Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2016) 24: 2525 – 2538

doi:10.3906/elk-1402-100

Implementation analyses of proteins and genes obtained from cancer patients

Murat DEMİR[1], **Ali KARCI**[2,*]
[1]Vocational School, Muş Alparslan University, Muş, Turkey
[2]Department of Computer Engineering, Faculty of Engineering, İnönü University, Malatya, Turkey

Received: 11.02.2014 • **Accepted/Published Online:** 28.09.2014 • **Final Version:** 15.04.2016

Abstract: Proteins are an important research area for scientists who are interested in bioinformatics and computational molecular biology, since these studies may result in important results in the case of diseases. Due to this, in this study, bioinformatics data were analyzed based on nucleotides and motifs. Bioinformatics data for proteins were obtained from two different databases. The obtained data belonged to cancer patients, and the genes in these DNA and protein sequences, the proteins synthesized by these sequences, and motifs in these data were analyzed. In the analysis, the ABCB1, ALOX5AP, AKT1, BRCA1, BRCA2, TNF, TNFSF13B, TP53, TP63, TP73, and WT1 genes were used. The proteins synthesized by genes belonging to similar classes were analyzed based on amino acid distributions, atomic distributions, Ramachandran plot similarities, and motifs.

Key words: Genes, proteins, cancer genes, motif discovery, Ramachandran diagrams

1. Introduction

Today multidisciplinary study is obligatory because some methods from an area are applied to some problems in another area. Bioinformatics is such an area where methods in computer science are applied to problems in biology. Bioinformatics can thus be defined as a branch of biology where methods in computer science are applied to problems for finding solutions. Computer applications in biology began in the 1960s in parallel with technological development in both areas, progressed rapidly, and have become a popular area. This resulted in the new research area of bioinformatics. In other words, in the 21st century, biology is both a science in the laboratory and a science based on information technology. Today computer methods are used in all areas of biology. As a result of this, bioinformatics science has emerged. Some terms in biology can be briefly explained as follows.

Genes and proteins: DNA is a macromolecule where genes, organisms' hereditary characteristics, protein-synthesizing coding, enzymes, and other macro and micromolecules' hosts are hosted. Genes are DNA sequences. In other word, genes are carried on chromosomes where they are called very large DNA molecules [1]. A gene, forming part of a chromosome, is a specific nucleotide sequence. The number and sequence of the genes on the chromosomes for each organism are certain.

Each protein encoded by a gene is assigned to a functional purpose. This relationship is called the "gene of an enzyme" hypothesis. Genes encrypt the function of expressed protein. More than one enzyme can run for a reaction, and these enzymes run as a team.

Genes and proteins are the basic concepts of life in biology. Cellularity and vitality of life is a continuation

*Correspondence: adresverme@gmail.com

of the existence of cells that support vital body proteins. Good descriptions of the structure of proteins of organisms and definitions of cellular life for the right medication are also used in drug design [2]. These macromolecules are also effective in treatment of diseases and drug design. Today, without identifying the target proteins, drug improvement is rare. For this, a very good knowledge of the biological response mechanisms is required [3]. Identification of the causative genes for the detection of diseases is a key factor. In the same clinical phenotypes, genetic diseases usually cause the same function, revealed by gene with which they are associated [4]. Expression of thousands of genes simultaneously in a sequence as large as the disease genes for classification is a difficult task [5].

Proteins in an organism and cells are the basis of structural and functional reactions. For this, protein is coded in related DNA areas. This code is read in ribosomes. An appropriate organic structure is formed by amino acids.

Proteins are formed by constituting peptide bonds among amino acids and later they gain suitable conformations. Amino acids are aligned in a specific order for constituting proteins. There are 20 different amino acid types in bodies of living beings. These amino acids do not bind to each other in random forms. On the contrary, each protein has a distinguished amino acid order. The absence of a single amino acid in the protein structure or change of place makes that protein molecularly useless or makes it have different functions. Therefore, each amino acid must occur in exactly the right place [6]. Figure 1 depicts a basic structure of an amino acid.

The combination of different numbers of amino acids in different orders constitutes the structures of proteins. Only amino acids in the primary structure of a protein molecule consist of the adjacent and the chaining [7]. Some parts of DNA sequences have specific functions, and these parts are called genes. The problem of determining genes in DNA sequences is known as motif discovery. The method used for motif discovery in this paper was introduced in [8].

The primary structures of proteins synthesized by amino acid sequences, amino acid sequences, Ramachandran diagrams, and dipeptide sequences were analyzed in this study for patient samples obtained in 2012 from www.bioinformatics.org/pcgdb/Genes and www.genecards.org. Figure 2 illustrates a primary structure of a protein. Computational motif discovery means to investigate repeated patterns in a sequence. In this way, analysis of biological data may reveal some important points to determine the reasons for some diseases.

The aim of this paper is to analyze the biological data obtained from patients for revealing the important knowledge from biological data such as similarities of amino acids sequences, similarities of Ramachandran diagrams, similarities of dipeptide sequences, and similarities of motifs.

The organization of this paper is as follows. Section 2 describes the motif discovery algorithm, Section 3 gives some information about cancer and genes, Section 4 includes data analysis methods and results, and Section 5 concludes the paper.

2. Motif discovery algorithm

The algorithm used for motif discovery in this study was introduced in [8], and that algorithm will be described in this section. Assume that Γ is an alphabet where $\Gamma=\{a_1,a_2,a_3,\ldots,a_n\}$, $1\leq i\leq n$, a_i is called a singleton, and L is language defined over Γ and $L=\{x|x\in\Gamma^*\}$ such that $L=\Gamma^*=\{\varepsilon, a_1, a_2, a_3, \ldots, a_n, a_1a_1, a_1a_2, \ldots, a_1a_n, a_2a_1, a_2a_2, \ldots, a_2a_n, \ldots.., a_na_1, a_na_2, \ldots.., a_na_n, a_1a_1a_1, a_1a_1a_2, a_1a_1a_3, \ldots., a_1a_1a_n, a_1a_2a_1, a_1a_3a_1, \ldots.., a_1a_na_1, \ldots\ldots\ldots..\}$. For example, $\Gamma=\{A, C, G, T\}$ and $L=\Gamma^*=\{\varepsilon$, A, C, G, T, AA, AC, AG, AT, CA, CC, CG, CT, GA, GC, GG, GT, TA, TC, TG, TT, AAA, AAC, AAG, AAT, ACA, AGA, ATA, $\ldots\ldots\ldots..\}$.

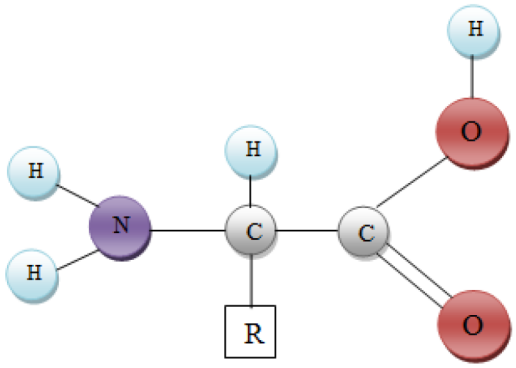

Figure 1. The basic structure of amino acid.

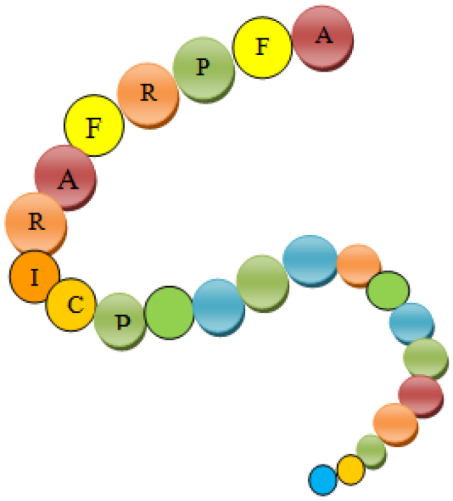

Figure 2. Primary structure of protein representation.

The words in L are sequences of DNA sequences. The sequences S∈L with small sizes do not belong to the scope of this study, and the sequences with large sizes belong to the scope of this study. Assume that S is a sequence of L. S=$s_1 s_2 \ldots s_n$ and k is an integer such that 2≤k≤n. A factor of S is a subsequence $s_{[i,j]}$ such that $s_{[i,j]}=s_i s_{i+1} \ldots . s_j$, 1≤i≤j≤n. The factors of k-lets are factors of size k. For example:

S=ACTCAGCTCTG, and set of factors of k-lets is a multiset. So,

Multiset of 2-lets Ψ_2={AC, CT, TC, CA, AG, GC, CT, TC, CT, TG}

Multiset of 3-lets Ψ_3={ACT, CTC, TCA, CAG, AGC, GCT, CTC, TCT, CTG }

. . . . . .

Multiset of k-lets $\Psi_k = \left\{ \overbrace{AAA.....A}^{k}, \overbrace{AAA.....AC}^{k}, \overbrace{AAA.....AG}^{k}, \right.$

. . . . . $\vdots$ . . . . . .

$\left. \overbrace{TTT.....TA}^{k}, \overbrace{TTT.....TC}^{k}, \overbrace{TTT....TG}^{k}, \overbrace{TTT....TT}^{k} \right\}$

More formally, Ψ_k={$s_{[i,i+k-1]}$, $s_{[i+1,i+k]}$, $s_{[i+2,i+k+1]}$, ...} . The multiset can be defined as $\Psi = \bigcup_{i=1}^{n} \bigcup_{j=1}^{n-i+1} \{s_{[j,j+i-1]}\}$ where Ψ is a multiset. In other words, the multiset Ψ can be redefined as $\Psi = \bigcup_{i=2}^{n-1} \bigcup_{j=1}^{n-i+1} \{s_{[j,j+i-1]}\}$.

All elements of Ψ are not motifs; they are candidate motifs, and we assume that φ is a CanMotif (Candidate Motif) where CanMotif is a triple structure $\varphi=(S_\varphi, I_\varphi, L_\varphi)$. The CanMotif φ is called the L_φ-lets CanMotif and S_φ is a L_φ-lets factor, since it has L_φ singletons. The set of CanMotifs is Φ, and Φ_k, 2≤k≤n-1, is the set of k-lets CanMotifs, and φ_i, $\varphi_j \in \Phi_k$. If $S_{\varphi_i} = S_{\varphi_j}$, and $I_{\varphi_i} = I_{\varphi_j}$, then CanMotifs φ_i,

φ_j are equal CanMotifs. If $S_{\varphi_i} = S_{\varphi_j}$, and $I_{\varphi_i} \neq I_{\varphi_j}$, then CanMotifs φ_i, φ_j are equivalent CanMotifs. A Motif is an instance in the set of CanMotifs.

Brute-force motif discovery algorithm

1. k←2
2. Generate k-lets factors from the given sequence S.
   Compute $\Psi_2 \leftarrow \bigcup_{j=1}^{n-1} \left\{s_{[j,j+1]}\right\}$
3. Obtain Φ_k from Ψ_k by concatenation of each element of Ψ_k with its starting position in S and k. So, there is a triple structure such as (γ_i,I_i,k) where $\gamma_i \in \Psi$ and I_i is the starting position of γ_i in S.
4. While $\Phi_k \neq \emptyset$ do the following steps.
5. For i←1, . . . , n-k+1
   a) $\varphi_i \in \Phi_k$ and $\#(\varphi_i)$=1 where $\#(\varphi_i)$ is number of occurrences of φ_i in Φ_k.
   b) For i←1, . . . , n-k+1
   c) For j←i+1, . . . , n-k+1
   d) Assume that φ_i, $\varphi_j \in \Phi_k$, φ_i=$(S_{\varphi_i}, I_{\varphi_i}, L_{\varphi_i})$ and φ_j=$(S_{\varphi_i}, I_{\varphi_i}, L_{\varphi_i})$. If $S_{\varphi_i} = S_{\varphi_j}$and $I_{\varphi_i} \neq I_{\varphi_j}$, then $\#(\varphi_i) \leftarrow \#(\varphi_i)$+1, $\#(\varphi_j) \leftarrow \#(\varphi_j)$+1.
   e) For i←1, . . . , n-k+1
   f) if$\varphi_i \in \Phi_k$ and $\#(\varphi_i)$=1, then $\Phi_k \leftarrow \Phi_k$-φ_i.
   g) Φ_{k+1}=$\emptyset$
   h) For i←1, . . . , n-k+1
   i) if $\varphi_i \in \Phi_k$, φ_i=$(S_{\varphi_i}, I_{\varphi_i}, L_{\varphi_i})$, then Φ_{k+1}=Φ_{k+1}+$\{(S_{\varphi_i}||S\left[I_{\varphi_i} + L_{\varphi_i}\right], I_{\varphi_i}, L_{\varphi_i} + 1)\}$
   j) k←k+1

The Delphi source code of this algorithm is as follows.

Procedure 1. MainProcedure

```
procedure TForm1.Button1Click(Sender: TObject);
begin
  GenerateCanMotifs;
  Continue:=True;
  while (Continue) do
  begin
    CanMotifSorting;
    EliminateDuplication;
    GenerateNextCanMotifs;
  end;
end;
```

Procedure 2. Generate_2-lets_CanMotifs

```
procedure TForm1.GenerateCanMotifs;
var
  Local variables list
begin
  Read_File_of_AminoAcids;
  GeneNumber:=0;
  while not(eof_File) do
  begin
    if Read_Character in ['A','C','G','T','F','L','S','Y','W','P','H','Q','R','I','M','N','K','V','D','E']
then
    begin
        GeneNumber:=GeneNumber+1;
        AminoAcids[GeneNumber]:= Read_Character;
    end;
      Read_File_of_AminoAcids;
  end;
  for i:=1 to (GeneNumber-1) do
  begin
    Motif2[i].Motif:= AminoAcids[i]+DNA[i+1];
    Motif2[i].Position:=i;
  end;
  MotifNumber:=GeneNumber-1;
  MotifLength:=2;
end;
```

3. Cancer and genes

Cancer is a disease of cell division without control. Cancer occurs while changes occur in genetic structures taking active processes in the birth and death of cells. There are many genes causing cancer, and there are many mutation points on these genes. On the contrary, the changes in genetic structures cause different cancers. In addition, mutations in the same region of a gene can trigger different types of cancer. Types of cancer in people with the same disease can be caused by mutations in different genetic regions. As a result, cancer develops, a disease that has to be understood because it is difficult and complex.

Symptoms and diagnosis of tumors follow traditional methods of tumor morphological (formal) structure based on pathologic examination and tissue operations. This morphological analysis obtained from the limited information is often insufficient to diagnose cancer. However, gene expression, i.e. the gene-based diagnostic method, is more accurate and more reliable. Gene sequences obtained from gene expression, even for heterogeneous types of cancer, can be divided into subclasses. Recently, artificial neural networks, evolutionary computations, neighborhood-based methods, etc. are used for gene analysis and some of these methods have the potential to be applied to bioinformatics problems [9–22].

Procedure 2. Expand CanMotifs

```
procedure TForm1.GenerateNextCanMotifs;
var
   Local variables list;
begin
  k:=1;
  MotifLength:=MotifLength+1;
  i:=1;
  while (i<=MotifNumber) do
  begin
    if Motif3[i].Position>1 then
    begin
      Motif2[k].Motif:= AminoAcids[Motif3[i].Position-1]+
                Motif3[i].Motif;
      Motif2[k].Position:=Motif3[i].Position-1;
      k:=k+1;
    end;
    if (Motif3[i].Position+MotifLength-1)<=GeneNumber then
    begin
      Motif2[k].Motif:=Motif3[i].Motif+
           AminoAcids[Motif3[i].Position+MotifLengthk-1];
      Motif2[k].Position:=Motif3[i].Position;
      k:=k+1;
    end;
    i:=i+1;
  end;
  MotifNumber:=k-1;
end;
```

4. Data analysis

The data analysis process is handled in three ways. The first one is the primary protein structure, number of amino acids, etc.; the next one is the Ramachandran diagrams; and the last one is the discovery of motifs in the obtained data. In this part of the study, analysis of the patient data from the two different websites mentioned in Section 1 is performed. The MATLAB toolbox was used for the first part of data analysis. The working data obtained of different genes from different patients in the database related to cancer patients are examples of the proteins that have been synthesized. Protein prefixes on the basis of their amino acid sequences in the primary structure are discussed.

As seen from Figure 3, the primary amino acid sequence of a protein in the amino acid sequences is provided, and their percentages are also provided. ABCB1, BRCA1, and BRCA2 are gene expressions of different patients and the SEQTOOL results for the number of amino acids and percentages are shown in Figure

4. ABCB1, BRCA1, and BRCA2 genes and protein samples were evaluated together because they trigger the same types of cancer. Although breast cancer genes BRCA1 and BRCA2 in the literature have been focused on, it has also been proven that those genes are effective in pancreatic cancer.

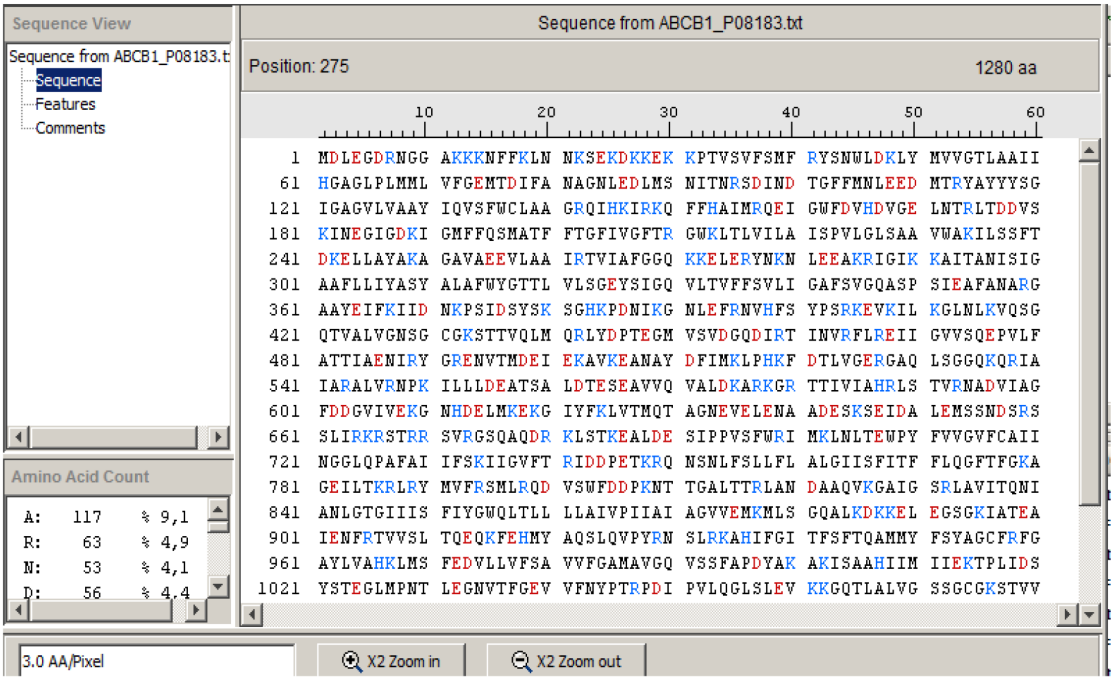

Figure 3. The result of the SEQTOOL command.

	ABCB1				BRCA1				BRCA2				ALOX5AP	
Tür	Sayı	Yüzde		Tür	Sayı	Yüzde		Tür	Sayı	Yüzde		Tür	Sayı	Yüzde
A	117	9.140625		A	84	4.50886		A	175	5.11995		A	11	6.8323
C	7	0.546875		C	44	2.36178		C	76	2.22352		C	2	1.2422
D	56	4.375		D	85	4.56253		D	171	5.00293		D	3	1.8634
E	75	5.859375		E	198	10.628		E	293	8.57226		E	6	3.7267
F	70	5.46875		F	49	2.63017		F	136	3.97894		F	16	9.9379
G	99	7.734375		G	87	4.66989		G	122	3.56934		G	12	7.4534
H	19	1.484375		H	49	2.63017		H	84	2.45758		H	2	1.2422
I	104	8.125		I	77	4.13312		I	187	5.47104		I	11	6.8323
K	85	6.640625		K	137	7.35373		K	322	9.42071		K	4	2.4845
L	119	9.296875		L	156	8.37359		L	281	8.22118		L	18	11.18
M	32	2.5		M	30	1.61031		M	45	1.31656		M	3	1.8634
N	53	4.140625		N	121	6.4949		N	229	6.69982		N	7	4.3478
P	29	2.265625		P	96	5.15298		P	149	4.35927		P	5	3.1056
Q	53	4.140625		Q	97	5.20666		Q	154	4.50556		Q	8	4.9689
R	63	4.921875		R	76	4.07944		R	110	3.21826		R	7	4.3478
S	88	6.875		S	224	12.0236		S	381	11.1469		S	10	6.2112
T	69	5.390625		T	111	5.95813		T	220	6.43651		T	12	7.4534
V	96	7.5		V	101	5.42136		V	184	5.38327		V	14	8.6957
W	11	0.859375		W	10	0.53677		W	20	0.58514		W	1	0.6211
Y	35	2.734375		Y	31	1.66398		Y	79	2.31129		Y	9	5.5901
	1280				1863				3418				161	

Figure 4. The distribution of amino acids and their percentages for ABCB1, BRCA1, BRCA2, and ALOX5AP genes.

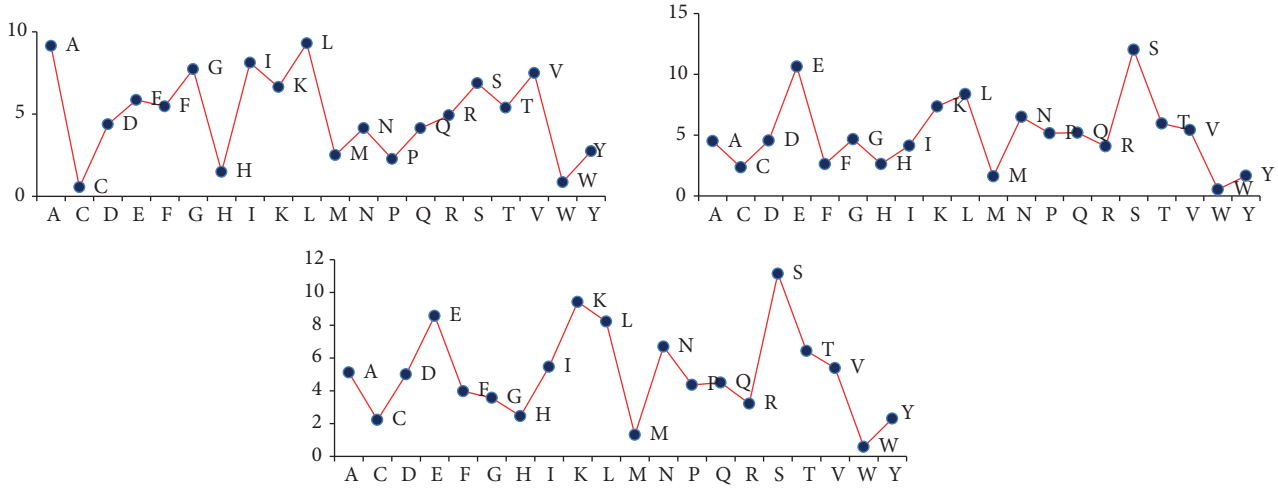

Figure 5. The graphical representations of amino acids of proteins corresponding to genes ABCB1, BRCA1, and BRCA2.

If Figure 5 is examined closely, although the values of points look different, the characteristics of data are similar. The graphic shows the similarity in the increasing and decreasing regions. For example, the number of amino acids make the graph be different. Ultimately, it is considered that these three gene expressions trigger the same cancer, and the primary structures of proteins synthesized by these gene expressions have similar characteristics.

Figure 6 depicts the number of amino acids in proteins that are synthesized by TP53, TP63, and TP73 gene expressions. These genes are genes also proven to induce cancers. With a careful examination of the characteristics of the chart above, the graphic will be seen as very similar. Figures 7–10 illustrate that the distributions of atoms demonstrate similar characteristics.

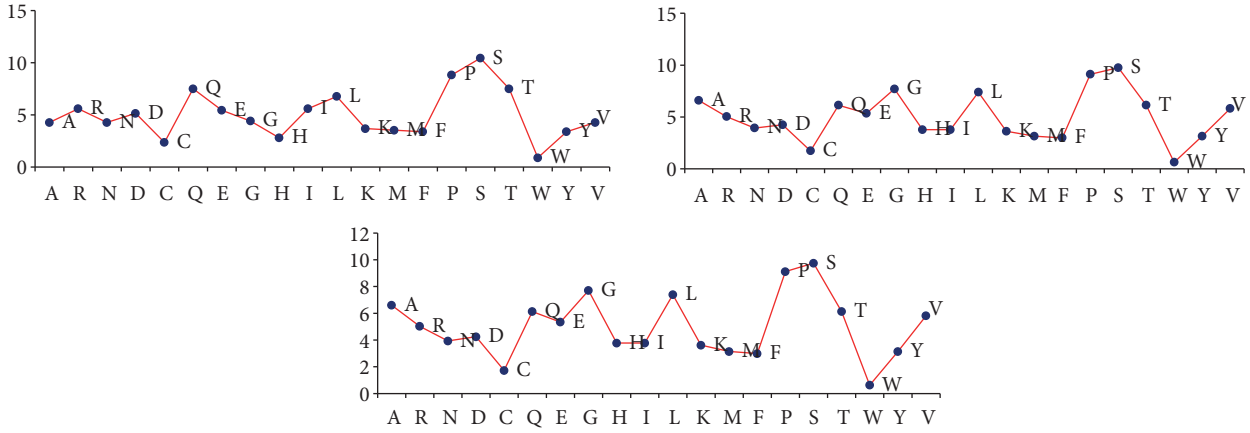

Figure 6. The graphical representations of the number of amino acids for TP53, TP63, and TP73 gene expressions.

Ramachandran diagrams: Ramachandran diagrams contain all possible combinations of bond angles (phi and psi angles) where angles occur in the polymerized peptide structures. These angles are used in the determination of their specific protein secondary structures. Peptide linkages formed by molecules vary depending on the power and energy of this aspect of this psi and phi. The diagrams drawn for these angles versus energy are called Ramachandran diagrams.

ALOX5AP(-)				BRCA2(-)				BRCA1(-)		
Atom	Sayı	Yüzde		Atom	Sayı	Yüzde		Atom	Sayı	Yüzde
C	849	32.98368		C	16802	31.28165		C	8908	30.93485
H	1288	50.03885		H	26738	49.78031		H	14246	49.47215
N	206	8.003108		N	4640	8.638665		N	2554	8.869287
O	226	8.780109		O	5411	10.0741		O	3014	10.46673
S	5	0.19425		S	121	0.225276		S	74	0.25698
	2574				53712				28796	

Figure 7. The distributions of atoms in proteins synthesized by ALOX5AP, BRCA2, and BRCA2 gene expressions.

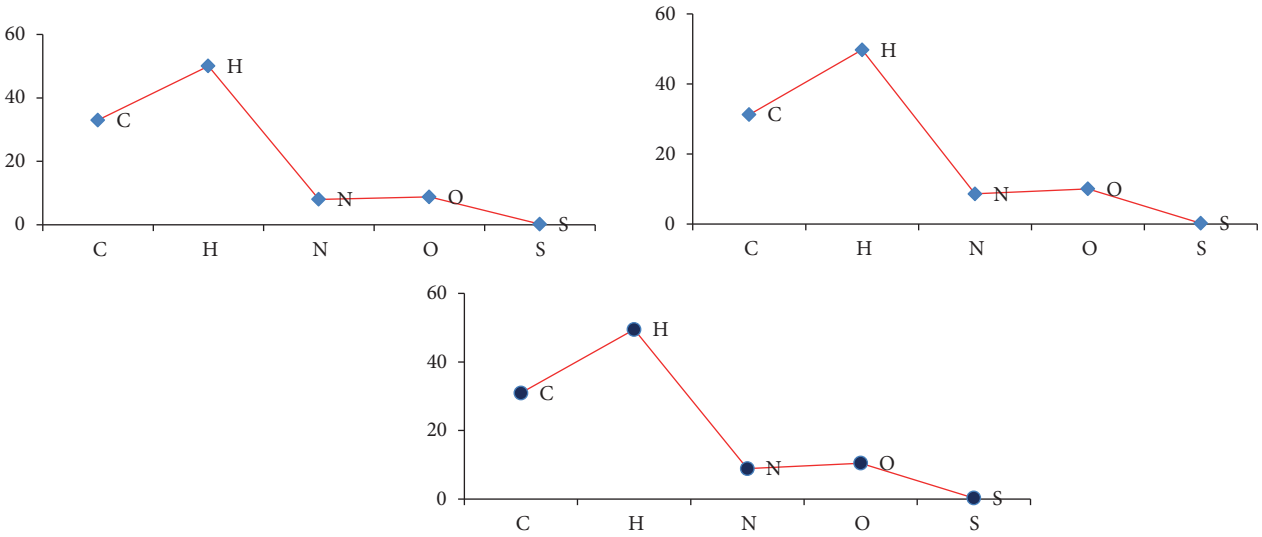

Figure 8. The graphical representations of distribution of atoms in proteins synthesized by ALOX5AP, BRCA2, and BRCA2 gene expressions.

TNF(+)			TNFSF13B(+)		
Sayı	Yüzde		Atom	Sayı	Yüzde
1152	31.70933		C	1293	31.33786
1825	50.23397		H	2080	50.41202
313	8.615469		N	350	8.482792
337	9.27608		O	393	9.524964
6	0.165153		S	10	0.242365
3633				4126	

Figure 9. The distributions of atoms in proteins synthesized by TNF and TNFSF13B gene expressions.

TP53(-)			TP73(-)		
Sayı	Yüzde		Atom	Sayı	Yüzde
1898	31.42384		C	2824	31.54602
2980	49.33775		H	4427	49.45264
548	9.072848		N	815	9.104111
592	9.801325		O	858	9.58445
22	0.364238		S	28	0.312779
6040				8952	

Figure 10. The distributions of atoms in proteins synthesized by TP53 and TP73 gene expressions.

Tumor necrosis factor (TNF), which is secreted by many cell types and provides the destruction of cancer cells, is a cytokine and a glycoprotein hormone. It is noticed that the genes causing tumor formation have their proteins' Ramachandran diagram angles concentrated in areas at the bottom left in the negative angle density regions, and the genes preventing tumor formation have their proteins' Ramachandran diagram angles concentrated in areas at the top left in the positive angle density regions (Figures 12 and 13).

In Figure 2, as shown in the represented form of the amino acid sequences of proteins, primary chains of amino acids are called dipeptide bonds. In our studies of the same gene family, for synthesized protein samples the similarity of the dipeptide bond also been shown to be very similar (Figures 14 and 15).

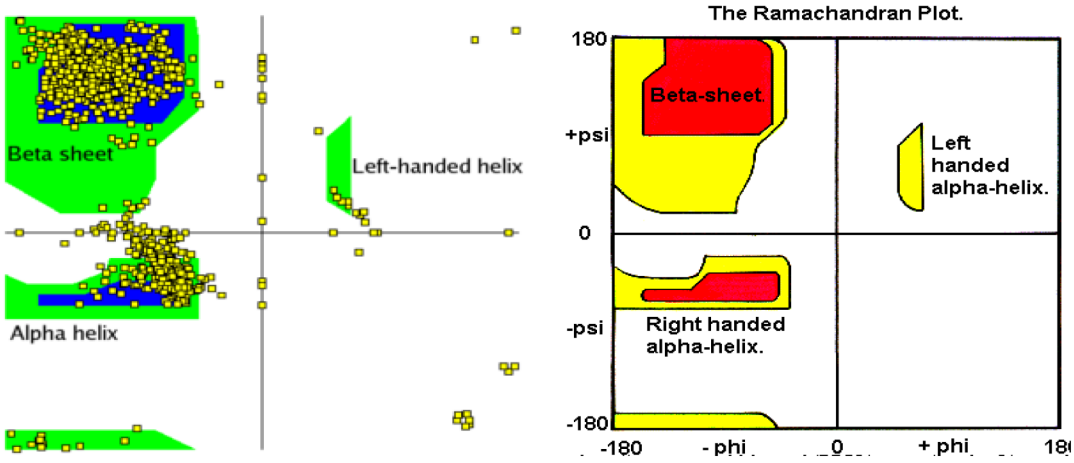

Figure 11. An example of Ramachandran diagrams [23,24].

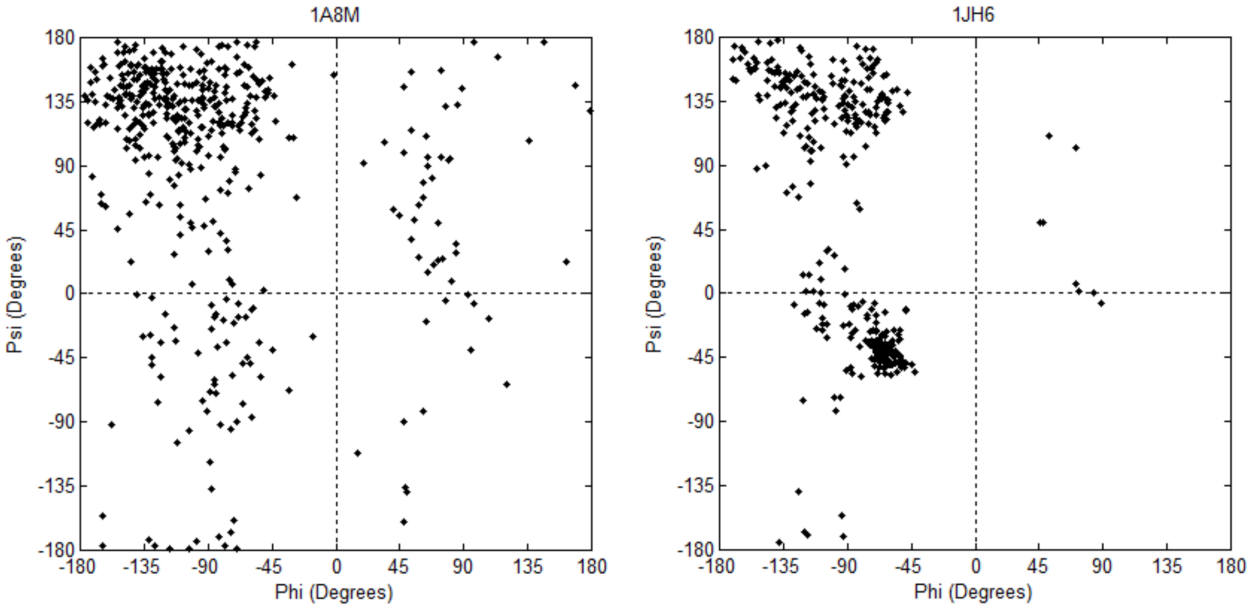

Figure 12. Some gene expressions thought to cause tumor formation: proteins' Ramachandran diagrams.

Figure 14 depicts the CYS dipeptide bonds. It is noticed that BRCA1 and BRCA2 have similar CYS dipeptide bonds, and TP53, TP63, and TP73 also have similar CYS dipeptide bonds. Figure 15 illustrates a similar case for amino acid MET. BRCA1 and BRCA2 again have similar dipeptide bonds, and TP53, TP63, and TP73 have similar dipeptide bonds. In Figures 14 and 15, the amino acids methionine and cysteine ??and other amino acids have been examined for frequency of dipeptide bonds. The same method can also be applied to the frequency of other dipeptide bonds.

Motif discovery: The motifs in the gene expressions of ABCB1, BRCA1, BRCA2, TNF, TNFSF13B, TP53, TP63, and TP73 are determined, and proteins formed by these gene expressions are obtained. The motif discovery is handled by using the algorithm in Section 2.

TP63 and TP73 have a lot of common motifs, and some of them are LYNFMCNSSCVGGMNRRPILII, KKLYCQIAKTCPIQIKV, IRAMPVYKKAEHVT, FQSSTAKSATWTYS, PYAQPSSTF, and RICACPGRDR. There are many motifs of length five, six, and seven belonging to these genes, but we did not write them here.

The genes TP53, TP63, and TP73 have common motifs such as VPYEPP, CACPGRDR, MCNSSC, GGMNRRPIL, GGMNRRPIL, HLIRVEGN, and AKTCP.

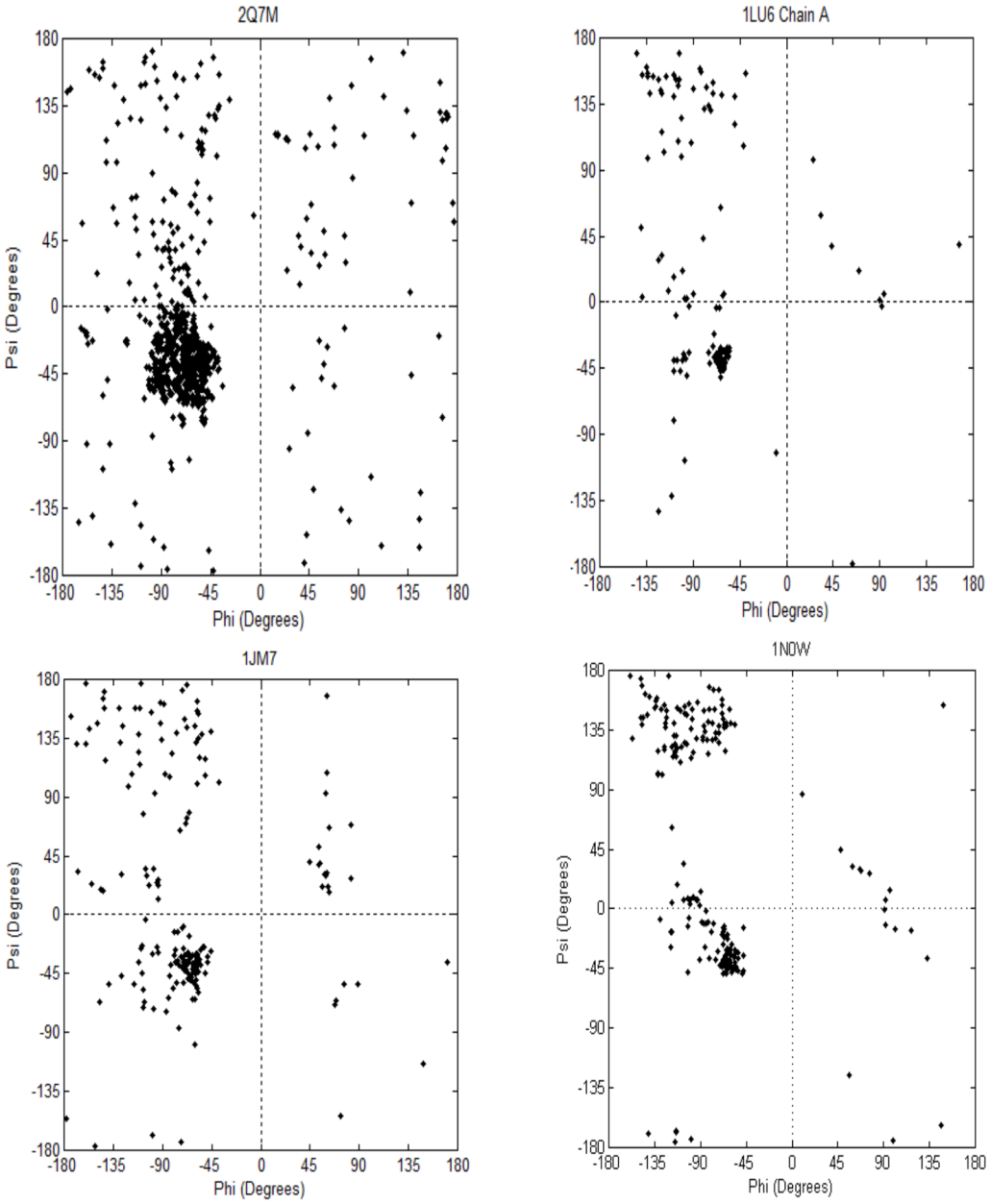

Figure 13. Several genes thought to eliminate the production of the tumor: proteins' Ramachandran diagrams.

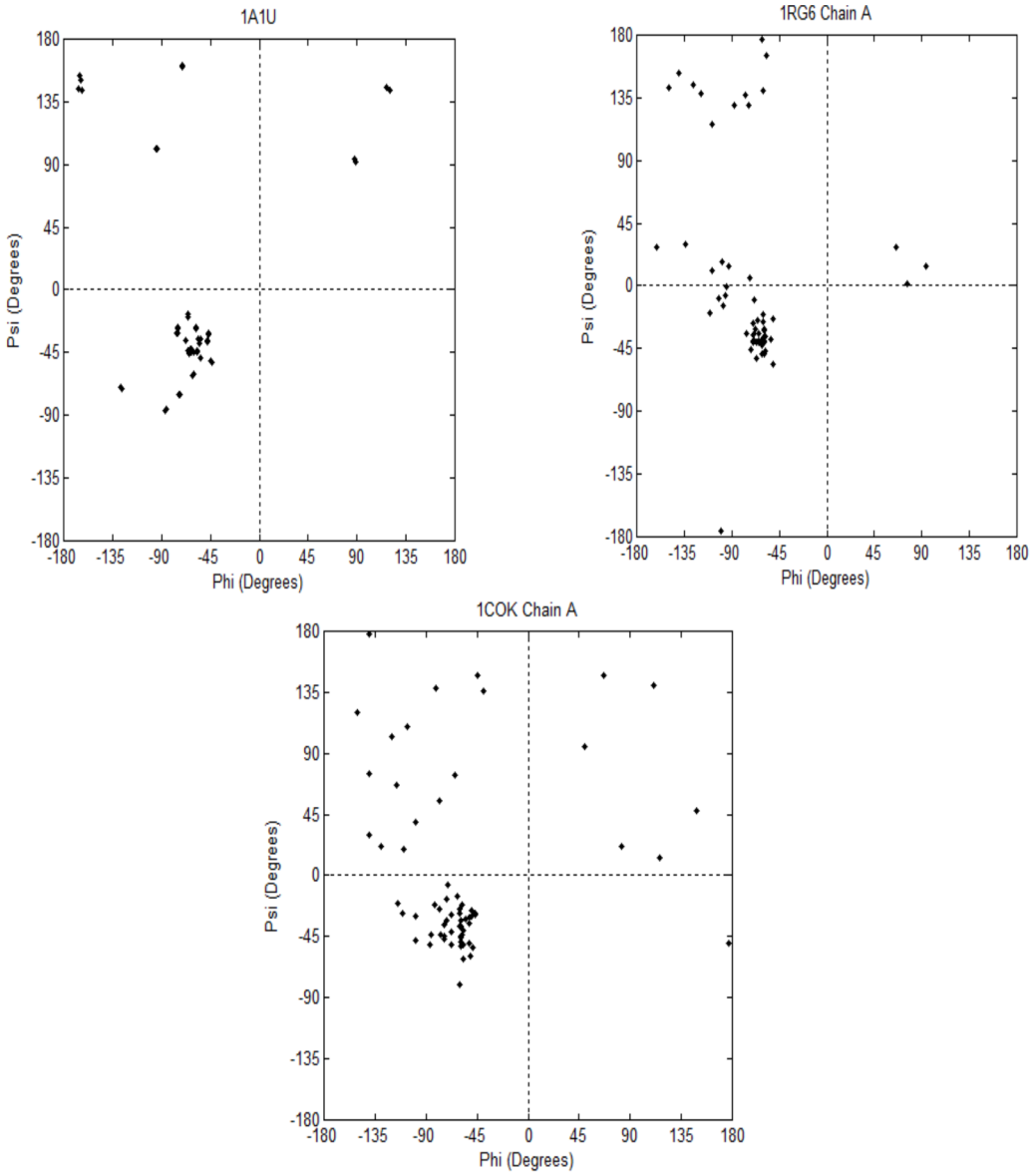

Figure 13. Continued.

The genes BRCA1 and BRCA2 have common motifs such as GSDDS, GSDSS, ISLLE, LEESG, and SLFSD.

The genes ABCB1 and RCA2 have common motifs such as EEVLAA, SIGQV, and TIAEN.

The genes ABCB1 and RCA1 have common motif LSSFT. This means that the proteins that may cause cancer have common motifs.

5. Conclusions

Genes and their synthesized proteins are important macromolecules in our life. Identifying these genes and proteins is necessary for a more healthy life. The genes and their synthesized proteins have roles in the diagnosis of diseases, so they attract researchers.

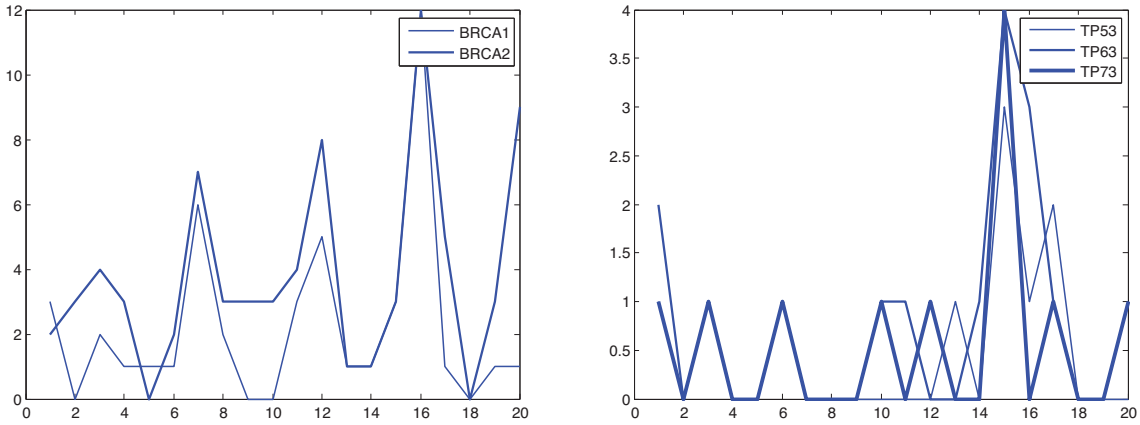

Figure 14. CYS dipeptide distributions for BRCA1, BRCA2, TP53, TP63, and TP73 genes.

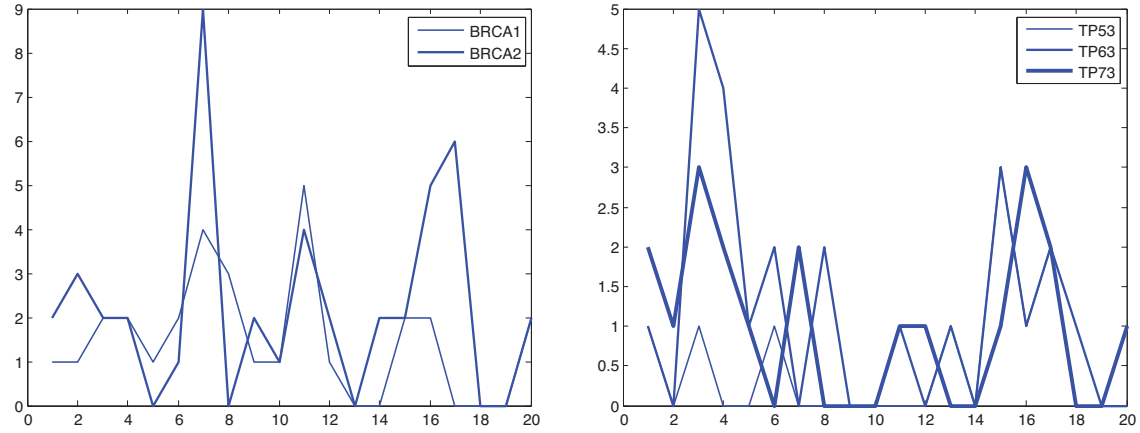

Figure 15. MET dipeptide distributions for BRCA1, BRCA2, TP53, TP63, and TP73 genes.

In this study, samples of genes and proteins obtained from cancer patients were examined. The similarities of these genes and proteins may help genetic treatments, and these similarities can be categorized with motifs, amino acid distributions, specified amino acid dipeptide bonds, and Ramachandran diagrams.

The proteins ABCB1, BRCA1, BRCA2, TNF, TNFSF13B, TP53, TP63, and TP73 were analyzed and they have similar motifs, Ramachandran diagrams, dipeptide bonds between specified amino acids, and other amino acid distributions. This means that these proteins hide more genetic information and should be the subjects of future research. Such studies would be especially important for genetic diseases such as diabetes, Down syndrome, and hemophilia.

References

[1] Chen HS, Lee G, Peng SL. On the gene team mining problem. In: 9th International Conference on Fuzzy Systems and Knowledge Discovery; 29–31 May 2012; Sichuan, China.

[2] Kim W, Li M, Wang J, Pan Y. Essential protein discovery based on network motif and gene ontology. In: IEEE International Conference on Bioinformatics and Biomedicine; 12–15 November 2011; Atlanta, GA, USA.

[3] Celal A. Yeni ilaç geliştirilmesi. In: Türk Farmakoloji Derneği Klinik Farmakoloji Çalışma Grubu Elektronik Bülteni; 2005. pp. 6-10 (in Turkish).

[4] Quan G, Du Y, Huang J, Wang Y. Prediction of human disease genes based on associations between phenome and proteins. In: 2nd International Conference on Biomedical Engineering and Informatics; 17–19 October 2009; Tianjin, China.

[5] İbrahim MAR, Jassim S, Cawthome MA, Langlands K. Pathway-based gene selection for disease classification. In: International Conference on Information Society; 27–29 June 2011; London, UK.

[6] Krawetz SA, Womble DD. Introduction to Bioinformatics: A Theoretical and Practical Approach. Totowa, NJ, USA: Humana Press; 2003.

[7] Onat T, Emerk K, Sözmen EY. İnsan Biyokimyası. Ankara, Turkey: Palme Yayıncılık, 2002 (in Turkish).

[8] Karcı A. Efficient automatic exact motif discovery algorithms for biological sequences. Expert Syst Appl 2009; 36: 7952-7963.

[9] Wang L, Chu F, Xie W. Accurate cancer classification using expressions of very few genes. IEEE ACM T Comput Bi 2007; 4: 40-53.

[10] Saraswathi S, Sundaram S, Sundararajan N, Zimmermann M, Nilsen-Hamilton M. ICGA-PSO-ELM approach for accurate multiclass cancer classification resulting in reduced gene sets in which genes encoding secreted proteins are highly represented. IEEE ACM T Comput Bi 2011; 8: 452-463.

[11] Canayaz M, Karcı A. A new metaheuristic cricket-inspired algorithm. In: Second International Eurasian Conference on Mathematical Sciences and Applications; 26–29 August 2013; Sarajevo, Bosna-Herzegovina.

[12] Karcı A. A new meta-heuristic algorithm based on chemical process: atom algorithm . In: 1st International Eurasian Conference on Mathematical Sciences and Applications; 3–7 September 2012; Kosovo.

[13] Karcı A. Theory of saplings growing up algorithm. In: International Conference on Adaptive and Natural Computing Algorithms; 11–14 April 2007. pp. 450-460.

[14] Karcı A. Saplings sowing and growing up algorithm convergence properties. In: International Symposium on Innovations in Intelligent Systems and Applications; 2007; İstanbul. pp. 322-326.

[15] Karcı A, Yiğiter M, Demir M. Natural inspired computational intelligence method: saplings growing up algorithm. In: International Kyrgyz-Kazak Electronics and Computer Conference; 18–19 October 2007; Bishkek-Almaty.

[16] Demir M, Yiğiter M, Karcı A. Application of saplings growing up algorithm to clustering medical data. In: International Kyrgyz-Kazak Electronics and Computer Conference; 18–19 October 2007; Bishkek-Almaty.

[17] Demir M, Karcı A. Veri kümelemede fidan gelişim algoritmasının kullanılması. In: Elektrik Elektronik Bilgisayar Biyomedikal Mühendisliği 12. Ulusal Kongresi ve Sergisi; 14–18 November 2007; Eskişehir, Turkey (in Turkish).

[18] Canayaz M, Karcı M. A new metaheuristic cricket-inspired algorithm. In: 2nd International Eurasian Conference on Mathematical Sciences and Applications; 26–29 August 2013; Sarajevo, Bosnia and Herzegovina.

[19] Özdağ R, Karcı A. The application of electromagnetism-like algorithm for the dynamic deployment problem in wireless sensor networks. In: 2nd International Eurasian Conference on Mathematical Sciences and Applications; 26–29 August 2013; Sarajevo, Bosnia and Herzegovina.

[20] Demir M, Karcı A, Özdemir M. Fidan gelişim algoritması yardımı ile DNA motiflerinin keşfi. Çankaya University Journal of Science and Engineering 2011; 8: 51-62 (in Turkish).

[21] Erdoğan Yıldırım A, Karcı A. Solutions of travelling salesman problem using genetic algorithm and atom algorithm. In: 2nd International Eurasian Conference on Mathematical Sciences and Applications; 26–29 August 2013; Sarajevo, Bosnia and Herzegovina.

[22] Karadoğan A, Karcı A. Artificial atom algorithm for reinforcement learning. In: 2nd International Eurasian Conference on Mathematical Sciences and Applications; 26–29 August 2013; Sarajevo, Bosnia and Herzegovina.

[23] http://www.cryst.bbk.ac.uk/PPS95/course/3_geometry/rama.html, 2014.

[24] http://de.wikipedia.org/wiki/Ramachandran-Plot, 2014.