Journal of Computational Methods in Molecular Design, 2017, 7 (1):1-6



ISSN : 2231- 3176 CODEN (USA): JCMMD

In Silico Analysis and 3D Modelling of Endoglin (ENG) Protein in Diabetic Retinopathy

Vidhya VG1* and Anusha Bhaskar²

¹Department of Biotechnology, Faculty of Science and Humanities, SRM University, Kattankulathur, Chennai, Tamilnadu, India ²Centre for Research and Development, PRIST University, Vallam, Thanjavur, Tamilnadu, India

Corresponding Email: vidhyavg@gmail.com

ABSTRACT

Diabetic retinopathy (DR), a microvascular disorder affects majorly the retina of the eye. It is the major cause of permanent blindness in type II diabetic patients. Many genes have been identified for its pathogenesis. Endoglin (ENG) is one of the candidate gene involved in the DR. The physicochemical properties and structure of the ENG protein was studied using computational tools. ENG protein belongs to the Zona-pellucida superfamily. The structure of the protein revealed that the presence of maximum number of random coils as its secondary structure elements. The 3D structure was modelled using Swiss model workspace and the same was validated. This study gives an outlook of endoglin protein to carry out further research on preventing the pathogenesis of disease.

Keywords: Diabetic retinopathy, ENG, Zona-pellucida, Homology modelling

INTRODUCTION

In both developed and developing countries, DR is the major socio economic problem affecting 80 % of patients over 20 years duration of uncontrolled type II diabetes. In India, it affects majorly individuals of age 20 to 64 [1]. Though remarkable advances exist in the diagnosis and treatment, DR still remains the major leading cause of blindness.

Many biochemical pathways have been proposed to explain the pathogenesis of DR. These include polyol pathway, advanced glycation end product (AGE) formation, Protein kinase C (PKC) pathway and hexosamine pathway. All these pathways in turn cause high levels of oxidative stress and upregulation of certain cellular factors (IGF (insulin growth factor), VEGF (vascular endothelial growth factor), TNF (tumor necrosis factor), bFGF-2 (basic fibroblast growth factor)) that eventually contribute to the progression of DR [2,3].

DR progresses slowly from NPDR (non-proliferative DR) to PDR (proliferative DR), characterized by the growth of new blood vessels (retinal neovascularization). Macula edema, characterized by retinal thickening from leaky blood vessels can develop during both stages of retinopathy. Pregnancy, puberty, blood glucose control, hypertension and cataract surgery can accelerate these changes [4].

The number of people worldwide at risk of developing vision loss from diabetes is predicted to double over within next 30 years [5]. Therefore, it is necessary to develop better means to identify, prevent and treat retinopathy in its earliest stages rather than wait for the onset of vision threating lesions.

Endoglin, a transmembrane glycoprotein expressed in vascular endothelial cells is essential for angiogenesis and retinal vascular development. It functions as a co-receptor for TGF- β and interacts with serine/threonine kinase receptors [6,7]. *In vitro* and *in vivo* studies demonstrated that besides affecting endothelial cell function, ENG is reported to regulate and inhibit angiogenesis [8]. Abu El-Asrar et al. have recently demonstrated a significant negative correlation between ENG levels and the levels of sVE-cadherin in the vitreous from patients from PDR [9].

Vidhya, et al.

These findings suggest a lower angiogenic activity in patients with higher levels of ENG and that the upregulation of ENG in the vitreous fluid from patients with PDR may be protective anti-angiogenesis eye response to suppress the progression of PDR.

The nucleotide sequence of the mRNA covers 7461 bp and the gene structure of ENG is divided into 14 exons and 22 introns. The gene was localized to chromosome 9 (9q33-q34.1) by in situ hybridization.

MATERIALS AND MEHODS

The endoglin protein sequence (accession no: AAC63386.1) was retrieved from NCBI

Primary structure prediction

Expasy protparm server was used for physicochemical characterization of the query protein ENG. These included calculating the pI (isoelectric point), molecular weight, total number of positive and negative residues, extinction coefficient [10], instability index [11], aliphatic index [12] and grand average hydropathy (GRAVY) [13].

Secondary structure prediction

SOPMA (self-optimized prediction method with alignment) tool was used for calculating secondary structure features of the protein sequence considered under this study. This method calculates the content of alpha helix, beta sheets, coils, strands and turns [14].

Protein motif scan

The motif regions of ENG protein was identified using FINGERPRINT tool.

Multiple sequence analysis

The MUSCLE tool based on Poisson method was used for performing multiple sequence alignment [15] for 10 sequences EHH62423.1 (*Macca fascicularis*). JAV38373.1 (*Castor canadensis*), NP_001010968.1 (*Rattus norvegicus*), CAA54917.1 (*Mus musculus*), NP_999196.1 (*Sus scrofa*), NP_035708.2 (*Mus musculus*), NP_001182613.1 (*Homo sapiens*), EHH24226.1 (*Macaca mulatta*) and NP_001126409.1 (*Pongo abelii*) along with ENG (AAC63386.1) whose blast hit was above 60%. End gap cost was not considered.

Phylogenetic analysis

Based on the alignment conserved sites were identified and tree was constructed using UPGMA method [16]. Bootstrapping analysis was performed, generating 500 replicates of the aligned sequences.

Tertiary structure prediction

The three dimensional structure of the ENG protein were generated using Swiss-model, an automated homology modeling server [17]. The templates were chosen based on sequence similarity (above 40 %) available in the Swiss model server itself.

Structure visualization and validation

The modeled 3d structure of ENG was visualized using Rasmol, molecular graphics visualization program and was validated using PSVS (structure validation suite) tool [18].

RESULTS

The similarity search of the sequence was carried out with the help of BLAST tool. The results indicated 100% similarity to endoglin isoform 2 precursor (*Homo sapiens*).

Primary structure prediction

The results of physicochemical characterization of ENG protein is shown in Table 1. ENG protein has high aliphatic index value and pI value less than 7. Therefore, ENG protein is a stable acidic protein and has better interaction with water (due to very low GRAVY value).

Secondary structure prediction

The secondary structure of the query protein was predicted using SOPMA. From the results it is seen that random coil conformation is predominant followed by strand and alpha helix orientation. The least observed secondary structure conformation was beta turn (Table 2).

J. Comput. Methods Mol. Des., 2017, 7 (1):1-6

	Mol. Wt.	No. of amino acids	Max. % of amino acid present	Least % of amino acid present	Number of positively charged amino acids (Asp+Glu)	Number of positively charged amino acids (Arg+Lys)	Isoelectric point (pI)	Aliphatic index	Instability index	GRAVY
Endoglin (ENG)	70578.07	658	Leucine (13.1 %)	Tryptophan (1.1%) and Tyrosine (1.4%)	54	46	6.14	95.29	50.29	0.076

Table 1: Physicochemical characterization of Endoglin protein

Table 2: Secondary structure of Endoglin protein

Secondary structure	SOPMA
Alpha helix	20.36 %
Extended strand	26.14 %
Beta turn	5.47 %
Random coil	48.02 %

Motif search

The FINGERPRINT scan of ENG protein showed 10 fingerprints having 2 motifs for each (Table 3).

Table 3: FINGERPRINT scan result of Endoglin protein

FINGERPRINT	No. of motifs
LIPOCALINIMR	2
ANNEXIN	2
INTEGRINB	2
BRADYKINNB1R	2
GPR153	2
SALSPVBPROT	2
NMDARECEPTOR	2
FMRFAMIDEPEP	2
JANUSKINASE3	2
LIGNINASE	2

Multiple sequence alignment and Phylogenetic tree construction

ENG protein from various organisms was chosen for multiple sequence alignment using MUSCLE tool. After complete alignment three conserved sites (100 % toggled) was identified (Figure 1 and Figure 2) and phylogenetic tree was inferred using UPGMA method with 500 bootstrap replicated to depict the evolutionary relationship between different taxons (Figure 3).



Figure 1: Multiple sequence alignment using MUSCLE tool

🐝 M6: Alignment Explorer (eng.mec	a)					
Data Edit Search Alignment	Web Sequencer	Display Help				
🗅 🥔 🖬 📽 🗮 🍓 🎆	W 😏 📜	*8.				n 🛙
Protein Sequences						
Species/Abbrv	Group Name	** * * * * *****	* * * * * * * *	* * * *	* * * * * * * * * *	* * * * * * *
l. Homo sapiens		GITFGAPL <mark>IGALLT</mark> A	<mark>ALWYI</mark> YSHIIRSPS	KRE PWVAVAA PASS	ESSSTN <mark>HSIGSTO</mark>	<mark> STPCS</mark> TSSMA
2. macaca fascicularis ENG		GITEGADL <mark>IGALLT</mark> A	<mark>ALWYI</mark> YSHIIRSPG	KRE PWVAVAA PASS	ESSSTN <mark>HSIGSTO</mark>	<mark> STPCS</mark> TSSMASPGKREPVVAVA#
3. Castor canadensis ENG		GITEGADL <mark>IGALLT</mark> A	<mark>ALUYI</mark> Y SHI RSPS	KRE PWVAVAA PASS	ESSSTN <mark>HSIGSTQ</mark>	ISTPCSTSSMA
4. Rattus norvegicus ENG		GITEGADL <mark>IGALLE</mark> A	<mark>ALWYI</mark> Y SHT RAPS	KRE POVAVAA PASS	ESSSTN <mark>HSIGSTQ</mark>	<mark> STPCS</mark> TSSMA
5. Mus musculus ENG		GITEGADL <mark>IGALLT</mark> A	<mark>ALWYI</mark> YSHTRGPS	KRE PWVAVAA PASS	B SSSTN <mark>HSIGSTO</mark>	STPCSTSSMA
6. Sus scrofa ENG		GITEGADL <mark>IGALLT</mark> A	<mark>ALWYI</mark> HSHTRHPG	KRE PWVAVAA PASS	D S <mark>S S</mark> T N <mark>H S I G S T O</mark>	<mark>) STPCS</mark> TSSMA
7. bTGF Mus musculus		GIAFAADV <mark>IGALLT</mark> IG	<mark>ALWYI</mark> Y <mark>SHT</mark> GETA	RRQQVPISPAS	DNSSAA <mark>HSIGSTO</mark>	STPCSSSTA
8. TGF homo		GIAFAADV <mark>IGALLT</mark> IG	<mark>ALWYI</mark> YSHIGETA	GRQQVPISPAS	DNSSAA <mark>HSIGSTO</mark>	STPCSSSTA
9. Macaca mulatta ENG		GITEGALLT.A	<mark>ALWYI</mark> Y SHT RSPG	REPWVAVAABASS	DSSSTN <mark>HSIGSTO</mark>	<mark>) STPCS</mark> T <mark>SS</mark> M <mark>8</mark>
10. Pongo abelii ENG		GITFGADL <mark>IGALLT</mark> ,A	<mark>ALWYI</mark> Y <mark>SHT</mark> RSPS	KREPWVAVAAPASS	ESSSTN <mark>HSIGSTO</mark>	<mark> </mark>

Figure 2: Three conserved sites (highlighted in yellow) identified in MUSCLE alignment



Figure 3: Phylogenetic tree inferred using UPGMA method with 500 bootstrap replicates

Tertiary structure

The tertiary structure of ENG protein was predicted using Swiss model workspace using the available PDBsum templates (Figure 4). The modeled structure showed 116 H bonds, 1 helices, 22 strands and 25 turns. The modelled structure was validated using PSVS tool and Ramachandran plot was plotted (Figure 5).

DISCUSSION

The physicochemical properties of endoglin protein were studied using protparm server. Aminoacid composition, molecular weight, extinction coefficient, theoretical isoelectric point (pI), grand average of hydropathicity (GRAVY),

aliphatic index (AI) and instability index (II) are the various parameters calculated using Protparm server. The total number of aminoacids in endoglin protein is 658. Based on the average isotopic masses of individual amino acids in the protein and the average isotopic mass of one water molecule, the protparm tool calculates the MW of the protein. The intrinsic property of protein is based on the molar extinction coefficient of three aminoacids tyrosine, tryptophan and cystine at a given wavelength. Based on this protparm computes the extinction coefficient of native protein in water using the following equation.



Figure 4: Three dimensional structure of ENDOGLIN protein modeled using Swiss model server



E (Prot)=No (Tyr) \times Ext (Tyr)+No (Trp) \times Ext (Trp)+No (Cys) \times Ext (Cys)

At 280 nm, the extinction coefficient of endoglin protein is 53035 M⁻¹cm⁻¹ and absorbance was 0.751 assuming all pairs of cys residues from cysteine. The charge of the protein affects its behavior in purification process especially in 2DE method. The isoelectric point (calculated using pKa of aminoacids) is defined as the pH at which a protein has no net charge. The pI of endoglin protein is below 7 which illustrates that most of the amino acids are acidic in nature. The stability of protein in test tube is based on II value. The computed results indicate that endoglin protein is unstable as the II value is above 40. However, the endoglin protein shows very high AI value which indicates its

high thermal stability at wide temperature. This might be due to high content of three amino acids alanine, valine and leucine. The GRAVY value defines the interaction of protein with water and it is calculated by adding the hydropathy values of all the amino acids and dividing it by the number of residues in the sequence. Endoglin protein has very low GRAVY value which deciphers that there is better interaction between protein and water. SOPMA tool was used to predict the secondary structure of the protein. It uses homology based method to compute the percentage of alpha helices, beta turns and random coils in protein. Based on SOPMA results, endoglin protein can be classified as coiled protein which expounds that it is involved in regulation of angiogenic receptors. A fingerprint is a group of conserved motifs taken from multiple sequence alignment- together, the motifs from a characteristic signature for the aligned protein family. The motifs themselves are not necessarily contiguous in sequence, but may come together in 3D space to define molecular binding sites or interaction surfaces. Therefore, the results of Fingerprint scan tool shows that endoglin protein can be used to identify the closest matching PRINTS sequence motif fingerprints in zona pellucida superfamily. Multiple sequence alignment is required for additional evaluation of protein families such as homology modelling and construction of phylogenetic tree. MUSCLE tool was used for aligning the sequences and after alignment 100 % conserved sites were toggled and evolutionary tree was inferred using UPGMA method. The optimal tree showed SBL (sum of branch length) 1.95545675. The percentage of replicate trees in which the associated taxa were clustered together in the bootstrap (500 replicates) is shown next to the branches. The tree was drawn to the scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction methods and are in the units of the number of amino acid substitutions per site. The homology modeling of endoglin protein was modeled using Swiss- model server using the available templates. The number of h bonds, strands and turns was predicted using Rasmol visualization tool.

CONCLUSION

In conclusion, endoglin is an important protein in TGF- β receptor system expressed predominantly on proliferating endothelial cells. In contrast to vascular endothelial growth factor, endoglin levels are high during the initial stages of retinopathy and subsequently decreased when the conditions worsen. In regard to this, this present study is a preliminary work, to suggest that monitoring endoglin gene expression in vivo is necessary for understanding its role in diabetic microvascular complications. Thus, endoglin protein could be used as a diagnostic biomarker for monitoring progression of diabetic retinopathy.

REFERENCES

- [1] Balasubramanyam, M., Rema, M. and Premanand, C., Curr Sci, 2002. 83: p. 1506-1511.
- [2] Pober, J.S. and Sessa, W.C., Nat Rev Imm, 2007. 7: p. 803-815.
- [3] Sang, D.N. and D'Amore, P.A., *Diabet*, 2008. 51: p. 1570-1573.
- [4] Fong, D.S., et al., *Diab Care*, 2003. 26: p. 226-229.
- [5] Wild, S., et al., Diab Care, 2004. 27: p. 1047-1053.
- [6] Barbara, N.P., Wrana, J.L. and Letarte, M., J Biol Chem, 1999. 274: p. 584-594.
- [7] Cheifetz, S., et al., J Biol Chem, 1992. 267: p. 19027-19030.
- [8] Abu El-Asrar, A.M., et al., Mediat Inflamm, 2012. P. 697489.
- [9] Castonguay, R., et al., J Biol Chem, 2011. 286: p. 30034-30046.
- [10] Stanley C.G. and Peter H.V., Anal Biochem, 1989. 182: p. 319.
- [11] Guruprasad, K., Reddy, B.V, and Pandit, M.W., Prot Eng, 1990. 4: p. 155-161.
- [12] Ikai, A., J Biochem, 1980. 88: p. 1895-1898.
- [13] Kyte, J. and Doolittle, R.F., J Mol Biol, 1982. 157: p. 105-132.
- [14] Ashokan, K.V., Mundaganur, D.S., and Mundaganur, Y.D., J Res Bioinfo, 2010. 1: p. 001-008.
- [15] Edgar, R.C., BMC Bioinfo, 2004. 19: p. 113.
- [16] Garcia-Vallve, S., Palau, J. and Romeu, A., Mol Bio and Evol, 1999. 16: p. 1125-1134.
- [17] Schwede, T., et al., Nucl Aci Res, 2003. 31: p. 3381-3385.
- [18] Sayle, R.A. and Milner White, E.J., Trends Biochem Sci, 1995. 20: p. 374.