January 2007 *Volume 1 Issue 1*

Trade Science Inc.

BIOCHEMISTRY
An Indian Journal

Regular Paper

BCAIJ, 1(1), 2007 [14-19]

In Silico Characterization of Ehlers-Danlos Syndrome Causing Human Collagen Proteins

Corresponding Author

K.Sivakumar
Department of Chemistry,
Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya (Deemed University),
Enathur, Kanchipuram – 631 561,
Tamilnadu, (INDIA)
Mobile : +91 98423 61378
Email : shivamk25@yahoo.co.in.

Co-Authors

S.Balaji[1], Ganga Radhakrishnan[2]
[1]Department of Chemistry, Sri Chandrasekharendra Saraswathi Viswa MahaVidyalaya (Deemed University), Enathur, Kanchipuram – 631 561, Tamilnadu, (INDIA)
[2]EXCEL and Polymer Science Labs, Central Leather Research Institute, Adyar, Chennai – 600 020, Tamilnadu, (INDIA)

Received: 26th August, 2006
Accepted: 31th Augest, 2006

Web Publication Date : 21th December, 2006

ABSTRACT

Bioinformatic tools and computational methods have been used to analyze, characterize and to provide more detailed definition of some selected Ehlers-Danlos syndrome (EDS) causing human collagen proteins retrieved from NCBI's Entrez Protein database. Defect in collagen causes various syndromes including Ehlers-Danlos syndrome (EDS). Here, EDS causing 11 human collagen proteins retrieved from NCBI's Entrez Protein database are characterized by using computational methods to give detailed definition of EDS causing collagen proteins. Primary structure analysis shows that all the eleven proteins are rich in glycine (23% – 34%) and proline (17% - 23%) amino acids. The computed pI values of EDS associated collagens reveals that the proteins NP_000080.2, AAB59374.1, AAB59383.1, and AAA52041.1 are basic, the protein AAB35615.1 is strongly basic (pI>7), and the proteins NP_000081.1, NP_000079.1, NP_000384.1 and NP_056534.1 are acidic, the protein NP_000084.2 is strongly acidic (pI<7) in character. The protein AAD13937.1 is classified as neutral (pI≅7) and it has no any acidic or basic amino acids. The low aliphatic index (0 – 52) infers that the collagen proteins may become unstable at high temperature. Secondary structure analysis shows that all the 11 collagen proteins are found to be of predominant coil structure content (>79%) and the secondary structure content prediction server (SSCP) classifies all the proteins as irregular secondary structure class. The irregular structure of collagen proteins is due to the rich content of more flexible glycine and hydrophobic proline amino acids. SOSUI server identified one transmembrane region in alpha 2 type V collagen preproprotein (NP_000384.1) and classifies all the other collagens as soluble proteins. The average molecular weight of EDS causing collagen proteins calculated is 105992 Da.

KEYWORDS Manganese(II) and Iron(II) complexes; Macrocyclic; Biological aspects.

Regular Paper

INTRODUCTION

Ehlers-Danlos syndrome [EDS][1-3] is a rare, heterogeneous group of genetic connective tissue disorder characterized by joint hypermobility (loose joints), easy bruising (contusion), skin extensibility (hyperelasticity or laxity of the skin), tissue fragility (weakness of tissues) and a bleeding diathesis[4]. Structural and genetic mutation studies of the skin in the EDS reveal that the syndrome is a disorder of the collagen fibril[5] due to the mutations in the collagen encoding genes[6-13]. According to the Ehlers-Danlos National Foundation[14], one in 5,000 to one in 10,000 people are affected by some form of EDS. Mutations in the collagen genes disrupt the production, processing, or assembly of type I, III or type V collagen proteins and cause Ehlers-Danlos syndrome[15]. There are different types of EDS and these were classified into six major types according to signs and symptoms. In order to provide a more detailed definition of Ehlers-Danlos syndrome causing collagen proteins, here we have characterized 11 EDS causing collagen proteins retrieved from the NCBI's Entrez Protein database[16].

MATERIALS AND METHODS

We retrieved the Ehlers-Danlos syndrome causing human collagen protein sequences from the NCBI's Entrez Protein database. The Entrez Protein database is searched for the key word "Ehlers-Danlos syndrome". The search result yielded 41 protein sequences. In this only 18 sequences were collagen proteins. From this, we have retrieved 11 EDS causing human collagen proteins. The accession number, database and the sequence details are listed in TABLE 1. The aim of the present study is to characterize the various Ehlers-Danlos syndrome causing human collagen proteins by In Silico methods. The amino acid composition of EDS causing human collagen proteins were computed using the tool CLC free WorkBench 1[17]. Percentages of hydrophobic and hydrophilic residues (TABLE 2) were calculated from the primary structure analysis results. The physico-chemical parameters theoretical isoelectric point (pI), molecular weight, total number of positive and negative residues, extinction coefficient[18], half-life[19-22], instability index[23], aliphatic index[24] and grand average hydrophathy[25] were computed using the Expasy's ProtParam prediction server and tabulated in TABLE 3. The computed theoretical isoelectric point (pI) and the total number of acidic and

TABLE 1: Ehlers-Danlos syndrome causing human collagen protein sequences retrieved from NCBI's Entrez Protein database

Accession Number	Database	Length	Sequence description
NP_000081.1	RefSeq	1466	procollagen, type III, alpha 1
NP_000084.2	RefSeq	1838	alpha 1 type V collagen preproprotein
NP_000080.2	RefSeq	1366	alpha 2 type I collagen
NP_000079.1	RefSeq	1464	alpha 1 type I collagen preproprotein
NP_000384.1	RefSeq	1496	alpha 2 type V collagen preproprotein
AAB59374.1	GenBank	1366	pre-pro-alpha-2 type I collagen
AAB59383.1	GeneBank	155	alpha-1 type III collagen
NP_056534.1	RefSeq	1745	collagen, type V, alpha 3 preproprotein
AAB35615.1	GeneBank	89	type III collagen alpha 1(III) chain
AAD13937.1	GeneBank	9	S62925_1 type III collagen
AAA52041.1	GeneBank	69	type III pro-collagen

TABLE 2: Hydrophilic and hydrophobic residues content

Accession number	Percentage of hydrophobic residues	Percentage of hydrophilic residues	Net hydrophobic residues content
NP_000081.1	50	50	-
NP_000084.2	46	54	Low
NP_000080.2	53	47	-
NP_000079.1	50	50	-
NP_000384.1	48	52	Low
AAB59374.1	53	47	-
AAB59383.1	51	49	-
NP_056534.1	49	51	-
AAB35615.1	56	44	-
AAD13937.1	33	67	Very low
AAA52041.1	45	55	Low

Regular Paper

TABLE 3: Parameters computed using EMBOSS Pepstats and Expasy's ProtParam tool

Accession number	M.Wt.	pI	-R	+R	EC	II	AI	GRAVY
NP_000081.1	138555	6.18	129	122	62225	30.18	37.31	-0.797
NP_000084.2	183591	4.96	224	169	98850	33.3	45.14	-0.876
NP_000080.2	129314	9.08	109	122	51840	23.38	47.67	-0.648
NP_000079.1	138911	5.66	141	129	53495	30.6	38.12	-0.785
NP_000384.1	144720	6.33	142	135	51880	26.11	43.05	-0.818
AAB59374.1	129468	9.08	109	122	51840	22.83	46.88	-0.661
AAB59383.1	13644	9.81	8	11	-	30.15	32.9	-0.795
NP_056534.1	172052	6.37	177	169	61140	29.11	52.68	-0.711
AAB35615.1	7862	11.3	4	8	-	22.64	31.01	-0.718
AAD13937.1	796	6.74	0	0	-	45.08	0	-1.1
AAA52041.1	6219	9.69	4	6	-	40.71	5.65	-1.265

M.Wt.-Molecular weight, pI – Isoelectric point, -R – Number of negative residues, +R– Number of positive residues, EC- Extinction coefficient at 280nm, II – Instability Index, AI – Aliphatic Index, GRAVY – Grand Average Hydropathy.

TABLE 5: Percentage of residues forming alpha, beta and coil structure

Accession number	Alpha	Beta	Coil
NP_000081.1	4	7	90
NP_000084.2	8	10	82
NP_000080.2	5	9	86
NP_000079.1	3	9	88
NP_000384.1	6	8	85
AAB59374.1	4	9	86
AAB59383.1	0	3	97
NP_056534.1	9	11	80
AAB35615.1	0	0	100
AAD13937.1	0	0	100
AAA52041.1	0	4	96

TABLE 6: Transmemebrane region identified by SOSUI server

Accession number	TMB region	Type
NP_000384.1	NWAEARPLLILIVLLGQFVSIK	Primary

basic amino acids were compared and shown in TABLE 4. The tools SOPM, SOPMA[26] and the Secondary Structural Content Prediction (SSCP method-I) server[27] were used for the Secondary Structure Prediction, secondary structure class identification and for the computation of percentages of α-helical, β-strand and coiled regions (TABLE 5). The SOSUI[28] server is used for the identification of transmembrane region (TABLE 6) in protein sequences. The predicted transmembrane helices were visualized and analyzed using helical wheel plots (Figure 1) generated by the program Pepwheel[29] included in the EMBOSS 2.7[30] suite.

TABLE 4: Computed theoretical isoelectric point (pI) and number of acidic and basic amino acids

Accession number	pI	No. of basic amino acids	No. of acidic amino acids	Property
NP_000081.1	6.18	122	129	Acidic
NP_000084.2	4.96	169	224	Strongly Acidic
NP_000080.2	9.08	122	109	Basic
NP_000079.1	5.66	129	141	Acidic
NP_000384.1	6.33	135	142	Acidic
AAB59374.1	9.08	122	109	Basic
AAB59383.1	9.81	11	8	Basic
NP_056534.1	6.37	169	177	Acidic
AAB35615.1	11.34	8	4	Strongly Basic
AAD13937.1	6.74	0	0	Neutral
AAA52041.1	9.69	6	4	Basic

RESULTS AND DISCUSSION

The following discussions are based on the results tabulated in TABLE 2,3,4,5 and 6. The results of primary structure analysis suggest that all the 11 EDS causing collagen proteins are rich in glycine (23% to 34%) and proline (17% to 23%) amino acids. The low hydrophobicity of NP_000084.2, NP_000384.1, AAD13937.1 and AAA52041.1 collagen proteins are due to the presence of highly polar residues (33% to 48%) content (TABLE 2). The average molecular weight of EDS causing collagen proteins calculated is 105992 Da. Although the Expasy's ProtParam computes the extinction coefficient for a range of (276, 278, 279, 280 and 282nm) wavelength, 280nm is favoured, because proteins absorb strongly there while other substances commonly in protein solutions do not. The extinction coefficient of EDS causing human collagen proteins at 280nm is ranging from 51840 to 98850 M^{-1} cm^{-1},

Regular Paper

with respect to the concentration of aromatic amino acids. The Expasy's Protparam did not calculated the extinction coefficient value for the proteins AAB59383.1, AAB35615.1 AAD13937.1 and AAA52041.1 due to the absence of Trp, Tyr and Cys residues. The extinction coefficient of NP_000084.2 is comparatively very high due to the high concentration of Tyr amino acid. The computed protein concentration and extinction coefficients help in the quantitative study of protein-protein and protein-ligand interactions in solution. The biocomputed half-life of most of the collagen is 30hrs and the Expasy's Protparam classifies all the collagen proteins as stable on the basis of instability index (II<40) except the proteins AAD13937.1 and AAA52041.1 whose instability index are greater than 40. The aliphatic index (AI) which is defined as the relative volume of a protein occupied by aliphatic side chains (Ala, Val, Ile and Lys) is regarded as a positive factor for the increase of thermal stability of globular proteins, the low aliphatic index (0 – 52) infers that the collagen proteins may become unstable at high temperature. The computed aliphatic index for the protein AAD13937.1 is zero due to the absence of Ala, Val, Ile and Lys amino acids. Grand average hydropathy (GRAVY) index of all the 11 EDS causing collagen proteins is ranging from –0.6 to –1.2, the very low GRAVY index infers its higher hydro solubility, which could result in a better interaction with water (TABLE .3). Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of the protein is zero. At pI proteins are stable and compact. The computed pI values of EDS associated collagens reveals that the proteins NP_000080.2, AAB59374.1, AAB59383.1, and AAA52041.1 are basic, the protein AAB35615.1 is strongly basic (pI>7), and the proteins NP_000081.1, NP_000079.1, NP_000384.1 and NP_056534.1 are acidic, the protein NP_000084.2 is strongly acidic (pI<7) in character. The protein AAD13937.1 is classified as neutral (pI≅7) and it has no any acid or basic amino acids. The number of basic and acidic amino acids in each collagen proteins correlates well with the corresponding pI computed (TABLE 4). The computed isolelectric point (pI) will be useful for developing buffer systems for purification by isoelectric focusing method. The secondary structure predicted with the help of programs SOPM, SOPMA and SSCP show that the 11 collagens are found to be predominant coil (>79%) structure content (TABLE 5). The secondary structure prediction server (SSCP) classifies all the EDS causing collagens as irregular secondary structure class. The irregular structure is due to the rich content of more flexible glycine and hydrophobic proline amino acids. Proline has a special property of creating kinks in polypeptide chains and disrupting ordered secondary structure. SOSUI server identified one transmembrane region in alpha 2 type V collagen preproprotein (NP_000384.1) and classifies all the other collagens as soluble proteins. The transmembrane region and its length are tabulated in TABLE 6. The helix (Figure 1) of collagen NP_000384.1 is found to have more hydrophobic residues (Figure 1) and it is also well documented by the Kyte and Doolittle mean hydrophobicity profile, computed using the BioEdit[31] tool in which all the peaks are above the zero line (Figure 2).

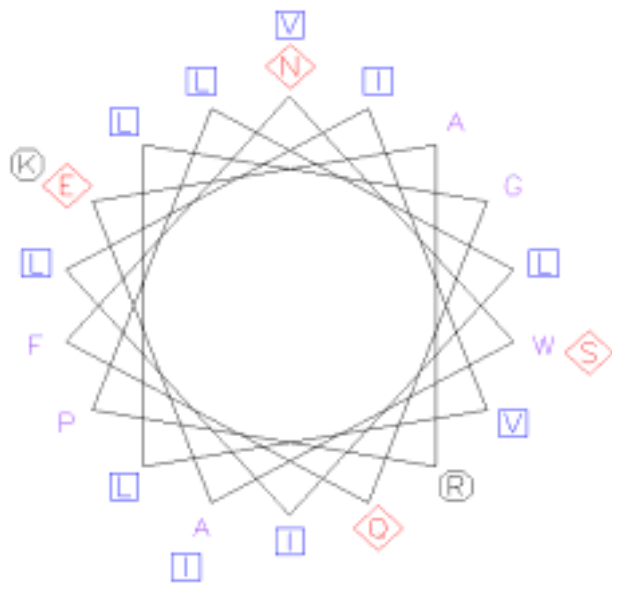

Figure 1: Helical wheel representation of predicted helix of NP_000384.1 collagen protein. Hydrophobic residues are represented as blue squares (V, L, I) and violet letters (F, G, A, P, W), polar residues (E, N, S, Q,) as red diamonds and positively charged residue (K, R) as octagons.

Regular Paper

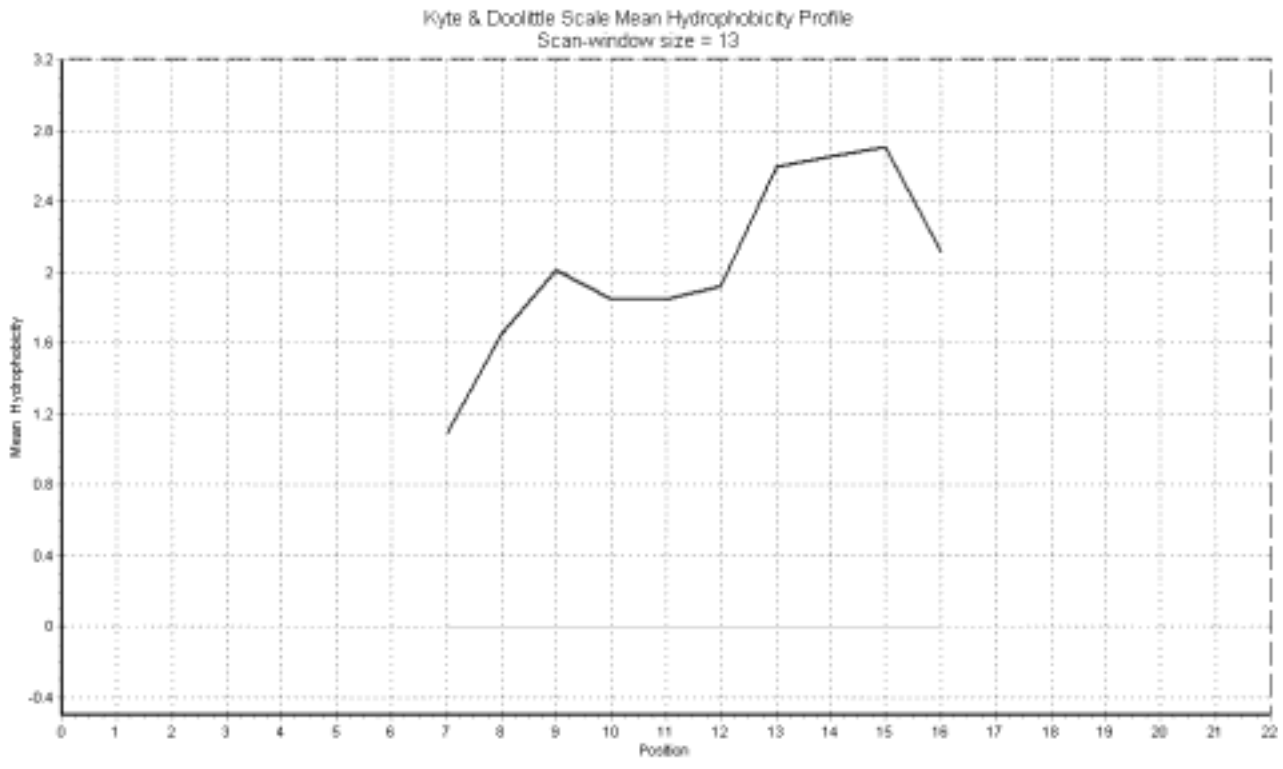

Figure 2: Kyte and Doolittle mean hydrophobicity profile, computed for the transmembrane region (NWAEARPLLILIVLLGQFVSIK) of NP_000384.1 collagen

REFERENCES

[1] N.L.Rudd, K.A.Holbrook, C.Nimrod, P.H.Byers; Lancet, **1**, 50-53 **(1983)**.

[2] P.Beighton, A.De Paepe, D.Danks; Am.J.Med.Genet., **29**, 581-594 **(1988)**.

[3] P.Beighton, A.De Paepe, B.Steinmann, P.Tsipouras, R.J.Wenstrup; Am. J.Med.Genet., **77**, 31-37 **(1998)**.

[4] S.R.Ainsworth, P.L.Aulicino; Clin.Orthop., **286**, 250-256 **(1993)**.

[5] A.Vogel, K.A.Holbrook, B.Steinmann, R.Gitzelmann, P.H.Byers; Lab. Invest., **40**, 201-206 **(1979)**.

[6] L.T.Smith, W.Wertelecki, L.M.Milstone; Am.J.Hum. Genet., **51**, 235-244 **(1992)**.

[7] L.T.Smith, U.Schwarze, J.Goldstein, P.H.Byers; J.Invest.Dermatol., **108**, 241-247 **(1997)**.

[8] P.H.Byers; J.Invest.Dermatol., **103**, Suppl:47S-52S **(1994)**.

[9] H.V.Toriello, T.W.Glover, K.Takahara; Nat.Genet., **13**, 361-365 **(1996)**.

[10] C.Giunta, B.Steinmann; Am.J.Med.Genet., **90**, 72-79 **(2000)**.

[11] A.De Paepe, L.Nuytinck, I.Hausser, I.Anton-Lamprecht, J.M.Naeyaert; Am.J.Hum.Genet., **60**, 547-554 **(1997)**.

[12] N.P.Burrows, A.C.Nicholls, A.J.Richards; Am.J.Hum. Genet., **63**, 390-398 **(1998)**.

[13] R.J.Wenstrup, G.T.Langland, M.C.Willing, V.N. D'Souza, W.G.Cole; Hum.Mol.Genet., **5**, 1733- 1736 **(1996)**.

[14] Ehlers-Danlos National Foundation. http://www.ednf.org, (27/03/2006).

[15] Genetics Home Reference. http:// ghr.nlm. nih.gov/ condition=ehlersdanlos syndrome (27/03/2006).

[16] NCBI Entrez protein database. http://www.ncbi.nlm.nih.gov/entrez/query.fcgidb=Protein. (27/03/2006).

[17] CLC bio., 2005. CLC free Workbench. http://www.clcbio.com/index.phpid=28, (27/10/2005).

[18] S.C.Gill, P.H.Von Hippel; Anal.Biochem., **182**, 319-326 **(1989)**.

[19] A.Bachmair, D.Finley, A.Varshavsky; Science, **234**, 179-186 **(1986)**.

[20] D.K.Gonda, A.Bachmair, I.Wunning, J.W.Tobias, W.S. Lane, A.Varshavsky; J.Biol.Chem., **264**, 16700-16712 **(1989)**.

[21] J.W.Tobias, T.E.Shrader, G.Rocap, A.Varshavsky;

Science , **254**, 1374-1377 **(1991)**.

[22] A.Ciechanover, A.L.Schwartz; Trends Biochem.Sci., **14**, 483-488 **(1989)**.

[23] K.Guruprasad, B.V.B.Reddy, M.W.Pandit; Protein Engineering, **4**, 155-161 **(1990)**.

[24] A.Ikai; J.Biochem., **88**, 1895-1898 **(1980)**.

[25] J.Kyte, R.F.Doolittle; J.Mol.Biol., **157**, 105-132 **(1982)**.

[26] C.Combet, C.Blanchet, C.Geourjon, G.Deleage; TIBS, **25(3) [291]**, 147-150 **(2000)**.

[27] F.Eisenhaber, F.Imperiale, P.Argos, C.Froemmel; Proteins Struct.Funct.Design., **25(N2)**, 157-168 **(1996)**.

[28] Takatsugu Hirokawa, Seah Boon-Chieng, Shigeki Mitaku; Bioinformatics Applications Note, **14(4)**, 378-379 **(1998)**.

[29] G.N.Ramachandran, V.Sasiskharan; Advan.Protein Chem., **23**, 283-437 **(1968)**.

[30] P.Rice, I.Logden, A.Bleasby; Trends Genet., **16**, 276-277 **(2000)**.

[31] T.A.Hall; Nucl.Acids.Symp.Ser., **41**, 95-98 **(1999)**.