also measures the deviation of total energy of the structure with respect to an energy distribution derived from random conformations. The scores indicate a highly reliable structure and are well within the range of scores typically found for proteins of similar size. The energy plot shows the local model quality by plotting knowledge-based energies as a function of amino acid sequence position.QMEAN analysis was also used to evaluate and validate the model. The QMEAN score of the model obtained was 0.69 and the Z-score was -0.88 which is very close to the value 0 and shows the good quality of the model because the estimated reliability of the model was expected to be in between 0 and 1 and this can be inferred from the density plot for QMEAN scores of the reference set **(Figure 6)**. Comparision with non-redundant set of PDB structures in the plot between normalized QMEAN score and protein size revealed different set of Z-values for differnet parameters such as C-beta interactions, interactions between all atoms, solvation, torsion, SSE agreement and ACC agreement which can be clearly observed **(Figure 7).** Some local error were also obtained for the model of TFPI which was higher somewhere in between the residue from 150 and 170. (**Available with authors).**

Conclusion:

In silico studies in general and molecular modeling with molecular dynamics studies based on simulations have been of great help in understanding the structure, function and mechanism of the action of proteins, particularly the membrane proteins. The present investigation was carried out with major objectives to model the TFPI protein and simulate the modeled TFPI protein, thus obtained to understand the actual mechanism of interaction between TFPI and FVIIa: TF complex and between TFPI and FXa. The present study generated a well-defined structure of TFPI protein and its simulation results indicate the validity of the model. The acidic recognition site was found to be present at Asp19 and Glu 48. The signal peptide region is present from residue number 1-28, region in Kunitz domain1 is present from amino acid number 54-104, in domain 2 from 125-175 and in domain 3 from 217-267. Also,

© 2013 Biomedical Informatics

heparin recognition site was also found which was present from residue number 254-263. The energy trajectory of simulation well supports the simulated complex. The trajectory of time with respect to time, due to RMSDs [A]: CA and Backbone differed from other contributing parameters and needed further computational time for achieving ideal plot of plateau phase. This may be attributed to the hardware with slow computational speed available and lack of time to carry out further simulations.

The model generated was also subjected to structural validation. Structure validation by WHATIF, PROCHECK-NMR, ProSA and QMEAN confirmed the reliability of model. The model showed good stereochemical property in terms of overall G-factor value of -0.68 indicating that geometry of model corresponds to the probability conformation with 97.4% residue in the core region of Ramachandran plot showing high accuracy of model prediction. Z score of -5.02 predicted by ProSA represents the good quality of the model. Our results provide insight in understanding structure of TFPI protein. The results has given a good platform for further investigation into deriving the putative drug binding sites of TFPI-FVIIa:TF-FXa quaternary complex. This will further aid in deriving the suitable pharmacophore for ligand search and designing, which will help designing drugs for myriad of diseases attributed to TFPI. The simulation was TFPI here is of preliminary nature and needs further computational timing and refinement. This will also help in understanding the basic molecular biology of TFPI-FVIIa: TF-FXa interactions.

Author contributions:
The presented work was carried out by PA whilst doing Master's dissertation work under the supervision of MK in MD University. The authors acknowledge the support of Centre for Bioinformatics, MD University for the work.

References:

[1] Petrillo G *et al. Curr Cardiol Rev*. 2010 **6:** 325 [PMID: 22043208]
[2] Novotny WF *et al. Blood*. 1988 **72**: 2020 [PMID: 3143429]
[3] Sandset PM & Abildgaard U, *Haemostasis*. 1991 **21:** 219 [PMID: 1794748]
[4] Girard TJ *et al. Nature*. 1989 **338**: 6215 [PMID: 2927510]
[5] Wun TC *et al. J Biol Chem*. 1988 **263:** 6001 [PMID: 2452157]
[6] Novotny WF *et al. J Biol Chem*. 1989 **264**: 18832 [PMID: 2553722]
[7] Broze GJ Jr, *Blood Coagul Fibrinolysis*. 1995 **1**: S7 [PMID: 7647225]
[8] Broze GJ Jr *et al. Biochemistry*. 1990 **29:** 7539 [PMID: 2271516]
[9] Bajaj MS *et al. Thromb Haemost*. 2001 **86:** 959 [PMID: 11686353]
[10] Zhang E *et al. J Mol Biol*. 1999 **285**: 2089 [PMID: 9925787]
[11] Burgering MJ *et al. Mol Biol*. 1997 **269:** 395 [PMID: 9199408]
[12] Carlsson K *et al. Eur J Biochem*. 2003 **270:** 2576 [PMID: 12787023]
[13] Winckers K *et al. Blood Rev.* 2013 **27**:119 [PMID: 23631910]
[14] Krieger E *et al. Proteins*. 2004 **57**: 678 [PMID: 15390263]
[15] Vriend G, *J Mol Graph*. 1990 **8: 52** [PMID: 2268628]
[16] Benkert P *et al. Proteins*. 2008 **71**: 261 [PMID: 17932912]
[17] Benkert P *et al. Nucleic Acids Res*. 2009 **37:** W510 [PMID: 19429685]
[18] Novotny WF *et al. Blood*. 1988 **72**: 2020 [PMID: 3143429]
[19] Wiederstein M & Sippl MJ, *Nucleic Acids Res*. 2007 **35**: W407 [PMID: 17517781]

Edited by P Kangueane
Citation: Agrawal *et al.* Bioinformation 9(16): 808-812 (2013)
License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited