Characterization Of A New Collagen-Like Protein From *Trichodesmium erythraeum*

By

RUCHIT GIRISHCHANDRA SHAH

A thesis submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Biomedical Engineering

written under the direction of

Professor Frederick H. Silver

and approved by

_____

_____

_____

New Brunswick, New Jersey

January 2012

ABSTRACT OF THE THESIS

Characterization of a New Collagen-Like Protein from *Trichodesmium erythraeum*

By RUCHIT GIRISHCHANDRA SHAH

Thesis Director:

Professor Frederick H. Silver

In order to meet the exponentially rising demands of collagen in the fields of tissue engineering, cosmetic surgery and drug delivery systems, there is an increasing demand to find novel sources. Recently, an intron-free collagen-like gene found in *Trichodesmium erythraeum*, a colonial marine cyanobacterium, was reported.

This thesis aims at characterizing this new collagen-like sequence and theoretically comparing it with human collagens (especially type I human collagen) and few other bacterial proteins in terms of amino acid composition, flexibility, hydropathicity and charge density. Enzyme degradation and polarized light microscopy were utilized to support the hypothesis of presence of collagen-like protein in *T. erythraeum*'s cell. The initial step of analysis involved running the blastp algorithm using this new sequence as the query sequence and studying its resemblance to various proteins in RefSeq protein database. Comparison with type I human collagen revealed that the

assumed homotrimer molecule of *Trichodesmium erythraeum*'s collagen-like protein is almost as relatively stable as the former. The alternate flexible and rigid domains in the molecular structure can be predicted to coincide and form a highly flexible fibrillar structure that may aid the aggregates to survive stresses from ocean. It is suggested that this collagen-like protein is primarily responsible for transmission of energy during loading and not as much for storage of elastic energy. This new sequence was found to be highly hydrophobic and sparsely charged unlike any of the other sequences used in analyses. Based on this unusual hydrophobic nature, this collagen-like protein may find its applications for procedures like guided tissue reconstruction in dentistry as also in hernia repair and topical wound dressings. In order to determine the potential of the collagen gene of *Trichodesmium erythraeum* for tissue engineering, further studies will be required that include experimentally measuring its thermal stability, enzymatic stability of the triglycine repeat region and the globular N and C termini and its ability to undergo fibrillogenesis using established methods.

ACKNOWLEDGEMENTS

Foremost, I would like to express my heartfelt gratitude to Dr. Frederick H. Silver, my research and thesis advisor. His support and guidance proved to be of invaluable help for me in the completion of this thesis. I would also like to thank Dr. David I. Shreiber and Dr. Joseph W. Freeman for being on the presentation committee.

I extend my sincere thanks to Prof. Shivanthi Anandan and her graduate student Simara Price, Department of Biology at Drexel University for providing us with the *Trichodesmium erythraeum* cells. I am also grateful to Rajesh Patel, Core Imaging Laboratory at Robert Wood Johnson Hospital for his help with microscopy.

I dedicate this thesis to my family, especially my parents, Smita and Girish Shah. My family members have been the pillars of my strength and this degree would not have been possible without their trust, support, prayers, sacrifices and blessings. I will be indebted to them for my life. And above all, I am thankful to God for giving me the ability, inspiration and endurance to pursue my dreams.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1

INTRODUCTION

1.1 Collagen Types and Functions

Collagens are a large family of structural proteins found in the extracellular matrix (ECM) of all vertebrates and invertebrates and account for one-third of their ECM's total protein mass. In its role as a major protein of all connective tissues including bone, skin, tendon and cartilage, collagen not only provides structural support but also influences cell behavior and gene expression via interactions with other matrix proteins and receptors. There are at least 28 different types of collagens that have been identified in vertebrates and reported in the literature [8]. Though each of these types of collagen plays a distinct structural role in the extracellular matrix, all of them have the characteristic triple helical motif as a part of their structure. The most abundant collagens that have continuous gly-X-Y amino acid repeats of ~1000 amino acid residues (types I, II and III) form structural fibrils in ECM. Other types contain smaller continuous regions of gly-X-Y in their triple helices and are collectively termed as non-fibrillar collagens. These non-fibrillar collagens play roles either in association with the fibers (fibril-associated collagens with interrupted triple helices - FACITs) or in networks (e.g.: types IV and VIII) or in linking structures (e.g. type VII).

1.2 Collagen Structure

The triple helix motif is the characteristic of all collagens. In this structure, three parallel chains (α chains), each with a left handed polyproline II-like helical conformation

wind together with one residue stagger to form a right-handed, rope-like super-coiled triple helix [5]. This trimeric collagen molecule results in the collagen's characteristic $(gly\text{-}X\text{-}Y)_n$ repeats. Glycine is found as every third residue because it has the only amino acid residue group that is small enough to fit in the centre of the triple helical structure. In this triplet, proline generally occupies the X and Y positions. Proline in some positions is altered by post-translational modification into hydroxyproline to increase the stability of the triple helix [22]. Other amino acids may also occupy the X and Y positions. This offers exceptional potential for lateral interactions and also helps in maintaining the stability. Some collagen molecules like collagen II are formed by homotrimers (i.e. by three same type of α chains) while others like collagen I are formed by heterotrimers (i.e. by more than one type of α chains).

Lateral interaction between homologous regions within the triple helical domains is the basis for fibril formation [29]. For fibrillar collagens, the ~ 300nm long triple helical molecules are staggered by ~ 67nm, a length known as D period, to form a quarter-staggered structure known as a collagen microfibril. The distance D, a characteristic fingerprint of a fibrous collagen, is made up of a hole region of about 0.6D and an overlap region of about 0.4D [25]. Lateral interactions of microfibrils result into formation of fibrils and eventually continue to form fibers. Non-fibrillar collagens do not form quarter-staggered structures themselves. Figure 1 shows the biosynthetic route followed from collagen genes to collagen fibers.

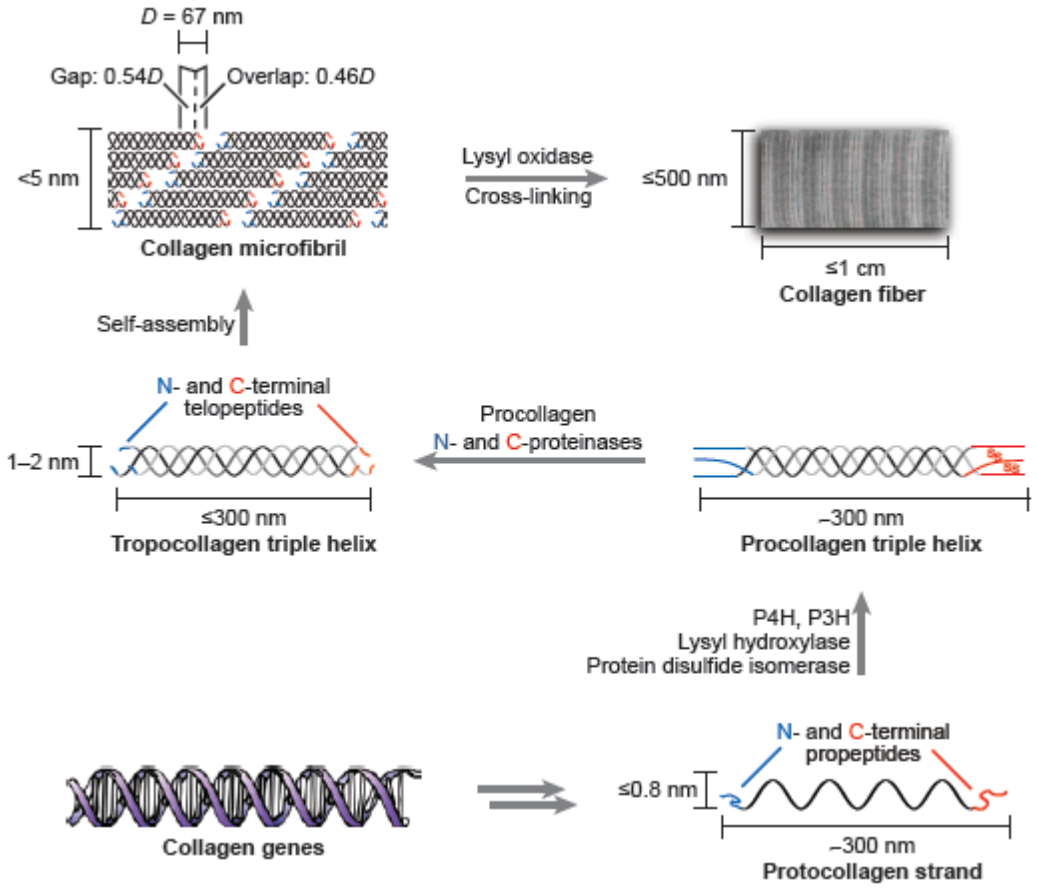

Figure 1: Biosynthetic route to collagen fibers (Hulmes et al., 1981), which are the major component of skin. Size and complexity are increased by posttranslational modifications and self-assembly. Oxidation of lysine side chains leads to the spontaneous formation of hydroxylysyl pyridinoline and lysyl pyridinoline cross-links. [Reproduced from Shoulders and Raines, 2009]

1.3 Collagen Properties

Collagen fibrils show a characteristic banded structure on being examined by transmission electron microscopy at high magnification. Collagen molecules on being stained with heavy metal ions show a series of light and dark bands across the axis of the fibril when viewed under an electron microscope [25]. Phosphotungstic acid (PTA) and uranyl acetate are commonly used as negative stains for electron microscopy of collagen.

The staining pattern of a collagen molecule (Segment Long Spacing – SLS pattern) is directly related to its charge distribution [3]. The lighter bands of the SLS banding pattern have been correlated to the regions of clusters of high hydrophobic amino acids [13]. Figure 2 shows the diagram of a procollagen molecule of Type I collagen. Picrosirius red-stained collagen is known to exhibit birefringence on being examined by polarized light microscopy [12].

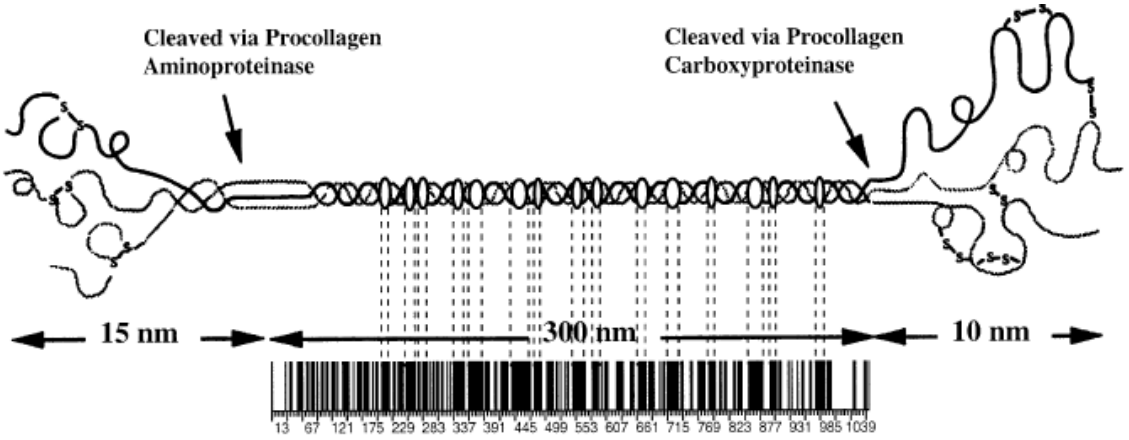

Figure 2: Type I procollagen molecule. The procollagen molecule is shown at the top and consists of aminopropeptides (left-hand portion of molecule at top), an amino non-helical end (straight portion), a triple helical region, a carboxylic non-helical end and a carboxylic propeptide (righthand side end of molecule). The amino (N-) and carboxylic (C-) propeptides are cleaved by specific proteinases during collagen self-assembly. The circles in the triple helix represent major sequences devoid of proline and hydroxyproline that are the likely sites of folds where flexibility is introduced into a normally rigid helix. The striated pattern shown below the helical portion of the molecule is a diagrammatic representation of the flexible (dark bands) and rigid regions (light bands) found in the triple helix. Note the ends of the triple helix are rigid while sequences towards the center of the molecule are more flexible. [Reproduced from Silver et *al.*, 2003]

Of the various functions of collagen, elastic energy storage is a key to the normal functioning of skin, tendon, cartilage and vessel walls. The flexibility of the collagen

molecule is implicated in this storage of elastic energy. Alternate flexible and rigid domains along the collagen molecule have been proposed [26].

1.4 Collagen as a Biomaterial

Collagen is non-toxic, can be resorbed into the body, promotes cell attachment and interaction within cells, directs cell orientation and produces minimal immune response even when transplanted into hosts from different species. Besides, collagen can be processed into various formats like sheets, gels etc., can be chemically crosslinked and also be combined with other biological or synthetic materials to improve its mechanical properties or to tailor cell behavior to specific applications of interest [9]. All these properties make collagen the most widely used biomaterial.

Collagen is a commercial medical product that can be a part of a medical device or it can be fabricated as a reconstituted product from different sources and be used for various biomedical applications like tissue engineering and cosmetic surgery. Usage of collagen can be seen across various medical fields like wound dressings in wound management, tissue augmentation in dermal applications, bioprosthetic heart valves in cardiology, collagen based sutures in surgery and even in treatment of urinary incontinence [21]. Collagen is typically extracted from animal sources especially bovine for most of these applications. However, the risk of transmissible diseases and the inability to modify the biological properties of these sequences is a major disadvantage. Stability and low yields have restricted the practice of producing recombinant human collagens from other systems like use of human cell lines or yeast.

1.5 Eukaryotic Collagen vs. Prokaryotic Collagen

It is generally well agreed that collagen serves the role of aggregating cells allowing larger multicellular organisms to maintain spatial organization of tens or hundreds of trillions of cells, however, the role that collagen plays in the mechanics of prokaryotes is somewhat underreported. Prolyl hydroxylase is considered to be present only in multicellular organisms, although a recent report indicates its presence in the yeast *Hansenula polymorpha* [7]. Bacteria lack this enzyme and cannot post-translationally modify Pro in the X position of the (gly-X-Y)*n* sequence to form Hyp. Despite the absence of Hyp, several expressed bacterial proteins with collagen-like sequences have been shown to form stable triple-helix structures [4, 8, 31]. For example, two collagen genes were identified within the genome of Streptococcal bacteria that were reported to be able to form stable triple helices [31].

1.6 Trichodesmium Erythraeum

Trichodesmium erythraeum (*T. erythraeum*), a colonial cyanobacterium, gets its name for its appearance: Greek words “trichoma” meaning hair and “desmus” meaning bonded. *Trichodesmium* e*rythraeum* produces more nitrogen than any other macroscopic cyanobacteria and is responsible for fixing majority of the N_2 of the biosphere [6]. It is known for its blooms of aggregations spanning tens of thousands of kilometers during periods of low wind stress and warm temperatures [6]. These blooms have been imaged from space as well. On aging the cells become positively buoyant and rise to the surface [1]. A regular array of gas vacuoles in the form of a hollow cylinder in each cell is expected to help the cell float. On reaching a critical length, these cells fracture and allow

new growth to occur at the new set of free ends; the cells then enter an exponential growth phase [2]. A collagen-like gene was found in *T. erythraeum* genome by Orcutt et al. (2002), that has triglycine repeat region that is approximately 10% longer than that found in most vertebrates [15].

1.7 Research Objective

A new collagen-like sequence was recently reported in the genome of *Trichodesmium erythraeum*. This thesis principally aims at two aspects: first, attempting to verify the presence of collagen in *T. erythraeum* cell using enzyme degradation and polarized light microscopy and secondly, theoretical comparison of *T. erythraeum*'s collagen-like sequence with other collagen sequences, mainly human collagen sequences.

T. erythraeum cell samples will be treated with collagenase, trypsin and pepsin individually and their effects will be noted by comparing their cellular structures to that of the control samples. Polarized light microscopy will be used to observe any birefringence exhibited by the cell samples. Local alignment of *Trichodesmium erythraeum*'s collagen-like sequence with all the sequences of NCBI's Reference Sequence (RefSeq) proteins database will be performed to find the closest matches from different species along with various human collagens. These sequences will then be compared in terms of flexibility, hydropathicity, charge density, amino acid composition and proline content using Matlab programming. The results of this research along with future work that analyzes the thermal stability and attempts fibrillogenesis of *Trichodesmium erythraeum*'s collagen-like sequence should be used to determine its potential in the field of tissue engineering.

CHAPTER 2

MATERIALS AND METHODS

2.1 Transmission Electron Microscopy and Polarized Light Microscopy

Fixed *Trichodesmium erythraeum* cells were obtained from the laboratory of Dr. Shivanthi Anandan, Department of Biology at Drexel University. The cells were fixed when they were at the plateau of their life cycle. The cell samples were divided into four sections, each containing two aliquots. One section was assumed to be the control while the others were treated with collagenase[1], trypsin[2] and pepsin[3] individually. 4.29 mg of collagenase powder was mixed with 1 ml of 1x pH 7.4 of Phosphate Buffered-Saline (PBS) solution to make 1ml of collagenase solution. 0.03 mg of trypsin powder was mixed into 1ml of permanence buffer solution to make 1 ml of trypsin solution. 1.2 mg of pepsin was added to 1 ml of pH 2.5 HCl to prepare pepsin solution. Two runs were peformed for each enzyme degradation test to test the consistency of results. For each run, 1ml of cell sample and 1ml of enzyme solution were mixed and pipetted onto a glass slide. All the slides were placed in a 37^{o}C incubator for an hour after which they were washed with deionized water and air dried.

For polarized light microscopy, the enzyme degraded slides were stained using picrosirius red. 0.5 g of Sirius red F3B and saturated aqueous solution of picric acid were used to prepare the picrosirius red stain. The slides were stained for one hour with picrosirius red stain that was previously filtered through a 0.22μm syringe filter. Samples were then washed with deionized water and were allowed to dry at room temperature

[1] Collagenase from Clostridium histolyticum, Type I, Sigma-Aldrich
[2] Trypsin from porcine pancreas, Type II-S, Sigma-Aldrich
[3] Pepsin from porcine gastric mucosa, Sigma-Aldrich

after pouring off most of the excess water from the slides. These slides were then observed under Leitz Laborlux 12 POL light microscope equipped with a ProgRes® CF digital microscope camera. With the aid of ProgRes® Image Capture software, images were captured at various magnifications.

2.2 Sequences: Source, Alignment and Composition

All the sequences used in this thesis were obtained from the NCBI protein database. The reference Fasta formatted text files of all the sequences are tabulated in Appendix 1. Firstly, protein blast (Basic Local Alignment Search Tool) was employed. *Trichodesmium erythraeum*'s collagen-like sequence was entered as the query sequence and RefSeq protein database was selected to find its hundred closest matches. This tool aligns and compares all the protein sequences in the chosen database with the query sequence and gives out a list of most similar sequences. Based on the blastp results, some of the most matched sequences from different species were selected to be used for further analysis. Also, since the thesis aims at comparing the *Trichodesmium erythraeum* collagen-like sequence to human collagens, various human collagen sequences were also downloaded from NCBI database. A total of 23 protein sequences including the *Trichodesmium erythraeum* sequence were selected. Only the gly-X-Y repeats or the triple helical domains of all the sequences were used for further analysis.

The *Trichodesmium erythraeum* collagen-like sequence was locally aligned with other above mentioned selected sequences using the Smith-Waterman algorithm as implemented in the Matlab R2011a bioinformatics toolbox. The resulting alignment scores were then recorded. The function uses BLOSUM50 as its default substitution

scoring matrix. The Smith-Waterman algorithm compares all possible length segments of sequences to be aligned and optimizes the alignment measure.

The amino acid composition of each sequence was determined using a Matlab function. Normalization of each sequence's composition was performed to obtain amino acid compositions per 1000 residues. The number of pro present in X and Y positions of gly-X-Y tripeptides of each chain was also calculated.

2.3 Flexibility Profile Computation

The flexibility profiles of collagen alpha chains and molecules were computed by using the flexibility indices for dipeptides that were obtained by the method previously described in Silver *et al*, 1987. The method used stereochemical plots of rotational freedom of dipeptide units that compose the alpha chains that eventually compose the collagen molecule. Stereochemical calculations involved rotating a dipeptide unit through all possible conformations of the dihedral angles that connect the units. For each set of dihedral angles, if the distance between non-bonded atoms was greater than the minimum allowed inter-atomic contact distances the conformation was allowed [23]. The area of the dipeptide Pro–Lys was chosen as the reference area. The flexibility index of a dipeptide was then determined by dividing the area of the allowed conformations on the conformational map by the reference area. The flexibility index of Pro–Lys was arbitrarily selected as 1 [23]. The flexibility indices assigned for different dipeptide combinations can be seen in Table 1. Assignment of indices for other sets of amino acids was based on their structural similarity to pairs listed in Table 1.

For this thesis, flexibility indices of various dipeptide forming amino acid combinations were entered into a Matlab program that took an amino acid sequence as input and gave a vector of flexibility indices as output. These output flexibility indices were assigned based on the dipeptide units forming the input sequence. The flexibility indices were computed for each of the selected alpha chain sequences and the mean flexibility of each chain was also calculated. For human type I collagen molecule, the α2-2α1 conformation was adopted with each α- chain staggered with respect to other by one residue. Flexibility was computed for each of these three chains and then an average was taken across the three staggered chains to obtain the flexibility profile of type I human collagen molecule. Other than using three identical α- chains to form a homotrimer, the same approach was used for obtaining the flexibility profile of *Trichodesmium erythraeum* collagen molecule.

Dipeptide	Flexibility Index	Dipeptide	Flexibility Index	Dipeptide	Flexibility Index
pro- pro	8	asp- gly	41	glu- glu	13
ala- ala	45	gly-ser	49	glu- ser	12
ala- gly	57	pro- lys	1	arg- gly	20
ala- pro	28	pro- hyp	8	gly- gly	81
arg- hyp	24	ala- asp	40	lys- pro	20
glu- asp	12	ala- hyp	29	hyp- arg	4
pro- glu	2	arg- arg	20	gly- arg	30

Table 1: Typical flexibility indices for dipeptides calculated from the area under conformational plots [Reproduced from Silver *et al*, 1987]

Since an average of the flexibility indices across the whole molecule will not portray the actual spatial distribution of the flexibility along the molecule, a histogram was created to serve the purpose. For our experiment, the flexibility indices of both, the human type-1 collagen molecule and the assumed homotrimer molecule of *T. erythraeum*, lied between 0 and 60. Thus, the X- axis of the histogram consisted of bins

that represented a range of flexibility indices from 0 to 60. A total of 240 equally spaced bins were used thereby allotting each bin a range of 0.25 units of flexibility. The Y-axis represented the frequency or the number of occurrences of any flexibility index falling within the range of each bin along the whole molecule. Due to the different lengths of both the molecules, the frequencies of both the histograms were normalized to obtain data for 1000 residues each thereby making them comparable.

2.4 Hydropathicity Profile Computations

The hydrophobic or hydrophilic nature of each alpha chain was determined using the hydropathy scale developed by Kyte and Doolittle, 1982. In this scale, each amino acid is assigned a value which is reflective of its hydrophobic or hydrophilic nature. Kyte and Doolittle decided on these values based on the assumption that the best hydropathy values should be based on the correlation of transfer free energies and the actual distribution of side chains [14, 30]. The values assigned to various amino acids can be seen in Table 2. A positive value indicates a hydrophobic amino acid whereas a negative value indicates a hydrophilic amino acid.

Alpha chain sequences were entered into the ‘proteinpropplot’ function as implemented in Matlab 2011a bioinformatics toolbox. This function determines the hydropathy of the amino acid sequence as it advances from the amino terminal to the carboxyl terminal of the sequence. A hydrophobic or hydrophilic index determined by Kyte and Doolittle method was assigned to each amino acid residue of the sequence. Average hydropathy of each chain was then determined based on its length. For type I human collagen molecule, the hydropathies of α1 and α2 chains were individually

computed and then the α2-2α1 conformation, with each chain staggered by one amino acid residue with respect to other, was adopted. Average across these three chains was taken to obtain the hydropathy profile of human type I collagen molecule. The *Trichodesmium erythraeum* homotrimer was formed the hydropathy profile of its molecule was determined in the same manner.

Amino Acid	Hydropathy Index	Amino Acid	Hydropathy Index
Ala	1.8	Leu	3.8
Arg	-4.5	Lys	-3.9
Asn	-3.5	Met	1.9
Asp	-3.5	Phe	2.8
Cys	2.5	Pro	-1.6
Gln	-3.5	Ser	-0.8
Glu	-3.5	Thr	-0.7
Gly	-0.4	Trp	-0.9
His	-3.2	Tyr	-1.3
Ile	4.5	Val	4.2

Table 2: Hydropathy Scale for amino acids [Reproduced from Kyte and Doolittle, 1982]

2.5 Charge Density Profile Computation

The charge density profile was derived using the charge distribution along each chain. The charged amino acid residues of each sequence were assigned specific weights. These charges were assigned by Meek et al., 1979 to correlate periodic banding pattern of stained collagen fibrils seen under electron microscope with the charge distribution deduced from amino acid sequence. The weights assigned were as follows: asp, glu, arg, lys =3; his, hyl =1; while all other amino acids received zero weighting [16]. A Matlab function assigned these weights to the amino acids that made up the input sequence. Average charge density of each chain was also computed. The charge density profiles of type I human collagen and *Trichodesmium erythraeum* collagen molecules were computed using an approach similar to that used for obtaining their flexibility and hydropathicity profiles.

2.6 Collagen Stability Calculator

The stabilities of human type I collagen and *T. erythraeum* collagen-like sequences were calculated using the online interactive collagen stability calculator developed by Persikov, Ramshaw and Brodsky, 2005. This algorithm uses an averaging approach that aids in discovering thermally stable and labile domains along the triple helix [19]. Every gly-X-Y triplet is assigned a stability value which is modified for interactions between triplets. Average relative stability values of tripeptide units are computed by using a five tripeptide window. The window consists of two tripeptide units on each side of the gly-X-Y triplet whose average relative stability is to be computed. The values of three individual chains are averaged to obtain the stability of a heterotrimer sequence.

Since the online collagen stability calculator computes the stability of sequences having uninterrupted gly-X-Y triplets, the stability of *Trichodesmium erythraeum* sequence was computed in two parts, one before and one after the interruption in gly-X-Y triplets. For stability calculations of type I human collagen molecule, α2-2α1 heterotrimer conformation was adopted and an option that allowed to convert pro residues in Y positions to hyp was chosen. For *Trichodesmium erythraeum* molecule, homotrimer conformation was adopted and pro residues in Y positions were not converted to hyp in order to mimic the absence of Prolyl hydroxylase in prokaryotic organisms that converts pro to hyp. The stability values of all triplets forming each chain and molecule were obtained.

CHAPTER 3

RESULTS AND DISCUSSION

3.1 Analysis of Polarized Light Microscopy

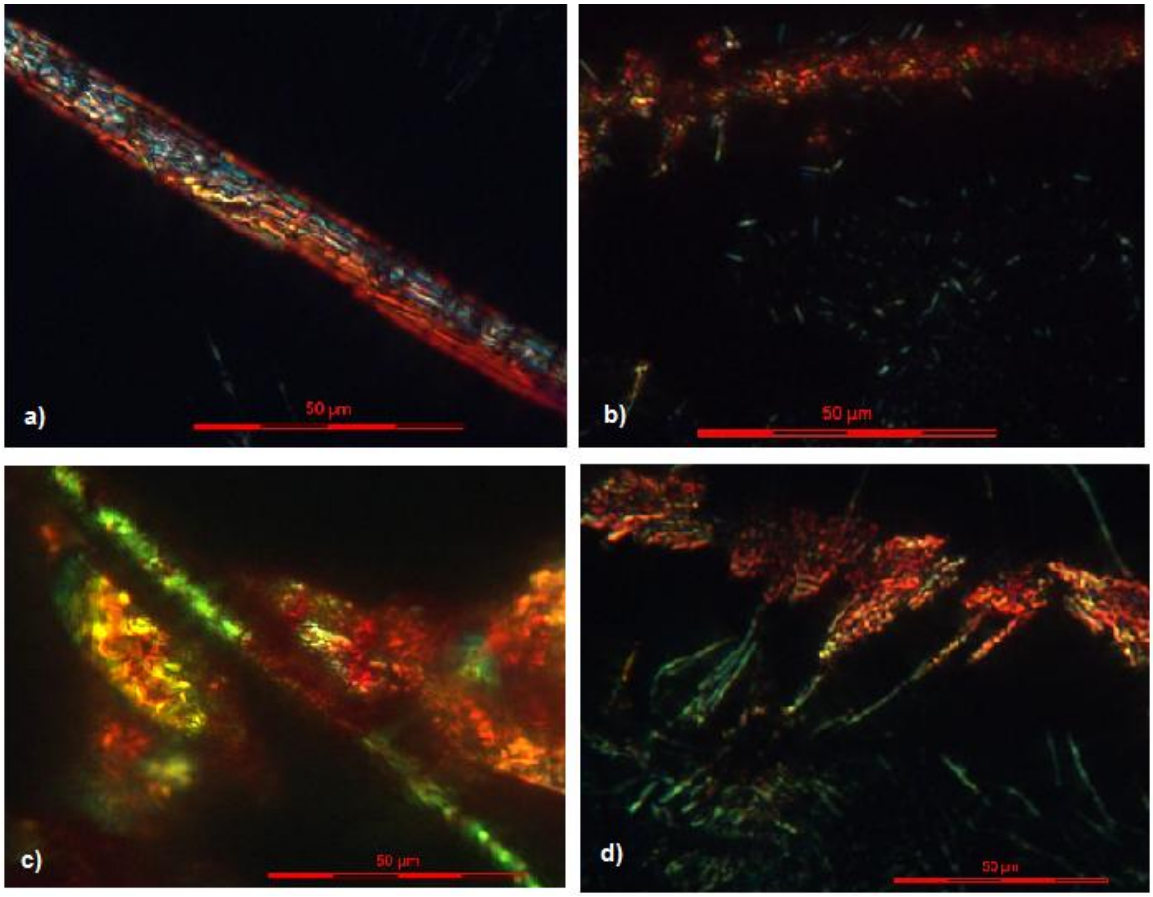

Figure 3: Polarized Light Microscopy images of *T. erythraeum* cell samples. Figure 3a shows a 0.5 mm long and 12µm wide trichome on a control slide under the polarized light microscope. Collagenase degraded the outer cell walls and also digested some materials within the cell (Figure 3b). Figure 3c depicts a pepsin-degraded trichome which looks to have intact inner-cellular structure. Cells forming a trichome are seen to be falling apart due to action of trypsin in Figure 3d. Red bar scale at the bottom of each figure denotes a length of 50µm.

Unstained samples were not clearly visible under polarized light microscope. Thus they were stained with picrosirius red stain. Under the polarized light microscope,

the control slides showed ~0.5 mm long filments formed by aggregation of *T. erythraeum* cells that are about 12µm each in length and width. These filaments, known as trichomes, have width of a single cell and are several cells long. As seen in Figure 3a, the partitions between the cells of a single trichome caused by the outer cell walls were also visible. Figure 3b shows a collagenase-degraded trichome that is about 7 µm in width. The length dimension of the trichome seems to be more or less intact; however, the cell partitions are not visible. Also, some degraded material looks to be falling apart from the cells. A pepsin-degraded trichome is visible in Figure 3c. Pepsin appears to have digested some lateral portions of the trichome; however, the internal cellular structure seems to be intact. As seen in Figure 3d, trypsin seems to have caused a lysis of the trichome structure. Degraded cellular structure appears to be falling apart from the main trichome body. The trichomes on all the slides exhibited birefringence.

On being stained with picrosirius red, collagen appears to give out reddish birefringence under polarized light microscope [12]. All the trichomes on each slide appeared to exhibit reddish birefringence at certain angles. On comparing the collagenase-treated trichome to the control, we observed that the former was laterally thinner as result of the digestion caused by collagenase. Certain cellular structures were also seen to be falling apart from the trichome body; however, the trichome remained intact along its length. Both the above observations insinuate the presence of collagen or a collagen-like protein on the outer and inner longitudinal cell walls. This would explain the collagenase-induced degradation of the cells both on the outside and inside. Trypsin degradation appears to have caused the cells forming a trichome to fall apart thereby indicating the presence of a trypsin-sensitive protein on transverse cell walls that aids in

building cell aggregates to form a trichome. Digestion along the length of the trichome of trichome as caused by pepsin indicates that some pepsin-sensitive protein may also be present on the outer longitudinal walls of cells. Apart form the digestion, further shrinkage of the cells can be attributed to the possible collapse of some of the gas vacuoles that occupy a major portion of *T. erythraeum*'s cell.

3.2 Analysis of Sequence Alignment

SEQUENCE	SMITH-WATERMAN SCORE	SEQUENCE	SMITH-WATERMAN SCORE
T. erythraeum	4172.7	**humancol9a2**	578
humancol1a1	1137.33	**humancol9a3**	603.33
humancol1a2	1088	**humancol10a1**	514
humancol2a1	1096	**humancol15a1**	475.33
humancol4a1	1104	**humancol17a1**	589
humancol4a2	1014.33	**humancol24a1**	913.33
humancol4a3	1188.33	**rattuscol2a1**	1140
humancol4a4	1209	**wssv**	1165.33
humancol4a5	1269	**cereus**	1170.67
humancol4a6	1128.67	***Clostridium b.***	1582.33
humancol7a1	1359.67	***Desulfotomaculum***	1550
humancol9a1	559.33		

Table 3: List of sequences and their Smith-Waterman alignment scores. The scores result from the alignment of these sequences with *T. erythraeum*'s collagen-like sequence. Blosum50 was used as scoring matrix. The first score in the table denotes the score obtained by self-alignment of *T. erythraeum*'s sequence.

The blastp results showed that the *Trichodesmium erythraeum*'s collagen-like sequence aligned most closely to various collagen and collagen-like sequences of diverse species amongst all the protein sequences in NCBI's RefSeq Protein database. For further analyses, protein sequences were chosen based on the magnitude of the scores resulting from blastp and the proximity of the organism to bacterial or human species. Along with

various human collagen sequences, collagen and collagen-like sequences from following species were selected for further comparison: *Clostridium beijerinckii*, shrimp white spot syndrome virus (wssv), *Bacillus cereus* B4264, *Desulfotomaculum acetoxidans* DSM 771 and *Rattus norvegicus*. The scores resulting from application of Smith-Waterman algorithm using Blosum50 scoring matrix are listed in Table 3.

The *Trichodesmium erythraeum* sequence aligned more closely to the bacterial sequences than the human collagen sequences (**Table 3**). Higher scores of type IV, type VII and type I collagens imply that the *Trichodesmium erythraeum* sequence resembles more closely to them than the other human collagen sequences. However, these scores are not too high denoting the fact that the amino acid compositions of these sequences are quite differen*T*. Thus, it can be inferred that even though the *T. erythraeum* sequence is somewhat similar to the human collagen sequences, phylogenetically it is divergent from the human collagen sequences.

3.3 Analysis of Flexibility Profiles

The mean flexibility index of *T. erythraeum*'s collagen-like sequence was found to be close to those of human α1 (I), human α2 (I) and human α1 (II) (**Table 4**). The flexibility indices of other human collagens were lower. The average flexibility of *T. erythraeum* (26.8) was found to be slightly lower than that of the human type I collagen molecule (27.28). A little drop in the flexibility of *T. erythraeum* was found in the region where an interruption in the gly-X-Y triplets occurs (**Figure 4**).

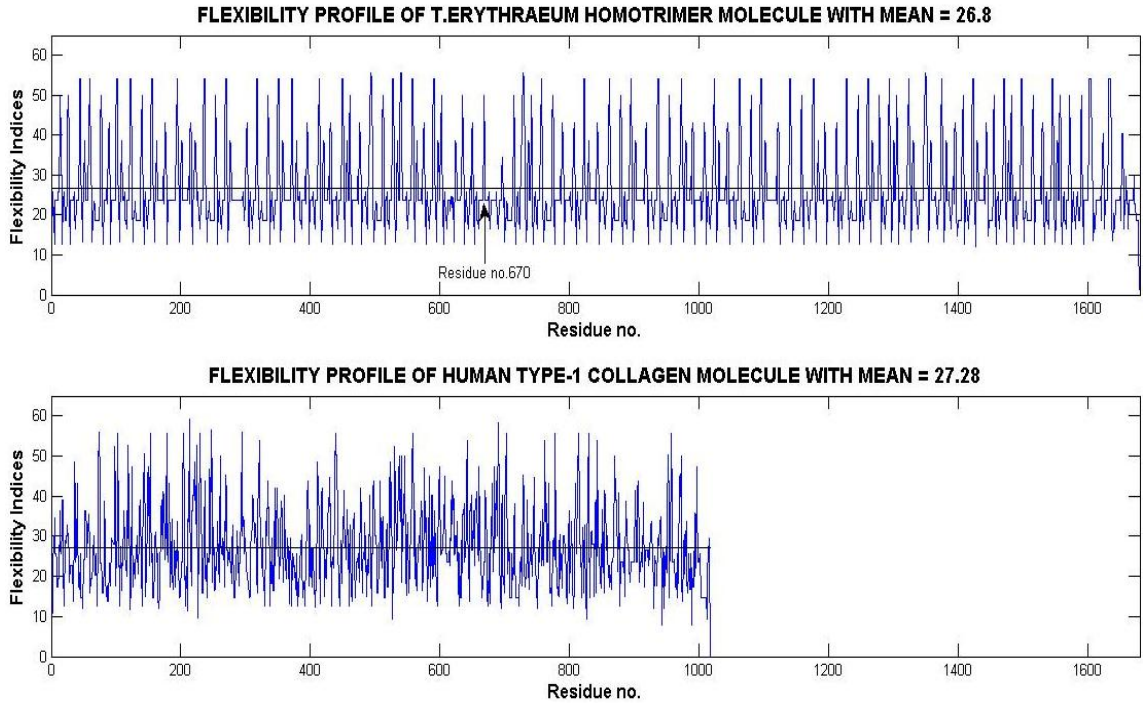

Figure 4: Flexibility profile of *Trichodesmium erythraeum*'s homotrimer molecule with its mean of 26.8 plotted as a black line (Upper). This upper figure also shows an arrow at residue number 670 that marks the interruption in gly-X-Y triplets of *T. erythraeum*'s collagen-like sequence. Flexibility profile of human type I collagen molecule with its mean of 27.28 plotted as a black line (Lower).

Flexible sites in type I human collagen have been reported to coincide with the regions of sequences in the triple helix that were devoid of Gly-Pro-Hyp [11]; whereas the regions of elastic storage were observed to overlap with the locations of the charged amino acid residues of type I collagen sequence [27]. Higher flexibilities of human types II and II collagen as compared to type I collagen has been believed to reflect their higher elastic energy storage during tensile deformation. It has been further suggested that the type I collagen, due to its limited flexibility, may be primarily responsible for transmission of stored energy as opposed to its dissipation [26]. The collagen-like sequence of *T. erythraeum* is rich in hydrophobic amino acids, has a low number of charged amino acid residues (**Table 5**) and consists of high proportions of pro in X and Y

positions of the gly-X-Y triplet. A slightly lower flexibility than type I collagen along with the above factors suggests that this collagen-like protein stores less amount of elastic energy and is principally responsible for energy transmission during loading thereby reducing local loads and stresses.

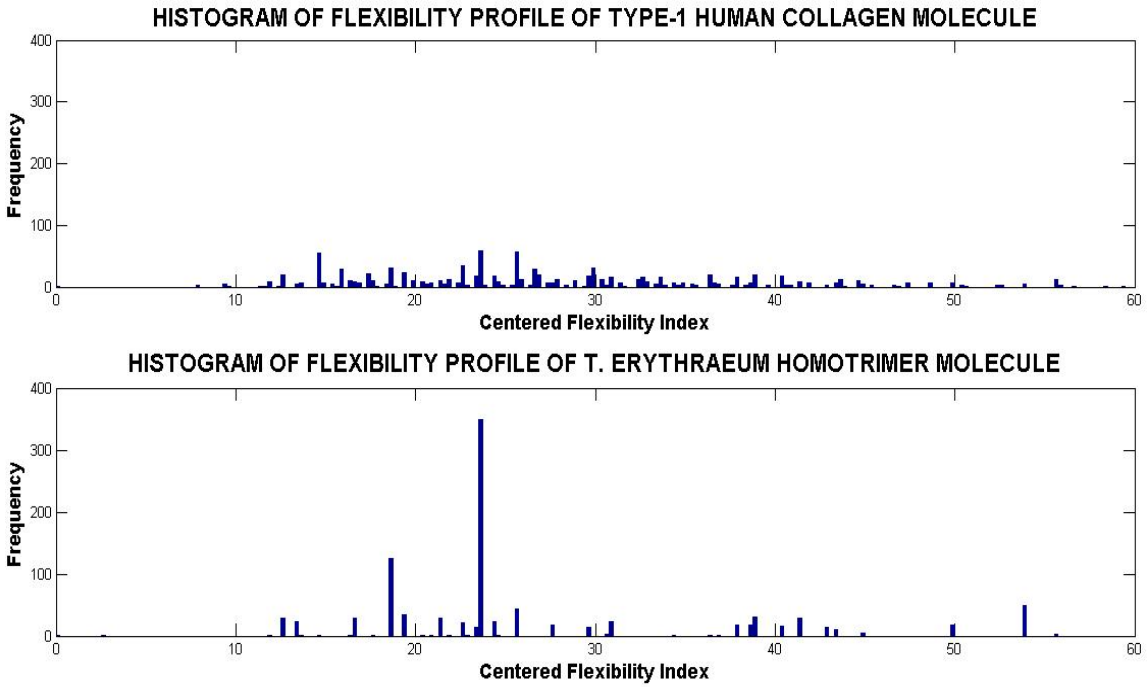

Figure 5: Histograms of type-1 human collagen molecule (upper) and *T. erythraeum's* homotrimer molecule (lower). The numbers on the x-axes represent the centered values of the bins that consist of 0.25 units of flexibility each. The y-axes show the frequency of each bin that has been normalized for 1000 amino acid residues.

The flexibility profile of *T. erythraeum's* homotrimer molecule in **Figure 4** appears to be made up of alternate flexible and rigid bands that are more periodic as compared to the flexibility profile human type 1 collagen molecule that appears to have a more random distribution of flexible and rigid sites. The histograms in **Figure 5** highlight the above fact. The frequency distribution of flexibility of human type 1 collagen has a larger spread as compared to that of the *T. erythraeum* molecule's histogram. Both the histograms have their highest peaks at the bin centered at the flexibility index of 23.625.

For every 1000 amino acids, *T. erythraeum* has 349 occurrences in the flexibility range of 23.5 to 23.75 whereas the human type 1 collagen molecule has 59. The periodic high peaks of about 53.9 flexibility units that are seen in the flexibility profile of *T. erythraeum* in **Figure 4** have 50 occurrences/ 1000 residues.

The fibril-forming capability of *T. erythraeum's* collagen-like molecule is unknown. However, if it can be assumed that it is capable to form fibers, the alternate flexible and rigid regions of individual molecules will coincide due to staggering and lead to a flexible structure. This flexible nature of *T. erythraeum's* collagen fibers will help the trichome aggregates in the ocean to withstand the loads caused by waves, wind currents etc. They can stretch on being loaded by waves and when the waves are gone they can contract back to their original structure and pass on the occupied energy back to the ocean.

3.4 Analysis of Hydropathicity Profiles

Among all the average hydropathicity indices computed by the method proposed by Kyte and Doolittle, only the average hydropathicity index of *T. erythraeum*'s collagen-like sequence was found to be positive; whereas the rest of the sequences were found to have negative average hydropathicity indices. This indicates that the *T. erythraeum*'s collagen-like sequence is highly hydrophobic in nature. This finding is in concurrence with the high number of hydrophobic amino acid residues and low number of hydrophilic amino acid residues present in *T. erythraeum*'s amino acid composition as seen in Table 5. The hydropathicity indices of other bacterial collagen-like sequences,

though lower than human collagen sequences, were also found to be negative i.e. these bacterial sequences are also hydrophilic in nature.

The computation of *T. erythraeum* molecule's hydropathicity profile gave a mean hydropathicity index of 0.35, indicating its hydrophobic nature. The heterotrimer human type I collagen molecule's hydropathicity profile resulted in an average hydropathicity index of -0.8, an indicative of its hydrophilic nature. The interrupted region near residue number 670 of *T. erythraeum*'s homotrimer molecule appeared to be less hydrophobic in comparison to its neighboring segments (**Figure 6**).

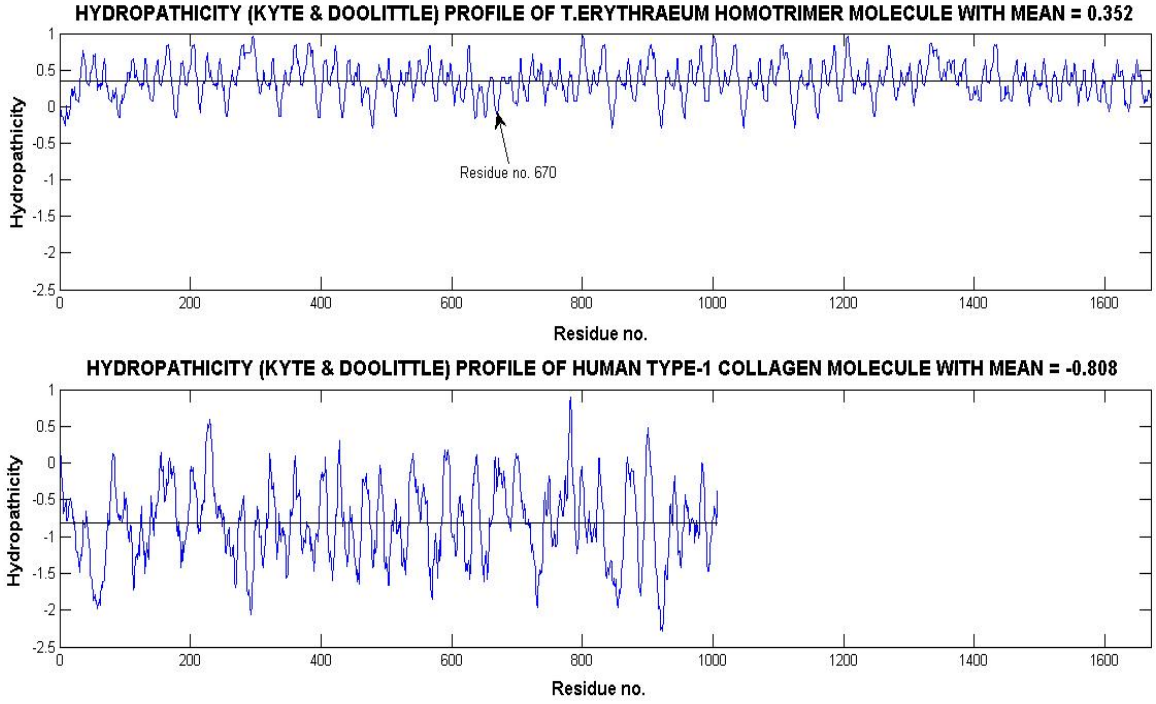

Figure 6: Hydropathicity profiles of *T. erythraeum* homotrimer collagen molecule (Upper) and human type I collagen molecule (Lower) computed by the method proposed by Kyte and Doolittle, 1982. The average hydropathicities of *T. erythraeum* molecule (0.352) and human type I collagen molecule (-0.808) are marked with a black line in their respective figures. The upper figure shows an arrow marked at residue number 670 denoting the break in the gly-X-Y triplets of *T. erythraeum*'s collagen-like sequence.

Hydrophobicity has been reported to be an imperative factor in the packing of molecules [20]. The highly hydrophobic nature of *T. erythraeum*'s collagen-like molecule

may play a significant role in increasing the molecular binding with the aid of hydrophobic interactions. It can also be proposed that this high hydrophobic content of *T. erythraeum*'s collagen-like sequence may help it in attaching to surfaces and form large aggregates that float in ocean water.

3.5 Analysis of Charge Density Profiles

Sequence	Length	Average Flexibility	Average Hydropathicity	Average Charge Density
T. erythraeum	1678	26.83	0.353	0.077
humancol1a1	1014	27.43	-0.876	0.490
humancol1a2	1017	27.14	-0.678	0.451
humancol2a1	1017	27.27	-0.916	0.503
humancol4a1	1268	22.22	-0.723	0.493
humancol4a2	1301	23.10	-0.625	0.534
humancol4a3	1396	21.87	-0.744	0.512
humancol4a4	1395	22.14	-0.779	0.533
humancol4a5	1415	20.55	-0.718	0.418
humancol4a6	1417	22.67	-0.492	0.452
humancol7a1	1530	24.88	-1.027	0.727
humancol9a1	655	23.43	-0.802	0.585
humancol9a2	638	22.63	-0.886	0.549
humancol9a3	634	24.30	-0.802	0.546
humancol10a1	463	22.30	-0.808	0.404
humancol15a1	1388	21.15	-0.375	0.565
humancol17a1	916	21.81	-0.594	0.425
humancol24a1	991	23.72	-0.814	0.570
rattuscol2a1	1081	26.97	-0.698	0.450
wssv	1521	25.31	-0.860	0.731
cereus	1124	28.79	-0.406	0.099
Clostridium b.	2100	24.34	-0.328	0.283
Desulfotomaculum	1265	31.89	-0.440	0.183

Table 4: Averages of the flexibility indices, hydropathicity indices and charge densities of all the sequences along with the lengths of the triple helical domains are stated.

The mean of *T. erythraeum*'s collagen-like sequence's charge density was the lowest among all the computed average charge densities (**Table 4**). The low number of charged amino acid residues in the amino acid composition of *T. erythraeum*, as seen in Table 5, supports its low charge density value. It was observed that the average charge densities of all the bacterial collagen-like sequences were lower than their human counterparts (**Table 4)**. *T. erythraeum* molecule's average charge density of 0.077 is significantly lower as compared to 0.476 of the human type I collagen molecule. A break in the somewhat regular pattern of charge density peaks was observed in the region of interruption of gly-X-Y triplets in *T. erythraeum*'s sequence (**Figure 7**).

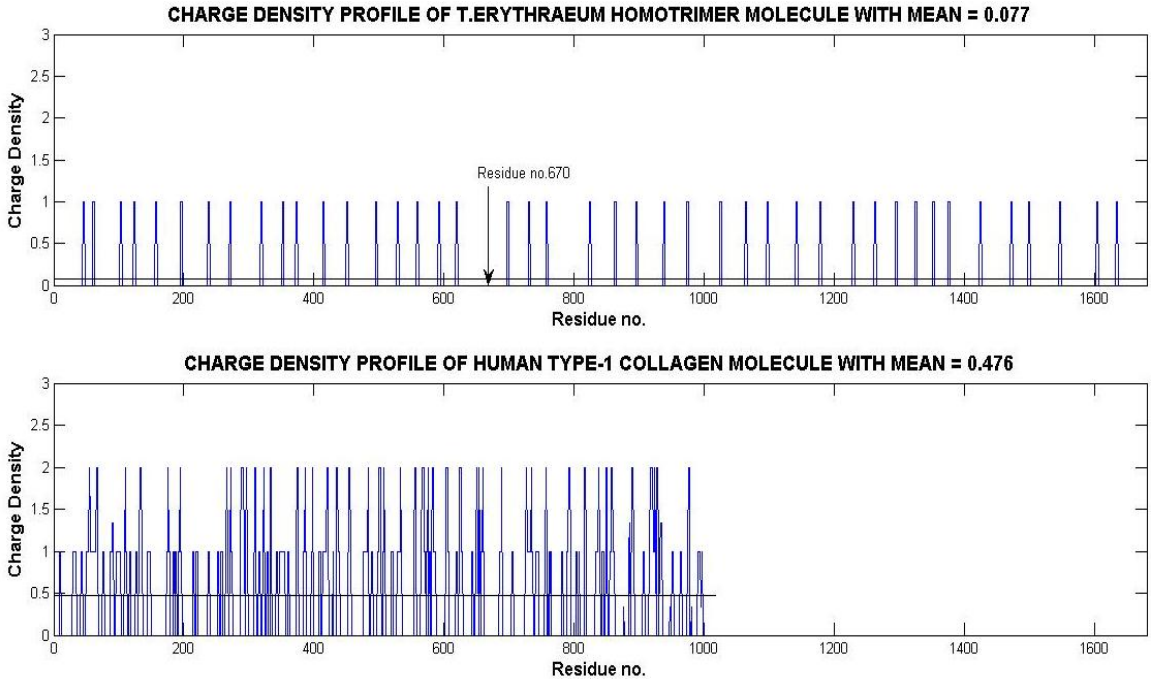

Figure 7: Charge density profile of *T. erythraeum*'s collagen-like molecule with its mean of 0.077 marked as a black line (Upper). Charge density profile of Human type I collagen molecule with its mean of 0.476 marked as a black line (Lower). The upper figure also shows an arrow marked at residue number 670 that represents the interruption in the gly-X-Y repeats of the *T. erythraeum*'s collagen-like sequence.

It was proposed that storage of elastic energy appears to coincide with the positive staining bands within collagen fibril which are formed by the charged amino acids of the collagen sequence [26]. Thus, the low content of charged amino acids in *T. erythraeum*'s collagen-like sequence denotes that this collagen-like protein will store less elastic energy. Also, the molecular structure of this protein will be held together majorly by hydrophobic interactions rather than hydrophilic interactions.

3.6 Analysis of Amino Acid Compositions and Collagen Stability Computations

The amino acid compositions show that gly is the major component of all the sequences. *T. erythraeum* is rich in pro like most of the other collagen sequences (**Table 5**). Apart from this, hydrophobic amino acid residues like ala, ile and val comprise large portions of *T. erythraeum*'s amino acid composition. Lower numbers of charged amino acid residues and higher number of hydrophobic residues, is a distinguishing factor of *T. erythraeum*'s collagen-like sequence that separates it from human collagen sequences.

Another parameter tabulated in Table 5 is the percentage of Pro in X and Y positions of the gly-X-Y triplets of each sequence. The highest percentage of Pro residues in X position (21.9%) was observed in the *T. erythraeum*'s collagen-like sequence among all the sequences and this percentage is equivalent to almost double of that present in any of the human collagens. Like most of the bacterial collagen-like sequences, *T. erythraeum* demonstrates a low percentage of Pro residues in the Y position of gly-X-Y triplet (5.6%). However, this low percentage of Pro in Y positions is still higher as compared to those of other bacterial collagen-like sequences. The total percentage of Pro in X and Y positions combined for *T. erythraeum* is greater than all other sequences.

SEQUENCES	Amino Acid Composition per 1000 residues																				% Pro in X	% Pro in Y
	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val		
T. erythraeum	135	0	0	26	0	0	0	334	0	50	27	0	0	1	275	15	25	0	0	111	21.93	5.60
humancol1a1	116	50	11	31	0	27	46	336	2	6	19	36	7	12	233	34	16	0	0	20	11.83	11.54
humancol1a2	105	53	24	20	0	21	44	338	12	18	32	29	5	10	200	31	19	0	1	38	10.72	9.24
humancol2a1	99	51	12	29	0	37	51	338	2	10	25	36	7	13	222	28	22	0	1	18	11.01	11.31
humancol4a1	26	21	3	39	2	47	41	326	3	34	47	62	14	26	218	34	20	1	6	29	5.99	15.62
humancol4a2	43	44	7	49	3	32	34	311	8	39	57	49	13	34	189	27	27	2	8	23	5.07	12.84
humancol4a3	33	36	14	34	6	27	41	312	9	29	59	57	15	21	214	38	28	1	8	20	6.45	14.47
humancol4a4	33	40	8	45	11	20	34	315	17	24	50	52	12	24	232	34	13	1	7	25	7.81	14.55
humancol4a5	16	17	17	33	3	40	37	322	4	42	64	52	11	20	265	24	17	0	3	13	8.13	18.02
humancol4a6	30	25	14	33	3	33	29	311	7	36	81	61	12	31	185	51	27	1	4	25	5.15	12.07
humancol7a1	49	67	7	54	1	25	68	324	1	12	49	54	5	7	199	31	14	1	0	34	8.04	11.83
humancol9a1	50	56	11	38	8	38	55	305	8	27	53	43	12	9	206	27	18	0	3	31	7.02	12.52
humancol9a2	47	38	9	38	5	50	45	313	13	33	31	58	16	6	210	20	20	0	8	39	8.46	12.38
humancol9a3	57	46	8	43	6	36	47	319	6	22	62	44	8	6	213	32	16	0	3	27	8.68	12.15
humancol10a1	54	30	13	13	0	26	35	337	9	26	37	54	11	9	274	24	19	0	11	19	11.23	16.20
humancol15a1	78	34	27	42	7	29	70	159	18	39	81	37	22	31	138	71	53	7	10	47	3.82	4.97
humancol17a1	44	38	9	35	0	39	39	246	12	27	60	25	22	17	192	106	28	1	25	34	9.72	9.39
humancol24a1	27	43	11	28	1	50	58	335	10	38	59	59	10	13	170	30	23	0	9	25	7.37	9.49
rattuscol2a1	100	54	21	22	0	23	42	337	8	19	33	30	5	10	201	40	20	0	2	33	10.64	9.62
wssv	76	93	20	50	3	35	64	269	3	42	32	36	16	11	118	30	54	1	11	37	7.82	3.09
cereus	72	2	1	11	0	144	20	340	0	76	9	0	4	0	126	23	133	0	0	39	11.39	0.80
Clostridium b.	46	0	1	93	0	0	1	334	0	32	0	0	0	0	56	35	337	0	0	64	5.57	0.00
Desulfotomaculum	163	0	5	56	0	0	1	333	1	2	0	4	0	0	134	7	284	0	0	10	13.44	0.00

Table 5: Amino acid composition of each sequence. The percentages of Pro in X and Y positions of gly-X-Y triplet of each sequence are also noted.

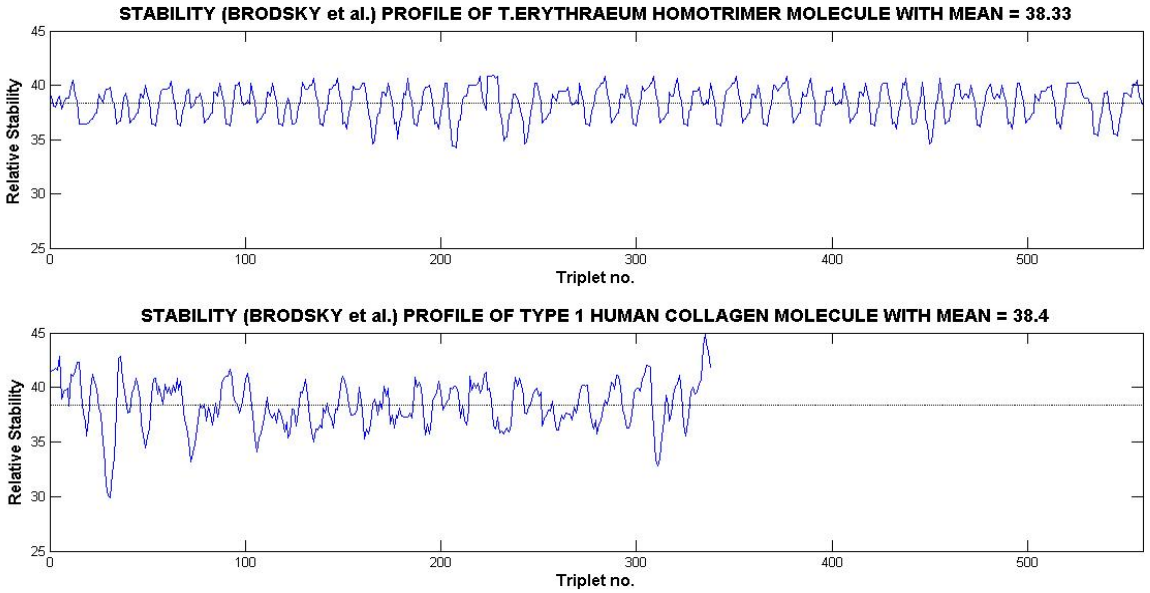

Figure 8: Collagen stability profiles of *T. erythraeum* homotrimer molecule (Upper) and human type I collagen (Lower) computed as per Collagen stability calculator, Brodsky et al. The stability profile of *T. erythraeum* was computed in two parts: before and after interruption in gly-X-Y repeats that is shown by an arrow in upper figure; and was then combined. The averages of the stabilities of *T. erythraeum* (38.33) and human type I collagen (38.4) are plotted as black lines in their respective figures.

The relative thermal stability values obtained from the collagen stability calculator for human type I collagen molecule and *T. erythraeum*'s homotrimer molecule are almost identical. The stability values for *T. erythraeum* molecule were obtained in two parts: before and after the interruption in the gly-X-Y triplets. The averages of stability values of the human type I collagen molecule, *T. erythraeum* molecule before interruption and *T. erythraeum* after interruption were found to be 38.4, 38.23 and 38.39 respectively. The average relative stability of *T. erythraeum* molecule, obtained by combining the two parts before and after the interruption, was found to be 38.33 which is almost equivalent to human type I collagen's average thermal stability. The range of relative stability values obtained for *T. erythraeum* molecule appears to be more compact than that of type I human collagen (**Figure 8**). The minimum relative stability value obtained for *T. erythraeum* was 34.24 in comparison to 29.87 for human type I collagen; while the maximum values obtained for *T. erythraeum* and human type I collagen were 40.96 and 44.95 respectively.

Gly-Pro-Hyp triplets play an important role in the stabilization of collagen. However, in bacteria, the hydroxylation of proline to hydroxyproline does not take place. It has been seen that in the regions of bacterial proteins that lack Hyp and have low imino acid composition, high KGE/D or GQN content leads to electrostatic stabilization [17, 19]. Also, GPP triplets have high melting temperature and these triplets can be used by bacterial collagen for their stability [28]. However, *T. erythraeum* sequence does not possess any KGE/D, GQN or GPP triplets. But, **Figure 8** shows that *T. erythraeum* homotrimer is almost as stable as human type I collagen. Unlike other bacterial collagen-like proteins, *T. erythraeum* sequence consists of a higher percentage of proline in Y

position of gly-X-Y triplets. Thus, the high content of proline in its sequence can be assumed to be the factor responsible for the relative stability of *T. erythraeum*'s collagen-like sequence. Regardless of the presence of high Pro in its amino acid composition, *T. erythraeum*'s collagen-like sequence can be assumed to lack in GPP tripeptides because such triplets may present toxicity problems in bacteria [17].

CHAPTER 4

CONCLUSIONS, APPLICATIONS AND FUTURE WORKS

The blastp results revealed that the collagen-like sequence of *Trichodesmium erythraeum* resembles various collagen sequences among all other protein sequences in the NCBI RefSeq protein database. The local alignments of *T. erythraeum*'s collagen-like sequence with collagen and collagen-like sequences of diverse species showed that this sequence matches more closely to bacterial collagen-like sequences than human collagens. Even though this sequence resembles collagen, the low alignment scores indicate that the composition of *T. erythraeum*'s sequence is different from that of human and other bacterial collagen sequences. The digestion of the cells forming trichomes caused by collagenase and the red birefringence exhibited by certain portions of *T. erythraeum* sample under polarized light microscope support the hypothesis of the presence of a collagen-like protein in *T. erythraeum*'s cells.

The assumed homotrimer molecule of *Trichodesmium erythraeum* was observed to be slightly less flexible as the type I human collagen molecule. Low number of charged amino acids and high gly-pro contents in its amino acid composition forms the basis of the proposal that this collagen-like protein is primarily responsible for transmission of stored energy during loading. The region with a discontinuity in the gly-X-Y triplets of *T. erythraeum*'s sequence coincides with a drop in flexibility and disruptions in somewhat regular patterns of hydropathicity and charge density distributions along the *T. erythraeum* molecule. This irregularity in the sequence is predicted to imply a possible anomaly in the molecular structure of this protein.

The *T. erythraeum* collagen-like sequence was found to be one of the most hydrophobic collagen or collagen-like sequences. This highly hydrophobic nature of collagen-like protein is assumed to aid in molecular binding, surface attachments and in forming large floating aggregates in oceans.

Relative thermal stability calculations demonstrate that *Trichodesmium erythraeum*'s assumed homotrimer molecule is as stable as human type I collagen molecule. Despite the absence of high contents of charged amino acids that are proposed to be responsible for stability of bacterial collagens, the high stability of *Trichodesmium erythraeum* molecule is believed to be a result of high Pro content in its amino acid composition.

In summary, the results of enzyme degradation and polarized light microscopy further asserted the speculation of the presence of a collagen-like protein in *Trichodesmium erythraeum*'s cells. Also, the collagen-like sequence of *T. erythraeum* resembled various collagen or collagen-like sequences. The homotrimer molecule of this sequence was found to be relatively as thermally stable as the human type I collagen. The flexibility histogram of *T. erythraeum* suggests probability of highly flexible fibers that help the aggregates to survive against loads imposed by ocean and winds. However, its highly hydrophobic nature that is assumed to aid it in molecular binding and bloom formation is unlike any vertebrate collagens.

Taking the unusual hydrophobic nature of *T. erythraeum's* collagen-like protein into consideration, some of its potential biomedical applications can be predicted. This protein can be used to make collagen membranes. These membranes can be used in dentistry for surgical procedures like guided tissue regeneration and aid in preventing

epidermal down-growth along the root surfaces during the initial stages of wound healing. These membranes may also be used in repairing abdominal hernias. In the form of capsules, this collagen may be used to contain any movements of the graft materials to maximize the formation of new bone around graft materials in procedures like tooth extraction or maxillary sinus elevation. This hydrophobic collagen may also be used in topical skin dressings effective for dry skin, skin burns and skin wounds as an insoluble collagen support matrix containing releasable soluble collagen. It can also be used as the base material for drug delivery systems.

In order to gauge the potential of this collagen-like gene of *Trichodesmium erythraeum* in tissue engineering, further analyses are required. One of the proposed experiments includes examining the enzymatic stability of the triglycine repeat region and the N and C termini of the protein expressed by this collagen-like gene followed by testing its ability to undergo fibrillogenesis. Another experiment that can be performed is Western blot analysis of *T. erythraeum* samples at various stages of its life cycle. This will help to determine the beginning of expression of collagen as well as the quantity of its expression at different phases of *T. erythraeum*'s life.

REFERENCES

[1] AE, Walsby. "Gas vesicles." *Microbiol Rev.*, 1994: 94–144.

[2] Bell,P.R.F.; Uwins, P.J.R; Elmetri,I.; Phillips, J.A.; Fu,F. & Yago, A.J.E. "Laboratory culture studies of *Trichodesmium* isolated from the Great Barrier Reef Lagoon, Australia." *Hydrobiologia*, 2005: 9-21.

[3] Bender, E.; Silver, F.H.; Hayashi, K. and Trelstad, R.L. "Type I collagen segment long spacing banding patterns. Evidence that the alpha 2 chain is in the reference or A position." *J Biol Chem.*, 1982: 9653-9657.

[4] Boydston, J. A.; Chen, P.; Steichen, C. *T.* and Turnbough, C. L., Jr. "Orientation within the exosporium and structural stability of the collagen-like glycoprotein BclA of Bacillus anthracis." *J. Bacteriol.*, 2005: 5310-5317.

[5] Brodsky, B. and Ramshaw, J.A.M. "The collagen triple-helix structure." *Matrix Biol.*, 1997: 545-554.

[6] Capone, D.G.; Zehr, J.P.; Paerl, H.W.; Bergman, B. and Carpenter, E.J. "*Trichodesmium*, a globally significant marine cyanobacterium." *Science*, 1997: 1221–1229.

[7] de Bruin, E. C.; Werten, M. W.; Laane, C. and de Wolf, F. A. "Endogenous prolyl 4-hydroxylation in Hansenula polymorpha and its use for the production of hydroxylated recombinant gelatin." *FEMS Yeast Res.*, 2002: 291-298.

[8] Gordon,M.K. and Hahn, R.A. "Collagens." *Cell And Tissue Research*, 2010: 247-257.

[9] Han, B.;Huang, L.L.H.; Cheung, D.; Cordoba, F.; Nimni, M. "Polypeptide growth factors with a collagen binding domain: Their potential for tissue repair and organ regeneration." In *Tissue engineering of vascular prosthetic grafts*, by R.G. Landes, 287-299. Austin, 1999.

[10] Han, R.; Zwiefka, A.; Caswell, C. C.,;Xu, Y.; Keene, D. R.; Lukomska, E.;Zhao, Z.; Hook, M. and Lukomski, S. "Assessment of prokaryotic collagen-like sequences derived from streptococcal Scl1 and Scl2 proteins as a source of recombinant GXY polymers." *Appl. Microbiol. Biotechnol.*, 2006: 109–115.

[11] Hofmann, H., Voss, *T.*, Kuhn, K. & Engle, J. "Localization of flexible sites in thread-like molecules from electron micrographs: comparison of interstitial, basement membrane and intima collagens." *J. Mol. Biol.*, 1984: 325-343.

[12] Junqueira, L. C. U.; Bignolas, G. and Brentani, R.R. "Picrosirius staining plus polarization microscopy, a specific method for collagen detection in tissue sections." *Histochemical Journal*, 1979: 447-455.

[13] Kobayashi,K.;Ito, *T.* and Hoshino, *T.* "Correlation between Negative Staining Pattern and Hydrophobic Residues of Collagen." *J. Electron Microscopy (Tokyo)*, 1986: 272-275.

[14] Kyte J. & Doolittle R.F. "A Simple Method for Displaying the Hydropathic Character of a Protein." *J. Mol. Biol.*, 1982: 105-132.

[15] Layton, B.E.; D'Souza, A.;Dmapier, W.; Zeiger, A.; Sabur, A. & Jean-Charles, J. "Collagen's Triglycine Repeat Number and Phylogeny Suggest an Interdomain Transfer Event from a Devonian or Silurian Organism into *Trichodesmium erythraeum*." *J. Mol. Evol.*, 2008: 539-554.

[16] Meek, K.M.; Chapman, J.A. & Hardcastle, A. "The Staining Pattern of Collagen Fibrils." *The Journal of Biological Chemistry*, 1979: 10710-10714.

[17] Mohs, A.;Silva, *T.*;Amin, R.;Lukomski, S.& Brodsky, B. "Mechanism of Stabilization of a Bacterial Collagen Triple Helix in the Absence of Hydroxyproline." *The Journal Of Biological Chemistry* 282, no. 41 (2007): 29757-29765.

[18] Orcutt, K.M.; Rasmussen, U.; Webb, E.A.; Waterbury, J.B.; Gundersen, K. and Bergman, B. "Characterization of *Trichodesmium* spp. by genetic techniques." *Appl. Environ. Microbiol.*, 2002: 2236–2245.

[19] Persikov, A.V.; Ramshaw, J.A.M & Brodsky, B. "Prediction of Collagen Stability from Amino Acid Sequence." *The Journal Of Biological Chemistry* 280, no. 19 (2006): 19343-19349.

[20] Piez, K.A. and Trus, B.L. "Microfibrillar structure and packing of collagen: Hydrophobic interactions." *Journal of Molecular Biology*, 1977: 701-704.

[21] Ramshaw, J.A.M; Vaughan, P. & Werkmeister, J.A. "Applications Of Collagen In Medical Devices." *Biomedical Engineering Applications,Basis & Communications*, 2001: 14-26.

[22] Rosenbloom, J.; Harsch, M. and Jimenez, S. "Hydroxyproline content determines the denaturation temperature of chick tendon collagen." *Arch. Biophys. Biochem.*, 1973: 478-484.

[23] Silver, F. H. *Biological Materials: Structure,Mechanical Properties and Modeling of Soft Tissues.* New York: New York University Press, 1987.

[24] Silver, F. H.; Freeman, J. and Seehra, G.P. "Collagen self-assembly and development of matrix mechanical properties." *J. Biomech.*, 2003: 1529–1553.

[25] Silver, F.H. and Landis, W.J. "Viscoelasticity, Energy Storage and Transmission and Dissipation by Extracellular Matrices in Vertebrates." In *Collagen: Structure and Mechanics*, by P. Fratzl, 133-154. Springer Science+Business Media, LLC, 2008.

[26] Silver, F.H.; Foran, D. & Horvath, I. "Mechanical Implications of the Domain Structure of Fiber-Forming Collagens: Comparison of the Molecular and Fibrillar Flexibilities of the a1-Chains Found in Types I–III Collagen." *J. theor. Biol.* 216 (2002): 243-254.

[27] Silver, F.H.; Seehra, P.; Freeman, J.W. and DeVore D. "Viscoelastic properties of young and old human dermis: Evidence that elastic energy storage occurs in the flexible regions of collagen and elastin." *J. Appl. Polym. Sci.*, 2002: 1978–1985.

[28] Sutoh, K. and Noda, H. "Conformational change of the triple-helical structure. IV. Kinetics of the helix-folding of (Pro-Pro-Gly)n (n = 10, 12, and 15)." *Biopolymers*, 1974: 2477-2488.

[29] Van Der Rest, M. & Garrone, R. "Collagen Family of Proteins." *The FASEB Journal*, 1991: 2814-2823.

[30] Wolfenden, R.V.; Cullis, P.M. & Southgate, C.C.F. "Water, protein folding and the genetic code." *Science*, 1979: 575-577.

[31] Xu, Y.; Keene, D.; Bujnicki, J.; Hook, M. & Lukomski, S. "Streptococcal Scl1 and Scl2 Proteins Form Collagen-like Triple Helices." *The Journal Of Biological Chemistry* 277, no. 30 (2002): 27312-27318.

[32] Thiagarajan, G.; Li, Y.; Mohs, A.; Strafaci, C.; Popiel, M.; Baum, J. and Brodsky, B. "Common interruptions in the repeating tripeptide sequence of non-fibrillar collagens: Sequence analysis and structural studies on triple-helix peptide models." *J Mol Biol.*, 2008: 736–748.

[33] Shoulders, M.D. and Raines,R.*T.* " Collagen Structure and Stability." Annu. Rev. Biochem. 2009.78:929-958.

[34] Hulmes, D.J.S.; Jesior, J-C.; Miller, A.; Berthet-Colominas, C. and Wolff, C. "Electron microscopy shows periodic structure in collagen fibril cross sections." *Proc. Natl. Acad. Sci. USA*, 1981. 78:3567–71.

APPENDIX 1

Collagen chains and National Center for Biotechnology Information (NCBI) reference numbers

Sequence	NCBI Reference No.
T. erythraeum	YP_720336.1
humancol1a1	NP_000079.2
humancol1a2	NP_000080.2
humancol2a1	NP_001835
humancol4a1	NP_001836
humancol4a2	NP_001837
humancol4a3	NP_000082
humancol4a4	NP_000083
humancol4a5	NP_000486
humancol4a6	NP_001838
humancol7a1	NP_000085
humancol9a1	NP_001842
humancol9a2	NP_001843
humancol9a3	NP_001844
humancol10a1	NP_000484
humancol15a1	NP_001846
humancol17a1	NP_000485
humancol24a1	NP_690850
rattuscol2a1	NP_445808.1
wssv	AAK77699.1
cereus	YP_002369524.1
Clostridium b.	YP_001309700.1
Desulfotomaculum	YP_003192498.1

APPENDIX 2

Relevant Matlab programs

1.

```
% Program to compute flexibility, hydropathicity, charge density, amino
% acid composition and thermal stability
clc
clear all
close all

%% Reading T.E. and human type 1 collagen sequences
tricho = fastaread('TE.txt');
tery = tricho.Sequence;
h1a1 = fastaread('humancol1a1.txt');
col1a1 = h1a1.Sequence;
h1a2 = fastaread('humancol1a2.txt');
col1a2 = h1a2.Sequence;

% Conversion to 3-letter code sequence for flexibility & charge density
computations
te = char(lower(aminolookup(tery')));
c1a1 = char(lower(aminolookup(col1a1')));
c1a2 = char(lower(aminolookup(col1a2')));

%% Computing Flexibility Profile
% Flexibility profile of single sequences
f_te = flexibility(te);
f_a1 = flexibility(c1a1);
f_a2 = flexibility(c1a2);

% Forming molecule conformation
flex_te = [[f_te;0;0] [0;f_te;0] [0;0;f_te]]; % homotrimer
flex_h1 = [f_a2 [0;f_a1;0;0] [0;0;f_a1;0]];   % heterotrimer

% Average flexibility of three chains to obtain molecule's flexibility
mol_flex_te = round(mean(flex_te,2)*100)/100;
mol_flex_h1 = round(mean(flex_h1,2)*100)/100;

edges = 0.125:.25:59.875;
[nhist_h1 xout_h1] = hist(mol_flex_h1,edges);
[nhist_te xout_te] = hist(mol_flex_te,edges);
nhist_h1 = round((nhist_h1/sum(nhist_h1))*1000);
nhist_te = round((nhist_te/sum(nhist_te))*1000);
figure
ax(1) = subplot(2,1,1);
bar(ax(1),xout_h1,nhist_h1)
xlabel('Centered Flexibility Index')
ylabel('Frequency')
title('HISTOGRAM OF FLEXIBILITY PROFILE OF TYPE-1 HUMAN COLLAGEN
MOLECULE')
ax(2) = subplot(2,1,2);
bar(ax(2),xout_te,nhist_te)
xlabel('Centered Flexibility Index')
```

```
ylabel('Frequency')
title('HISTOGRAM OF FLEXIBILITY PROFILE OF T. ERYTHRAEUM HOMOTRIMER
MOLECULE')
linkaxes([ax(2) ax(1)],'xy')

%% Charge Density Profile
% Charge Density profile of single sequences
c_te = charge(te);
c_a1 = charge(c1a1);
c_a2 = charge(c1a2);

% Forming molecule conformation
ch_te = [[c_te;0;0] [0;c_te;0] [0;0;c_te]];
ch_h1 = [c_a2 [0;c_a1;0;0] [0;0;c_a1;0]];

% Average charge density of three chains to obtain molecule's charge
density
mol_ch_te = mean(ch_te,2);
mol_ch_h1 = mean(ch_h1,2);
% Mean charge densities of molecules
mn_ch_te = mean(mol_ch_te);
mn_ch_h1 = mean(mol_ch_h1);

%% Hydropathicity (Kyte & Doolittle) Profile
% Hydropathicity profile of single sequences
h_te = proteinpropplot(tery,'propertytitle','Hydrophobicity (Kyte &
Doolittle)');
h_a1 = proteinpropplot(h1a1,'propertytitle','Hydrophobicity (Kyte &
Doolittle)');
h_a2 = proteinpropplot(h1a2,'propertytitle','Hydrophobicity (Kyte &
Doolittle)');

% Forming molecule conformation
hyd_te = [[h_te.Data;0;0] [0;h_te.Data;0] [0;0;h_te.Data]];
hyd_h1 = [[h_a2.Data] [0;h_a1.Data;0;0] [0;0;h_a1.Data;0]];

% Average hydropathicity of three chains to obtain molecule's charge
density
mol_hyd_te = mean(hyd_te,2);
mol_hyd_h1 = mean(hyd_h1,2);
% Mean hydropathicities of molecules
mn_hyd_te = mean(mol_hyd_te);
mn_hyd_h1 = mean(mol_hyd_h1);

%% Plotting of data
figure(1)
subplot(2,1,1)
plot(mol_flex_te)
hold on
plot(1:length(mol_flex_te),mn_flex_te,'k')
xlabel('Residue no.')
ylabel('Flexibility Indices')
title(['FLEXIBILITY PROFILE OF T.ERYTHRAEUM HOMOTRIMER MOLECULE WITH
MEAN = ',num2str(mn_flex_te)]);
axis([0 length(mol_flex_te) 0 65])
```

```
subplot(2,1,2)
plot(mol_flex_h1)
hold on
plot(1:length(mol_flex_h1),mn_flex_h1,'k')
xlabel('Residue no.')
ylabel('Flexibility Indices')
title(['FLEXIBILITY PROFILE OF HUMAN TYPE-1 COLLAGEN MOLECULE WITH MEAN
= ',num2str(mn_flex_h1)]);
axis([0 length(mol_flex_te) 0 65])

figure(2)
subplot(2,1,1)
plot(mol_ch_te)
hold on
plot(1:length(mol_ch_te),mn_ch_te,'k')
xlabel('Residue no.')
ylabel('Charge Density')
title(['CHARGE DENSITY PROFILE OF T.ERYTHRAEUM HOMOTRIMER MOLECULE WITH
MEAN = ',num2str(mn_ch_te)]);
axis([0 length(mol_ch_te) 0 3])
subplot(2,1,2)
plot(mol_ch_h1)
hold on
plot(1:length(mol_ch_h1),mn_ch_h1,'k')
xlabel('Residue no.')
ylabel('Charge Density')
title(['CHARGE DENSITY PROFILE OF HUMAN TYPE-1 COLLAGEN MOLECULE WITH
MEAN = ',num2str(mn_ch_h1)]);
axis([0 length(mol_ch_te) 0 3])

figure(3)
subplot(2,1,1)
plot(mol_hyd_te)
hold on
plot(1:length(mol_hyd_te),mn_hyd_te,'k')
xlabel('Residue no.')
ylabel('Hydropathicity')
title(['HYDROPATHICITY (KYTE & DOOLITTLE) PROFILE OF T.ERYTHRAEUM
HOMOTRIMER MOLECULE WITH MEAN = ',num2str(mn_hyd_te)]);
axis([0 length(mol_hyd_te) -2.5 1])
subplot(2,1,2)
plot(mol_hyd_h1)
hold on
plot(1:length(mol_hyd_h1),mn_hyd_h1,'k')
xlabel('Residue no.')
ylabel('Hydropathicity')
title(['HYDROPATHICITY (KYTE & DOOLITTLE) PROFILE OF HUMAN TYPE-1
COLLAGEN MOLECULE WITH MEAN = ',num2str(mn_hyd_h1)]);
axis([0 length(mol_hyd_te) -2.5 1])

% Stability Profile Plotting
load('te_stability.mat');
mnb_st_te = mean(st(1:223));
mn_st_te = mean(st);
l = length(st);
mna_st_te = mean(st(224:l));
```

```
figure(4)
subplot(2,1,1)
plot(no,st)
xlabel('Triplet no.')
ylabel('Stability')
title(['STABILITY (BRODSKY et al.) PROFILE OF T.ERYTHRAEUM HOMOTRIMER
MOLECULE WITH MEAN = ',num2str(mn_st_te)])
hold on
plot(1:l,mn_st_te,'k')
axis([0 l 25 45])
% subplot(3,1,2)
% plot(no,st)
% xlabel('Triplet no.')
% ylabel('Stability')
% title(['STABILITY (BRODSKY et al.) PROFILE OF T.ERYTHRAEUM HOMOTRIMER
MOLECULE AFTER INTERRUPTION WITH MEAN = ',num2str(mna_st_te)])
% hold on
% plot(1:l,mna_st_te,'k')
% axis([224 l 25 45])
clear st triplet no
load('type1_stability.mat')
mn_st_h1 = mean(st);
subplot(2,1,2)
plot(no,st)
xlabel('Triplet no.')
ylabel('Stability')
title(['STABILITY (BRODSKY et al.) PROFILE OF TYPE 1 HUMAN COLLAGEN
MOLECULE WITH MEAN = ',num2str(mn_st_h1)])
hold on
plot(1:l,mn_st_h1,'k')
axis([0 l 25 45])
```

2.

```
% FLEXIBILITY PROFILE OF AMINO ACID SEQUENCE
% This function calculates the flexibility profile of a column
containing
% the amino acid sequence of collagen protein

% the output is stored in variable 'c'. The input 'a' is just a single
% column of the sequence of the amino acids


function c = flexibility(a)
n = length(a);
c = zeros(n,1);
b = [a(2:n,:);'nil'];                          % Staggered amino acid
sequence

for i=1:n
x=strcmp(a(i,:),'ala');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
```

```
else x1=strcmp(b(i,:),'arg');
    if x1==1
        c(i)=30;
    else x1=strcmp(b(i,:),'asn');
        if x1==1
        c(i)=12;
    else x1=strcmp(b(i,:),'asp');
        if x1==1
        c(i)=40;
    else x1=strcmp(b(i,:),'cys');
        if x1==1
        c(i)=12;
    else x1=strcmp(b(i,:),'gln');
        if x1==1
        c(i)=13;
    else x1=strcmp(b(i,:),'glu');
        if x1==1
        c(i)=40;
    else x1=strcmp(b(i,:),'gly');
        if x1==1
        c(i)=41;
    else x1=strcmp(b(i,:),'his');
        if x1==1
        c(i)=20;
    else x1=strcmp(b(i,:),'hyl');
        if x1==1
        c(i)=20;
    else x1=strcmp(b(i,:),'hyp');
        if x1==1
        c(i)=29;
    else x1=strcmp(b(i,:),'ile');
        if x1==1
        c(i)=40;
    else x1=strcmp(b(i,:),'leu');
        if x1==1
        c(i)=40;
    else x1=strcmp(b(i,:),'lys');
        if x1==1
        c(i)=20;
    else x1=strcmp(b(i,:),'met');
        if x1==1
        c(i)=-1;
    else x1=strcmp(b(i,:),'pro');
        if x1==1
        c(i)=28;
    else x1=strcmp(b(i,:),'phe');
        if x1==1
        c(i)=-1;
    else x1=strcmp(b(i,:),'ser');
        if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'thr');
        if x1==1
        c(i)=20;
    else x1=strcmp(b(i,:),'trp');
        if x1==1
        c(i)=30;
```

```
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=44;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=12;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'arg');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=24;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=31;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'met');
            if x1==1
```

```
            c(i)=-1;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=12;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end

else x=strcmp(a(i,:),'asn');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=35;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=35;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=57;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=13;
```

```
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=29;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=35;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=14;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=12;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'asp');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
```

```
    c(i)=13;
else x1=strcmp(b(i,:),'cys');
    if x1==1
    c(i)=12;
else x1=strcmp(b(i,:),'gln');
    if x1==1
    c(i)=15;
else x1=strcmp(b(i,:),'glu');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'gly');
    if x1==1
    c(i)=41;
else x1=strcmp(b(i,:),'his');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'hyl');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'hyp');
    if x1==1
    c(i)=14;
else x1=strcmp(b(i,:),'ile');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'leu');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'lys');
    if x1==1
    c(i)=20;
else x1=strcmp(b(i,:),'met');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'pro');
    if x1==1
    c(i)=20;
else x1=strcmp(b(i,:),'phe');
    if x1==1
    c(i)=20;
else x1=strcmp(b(i,:),'ser');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'thr');
    if x1==1
    c(i)=20;
else x1=strcmp(b(i,:),'trp');
    if x1==1
    c(i)=30;
else x1=strcmp(b(i,:),'tyr');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'val');
    if x1==1
    c(i)=12;
else c(i)=0;
    end;end;end;end;end;end;end
```

```
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'cys');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=12;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=41;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=41;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'ser');
```

```
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=12;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'gln');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=57;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=24;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=13;
```

```
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=10;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=14;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=13;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end

else x=strcmp(a(i,:),'glu');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'glu');
```

```
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=29;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=12;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'gly');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=81;
```

```
else x1=strcmp(b(i,:),'arg');
    if x1==1
        c(i)=30;
    else x1=strcmp(b(i,:),'asn');
        if x1==1
        c(i)=12;
    else x1=strcmp(b(i,:),'asp');
        if x1==1
        c(i)=40;
    else x1=strcmp(b(i,:),'cys');
        if x1==1
        c(i)=49;
    else x1=strcmp(b(i,:),'gln');
        if x1==1
        c(i)=13;
    else x1=strcmp(b(i,:),'glu');
        if x1==1
        c(i)=40;
    else x1=strcmp(b(i,:),'gly');
        if x1==1
        c(i)=81;
    else x1=strcmp(b(i,:),'his');
        if x1==1
        c(i)=20;
    else x1=strcmp(b(i,:),'hyl');
        if x1==1
        c(i)=29;
    else x1=strcmp(b(i,:),'hyp');
        if x1==1
        c(i)=29;
    else x1=strcmp(b(i,:),'ile');
        if x1==1
        c(i)=8;
    else x1=strcmp(b(i,:),'leu');
        if x1==1
        c(i)=30;
    else x1=strcmp(b(i,:),'lys');
        if x1==1
        c(i)=20;
    else x1=strcmp(b(i,:),'met');
        if x1==1
        c(i)=20;
    else x1=strcmp(b(i,:),'pro');
        if x1==1
        c(i)=28;
    else x1=strcmp(b(i,:),'phe');
        if x1==1
        c(i)=20;
    else x1=strcmp(b(i,:),'ser');
        if x1==1
        c(i)=49;
    else x1=strcmp(b(i,:),'thr');
        if x1==1
        c(i)=20;
    else x1=strcmp(b(i,:),'trp');
        if x1==1
        c(i)=30;
```

```
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=29;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=12;
        else c(i)=0;
            end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end

else x=strcmp(a(i,:),'his');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=13;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=40;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=14;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=35;
```

```
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=14;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=-1;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'hyl');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=-1;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'his');
            if x1==1
```

```
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=35;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=-1;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'hyp');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=-1;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=4;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'asp');
```

```
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'cys');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'gln');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'glu');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'gly');
    if x1==1
    c(i)=8;
else x1=strcmp(b(i,:),'his');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'hyl');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'hyp');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'ile');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'leu');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'lys');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'met');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'pro');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'phe');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'ser');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'thr');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'trp');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'tyr');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'val');
    if x1==1
    c(i)=-1;
else c(i)=0;
```

```
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end

else x=strcmp(a(i,:),'ile');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=19;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=15;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
```

```
            c(i)=13;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=12;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'leu');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=35;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=29;
        else x1=strcmp(b(i,:),'ile');
```

```
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=35;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=35;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=35;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end

else x=strcmp(a(i,:),'lys');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
```

```
            c(i)=13;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=24;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=13;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end

else x=strcmp(a(i,:),'met');
if(x==1)
```

```
x1=strcmp(b(i,:),'ala');
if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'arg');
    if x1==1
        c(i)=13;
    else x1=strcmp(b(i,:),'asn');
        if x1==1
        c(i)=-1;
    else x1=strcmp(b(i,:),'asp');
        if x1==1
        c(i)=-1;
    else x1=strcmp(b(i,:),'cys');
        if x1==1
        c(i)=12;
    else x1=strcmp(b(i,:),'gln');
        if x1==1
        c(i)=-1;
    else x1=strcmp(b(i,:),'glu');
        if x1==1
        c(i)=13;
    else x1=strcmp(b(i,:),'gly');
        if x1==1
        c(i)=41;
    else x1=strcmp(b(i,:),'his');
        if x1==1
        c(i)=-1;
    else x1=strcmp(b(i,:),'hyl');
        if x1==1
        c(i)=20;
    else x1=strcmp(b(i,:),'hyp');
        if x1==1
        c(i)=29;
    else x1=strcmp(b(i,:),'ile');
        if x1==1
        c(i)=-1;
    else x1=strcmp(b(i,:),'leu');
        if x1==1
        c(i)=13;
    else x1=strcmp(b(i,:),'lys');
        if x1==1
        c(i)=13;
    else x1=strcmp(b(i,:),'met');
        if x1==1
        c(i)=-1;
    else x1=strcmp(b(i,:),'pro');
        if x1==1
        c(i)=20;
    else x1=strcmp(b(i,:),'phe');
        if x1==1
        c(i)=-1;
    else x1=strcmp(b(i,:),'ser');
        if x1==1
        c(i)=13;
    else x1=strcmp(b(i,:),'thr');
        if x1==1
        c(i)=20;
```

```
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=-1;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end

else x=strcmp(a(i,:),'pro');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=2;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=4;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=2;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=2;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'lys');
```

```
            if x1==1
            c(i)=1;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=2;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=2;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=2;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=2;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end

else x=strcmp(a(i,:),'phe');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
```

```
            c(i)=20;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=24;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=13;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'ser');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'asn');
```

```
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'asp');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'cys');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'gln');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'glu');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'gly');
    if x1==1
    c(i)=41;
else x1=strcmp(b(i,:),'his');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'hyl');
    if x1==1
    c(i)=30;
else x1=strcmp(b(i,:),'hyp');
    if x1==1
    c(i)=29;
else x1=strcmp(b(i,:),'ile');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'leu');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'lys');
    if x1==1
    c(i)=20;
else x1=strcmp(b(i,:),'met');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'pro');
    if x1==1
    c(i)=20;
else x1=strcmp(b(i,:),'phe');
    if x1==1
    c(i)=20;
else x1=strcmp(b(i,:),'ser');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'thr');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'trp');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'tyr');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'val');
```

```
            if x1==1
            c(i)=12;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end

else x=strcmp(a(i,:),'thr');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=45;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=29;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
```

```
            c(i)=20;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=29;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'trp');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=12;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=41;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
```

```
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=8;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=12;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'tyr');
if(x==1)
    x1=strcmp(b(i,:),'ala');
    if x1==1
        c(i)=12;
    else x1=strcmp(b(i,:),'arg');
        if x1==1
            c(i)=30;
        else x1=strcmp(b(i,:),'asn');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'asp');
            if x1==1
            c(i)=29;
        else x1=strcmp(b(i,:),'cys');
            if x1==1
            c(i)=12;
```

```
        else x1=strcmp(b(i,:),'gln');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'glu');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'gly');
            if x1==1
            c(i)=41;
        else x1=strcmp(b(i,:),'his');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'hyl');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'hyp');
            if x1==1
            c(i)=29;
        else x1=strcmp(b(i,:),'ile');
            if x1==1
            c(i)=13;
        else x1=strcmp(b(i,:),'leu');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'lys');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'met');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'pro');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'phe');
            if x1==1
            c(i)=12;
        else x1=strcmp(b(i,:),'ser');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'thr');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=20;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=12;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
else x=strcmp(a(i,:),'val');
```

```
if(x==1)
x1=strcmp(b(i,:),'ala');
if x1==1
c(i)=13;
else x1=strcmp(b(i,:),'arg');
    if x1==1
    c(i)=20;
else x1=strcmp(b(i,:),'asn');
    if x1==1
    c(i)=29;
else x1=strcmp(b(i,:),'asp');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'cys');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'gln');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'glu');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'gly');
    if x1==1
    c(i)=41;
else x1=strcmp(b(i,:),'his');
    if x1==1
    c(i)=-1;
else x1=strcmp(b(i,:),'hyl');
    if x1==1
    c(i)=29;
else x1=strcmp(b(i,:),'hyp');
    if x1==1
    c(i)=24;
else x1=strcmp(b(i,:),'ile');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'leu');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'lys');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'met');
    if x1==1
    c(i)=20;
else x1=strcmp(b(i,:),'pro');
    if x1==1
    c(i)=20;
else x1=strcmp(b(i,:),'phe');
    if x1==1
    c(i)=13;
else x1=strcmp(b(i,:),'ser');
    if x1==1
    c(i)=29;
else x1=strcmp(b(i,:),'thr');
    if x1==1
```

```
            c(i)=13;
        else x1=strcmp(b(i,:),'trp');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'tyr');
            if x1==1
            c(i)=-1;
        else x1=strcmp(b(i,:),'val');
            if x1==1
            c(i)=13;
        else c(i)=0;
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
            end;end;end;end;end;end;end
        end
    end

end;end;end;end;end;end;end
end;end;end;end;end;end;end
end;end;end;end;end;end;end
```

3.

```
%% FUNCTION TO CALCULATE CHARGE DENSITY OF A GIVEN COLLAGEN SEQUENCE
% Syntax
% c = charge(a)
% a is the input amino acid sequence whose charge density profile is
computed into
% variable c

function c = charge(a)
n=size(a,1);
c=zeros(n,1);
for i=1:n
x=strcmp(a(i,:),'asp');
if x==1
    c(i)=3;
end
end
for i=1:n-1
x=strcmp(a(i,:),'arg');
if x==1
    c(i)=3;
end
end
for i=1:n-1
x=strcmp(a(i,:),'his');
if x==1
    c(i)=1;
end
end
for i=1:n-1
x=strcmp(a(i,:),'glu');
if x==1
    c(i)=3;
```

```
end
end
for i=1:n-1
x=strcmp(a(i,:),'hyl');
if x==1
    c(i)=1;
end
end
for i=1:n-1
x=strcmp(a(i,:),'lys');
if x==1
    c(i)=3;
end
end
```

4.

```
% Program to compute Smith Waterman Alignment scores
clc
clear all
close all

files = dir('*.txt');
for i = 1:length(files)
    x = strcmp('TE.txt',files(i).name);
    if x == 1
        reffilename = files(i).name;
        break;
    end
end

files(i) = [];

clear i
ref = fastaread(reffilename);
score = zeros(length(files),1);
p1 = strfind(reffilename(1,:),'.');
reffile = reffilename(1,1:p1-1);

for i = 1:length(files)
    queryfilename = files(i).name;
    query = fastaread(queryfilename);
    p2 = strfind(queryfilename(1,:),'.');
    qfile = queryfilename(1,1:p2-1);
%     subplot(length(files)/2,2,i)
    [score(i) al] =
swalign(ref.Sequence,query.Sequence,'Showscore',true);
    xlabel([upper(num2str(reffilename)),' SEQUENCE'])
    ylabel([upper(num2str(queryfilename)),' SEQUENCE'])
    title(['Scoring Space & Winning Path for
',upper(num2str(reffile)),' vs. ',upper(num2str(qfile)),' with score =
',num2str(score(i))])
    clear queryfilename query al
end
```