Revista de Chimie
https://revistadechimie.ro
https://doi.org/10.37358/Rev.Chim.1949

ITSRepeats: Amino Acids Repeats and Composition Differences for Detecting Proteins Causing Bipolar Related Disorder

MARIYAPPAN MUTHUPANDI[1*], VAIRAMUTHU THANGARAJ[2]
[1]Faculty of Information and Communication Engineering, University College of Engineering Tiruchirappalli BIT Campus, Tiruchirappalli, 620024, India
[2]Department of Chemistry, University College of Engineering, Anna University (BIT Campus), Thiruchirappalli, 620024, India

Abstract: *Bipolar disorder is a threat that needs an efficient solution to detect in earlier times for leading the normal life of human beings. Simple Sequence Editor (SSE) is the tool used to analyze the nucleotide and amino acid sequence that helps in this proposed system to detect the real composition. This paper proposed the algorithm called Intelligent Rule-based Triplet and Single Residue Repeats (ITSRepeats) in proteins to classify the protein sequence which causes bipolar related disorder or not. The hamming distance was used to find the similarity between the protein strings which increase the efficiency initially, the patterns of the proteins are extracted and analyzed from the datasets SwissProt, PDBS25, Surface Residues and DisProt. The performance of the proposed algorithms was evaluated based on the performance among these four datasets for similar residues.*

Keywords: *bipolar disorder, single residue repeats, surface eesidues*

1.Introduction

Mental health is vital for the well-being of individuals, societies and countries. Living at fast paces and in a hectic society, people today suffer invisible pressures in family, and at work leading to a range of mental health problems. These are associated with a significant genetic component of risk and the interaction of multiple risk genes with environmental factors. Researches show such disorders as the result of a complex interaction between biological, psychological and social factors. In most parts of the world, mental health and mental disorders are not regarded as important as physical health. Even though large numbers of people are affected by mental and behavioral disorders they never receive treatment. Hence, there exists a widening "treatment gap". The treatment requires the use of high-performance computers and innovative software tools to manage and organize enormous quantities of genomic and proteomic data. It also comprises the development and application of innovative algorithmic techniques for the analysis and interpretation of sequential data. It is also used for the classification and prediction of sequential data which provide insight into the design and validation of experiments for science and technologies. Most human diseases are caused by protein disorders and proteins are changed based on the composition changes of amino acid sequence. The changes in amino acid sequence may lead to many diseases and that cause some serious diseases like bipolar disorder in neuro systems.

The earlier research on Bipolar disorder discovered single and multiple amino acid repeats as the root cause of neurological disorders. As per previous studies, not only the mutations may alter the protein structure and the function with adverse effects but also the natural variants would be responsible for such effects in bipolar disorder. As several life forms exist on earth, it is very difficult to identify all the repeats with the available techniques and software tools. Since such variations give strong impacts on molecule stability and function, an attempt is made to identify and develop a proper web server for repeat analysis. It is believed that the developed server would lead to improved protein stability predictions. This work proposed an Intelligent Rule-based Triplet and Single Residue Repeats (ITSRepeats) algorithm to identify the amino acid repeats. This technique efficiently identifies the Multiple Residue repeats (MRR) and Single residue repeats (SRR) for classifying the proteins which cause bipolar disorders.

**email:mariyappanmuthupandi97@gmail.com*

Literature survey

Based on the studies, many tools are explored to discover the composition of amino acids and their protein structure in various forms. Some of the tools and techniques are discussed in this section. Describe how amino acid sequence repeats are generated, assembled, and evaluated. Our analysis then turns to a paleontological record of an evolutionary and biological process which provides a detailed analysis of the DNA sequence, and its differences from those of other organisms as well as similarities with them [1]. These are repeats of amino acid sequences known as simple sequence repeats (SSRs). The functional potential of repeat elements is influenced by their size, base composition, and intragenic or intergenic location. Some inherited diseases are caused by SSRs, the most intensively studied group being trinucleotide repeats, which cause deleterious effects when they expand. There are more than 20 human disorders that have been proven to be caused by dynamic mutations, with the majority occurring as the result of extended trinucleotide repeats, though tetranucleotide repeat expansions and penta-nucleotide repeat expansions have also been described [2]. These repetitive sequences can occur in exons, introns, or intergenic regions, which is why they are also known as tandem repeats. There are many biological processes affected by tandem replication, including evolution, development, brain function and behavioral changes [3-5] such as Bipolar disorder, Alzheimer's disease and associated behavioral and/or psychological symptoms, Attention deficit hyperactivity disorder (ADHD), Schizophrenia. Repeating sequence of amino acids composed of several consecutive residues. The longer repeats are often caused by hydrophilic amino acids, such as glutamine. In the human genome, repeats are most often composed of poly (CAG) repeats consisting of glutamine repeats, which are encoded by CAG. Triplet repeats were initially observed in humanoid proteins and then in other transcription factors and human genetic diseases [6]. Ataxia is a type of brain disease caused by trinucleotide repeats that occur in non-protein-coding regions of the genome. In addition to SCA8, SCA10, and SCA12, six other spinocerebellar ataxias discussed above are characterized by CAG repeat expansions encoding polyglutamine tracts, in each case, the repeats are not coded. Hence, there may be some pathogenic mechanisms initiating before the translation of the disease proteins in polyglutamine diseases. It is now evident that polyglutamine tracts have a toxic effect on function [7, 8].

A polyglutamine disorder occurs when extended CAG repeats are present in the respective disease proteins and cause extended polyglutamine tracts in the respective disease proteins. Currently, there are nine polyglutamine disorders. This gene has an expanded and unstable trinucleotide repeat found on Huntington's disease chromosomes [9]. Despite their differences, all 'polyglutamine' diseases exhibit the same characteristic pathology: protein aggregates or inclusions. The pathological protein contained in intranuclear inclusions was originally proposed to promote the degeneration of neurons [10, 11]. Poly alanine tracts are associated with developmental disorders with rare occurrences whereas polyglutamine tracts are associated with developmental disorders with more frequency [12]. In this condition, there are usually nucleolar inclusions, indicating that the expansion of the alanine tract leads to protein aggregation, ultimately causing insoluble inclusions and cell death [13-15]. Tri-nucleotide repeat expansions cause loss of protein function in brain disorders such as fragile X syndrome (FRAXA), which is linked with long-range mutations in the FMR1 gene [16].

The compound polyglutamine is produced by a triplet repeat expansion of the CAG codon in most neurodegenerative proteins [17]. Unexpectedly, neuronal proteins contain only modest concentrations of alanine or histidine. Based on these results, they do not have an excess of glutamine-rich regions as a class. But many neurological disorders are associated with polyglutamine tracts and glutamine-containing proteins. These disorders are likely caused by the aggregation of proteins induced by the formation of polyglutamine tracts [18]. There are many neurodegenerative diseases associated with polyglutamine repeats, but polyglutamine repeats are not more common in neurodegenerative disease-specific proteins than they are within collections of proteins not related to neurodegenerative diseases. The sequences may be composed of a simple repetition of a single amino acid or may consist of long tracts of sequence that comport most or all of one amino acid [19]. A study analyzing how extended glutamine repeats in some neurological proteins produce their associated neurodegenerative diseases has

used this model as an explanation [20]. Multiple neurodegenerative diseases have been linked to repeating regions in proteins. Extensive glutamine residues are found in these proteins [21]. A triplet repeat in Huntington's disease encodes a polyglutamine tract that may cause neurons to become increasingly toxic as the repeat increases. The pathology of this disease is caused by an expanded CAG repeat those codes for polyglutamine [22, 23]. Based on their amino acid repeat patterns, amino acids have distinct evolutionary and functional backgrounds. In proteins, repeats with simple patterns, such as single AARs, are mostly found in intrinsically unstructured regions (IURs) [24, 25]. Many neurodegenerative diseases such as Huntington's disease, Alzheimer's and Parkinson's involve simple amino acid or trinucleotide repeats, such as polyQ [26].

2. Materials and methods

2.1. Bipolar detection

The algorithm for detecting SRR and MRR is given below and implemented using Python version 3.7 which also identifies the mutant changes in amino acids was shown in Figure 1 and classifies the stability to predict the possibility of bipolar disorder based on the mutant variants.

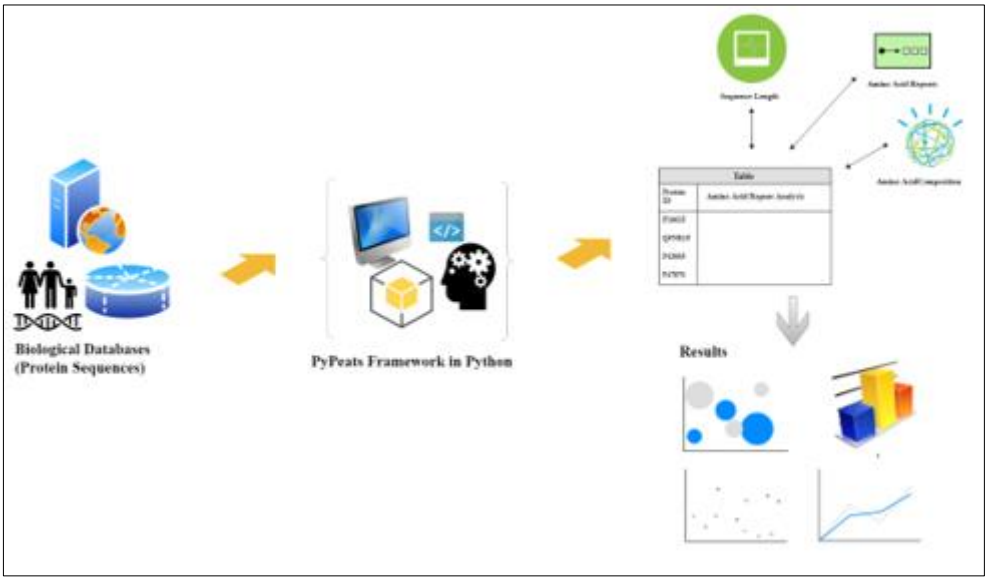

Figure 1. System architecture for PyRepeats Framework in Python

Algorithm: Intelligent Rule based Triplet and Single Residue Repeats (ITSRepeats) with Hamming Distance Similarity

Input: String (S) // String contains the amino acid sequence of protein
Output: SL, SRR, MRR
Step 1: Initialize the SL for the amino acid
Step 2: Compute the length of the sequence as per the obtained dataset
Step 3: Check the partial sequence < Null and Initial <0 // Repeat the sequence
Step 4: Enumerate (S) for the i and c
Step 5: Scan the pairs of nucleotide, Scan (Nucl)
Step 6: Plot the color code (Amino acid) to check similarity for each block
Step 7: Identify the substitutions in alignment of Amino acids
Step 8: Scan the occurrence of short sequence of acids as di-, tri or tetra SL (Nucl)
Step 9: Plots the graph for predefined sequences
Step 10: Base components of individual sequence retrieved from dataset dictionary
Step 11: Calculates the composition of individual (Seq) and Group (Seq)

Step 12: Extract expected value from codon usage biases and codon number
Step 13: Calculate the frequencies for nucleotide composition
Step 14: Check the variation in SL for Amino acids
Step 15: Compare SL (Dataset) with SL (Obtained Seq) using Hamming Distance
Step 16: IF SL (Dataset) >SL (obtained seq) then check for SRR or MRR
Step 17: Return the value for obtained Seq
Step 18: Else repeat the step 11 to 16.
Step 19: End.

Distance metrics are a key part of several machine learning algorithms. These distance metrics are used in both supervised and unsupervised learning, generally to calculate the similarity between data points. An effective distance metric improves the performance of our machine learning model, whether that's for classification tasks or clustering.

2.2. Hamming distance

Hamming Distance measures the similarity between two strings of the same length. The Hamming Distance between two strings of the same length is the number of positions at which the corresponding characters are different.

```
//Coding for Hamming Distance
defhamming_dist(s1, s2):
   assert len(s1) == len(s2)
hd = 0
   for b1, b2 in zip(s1, s2):
      if b1 != b2:
hd += 1
   return hd
defimap(function, *iterables):
iterables = map(iter, iterables)
   for it in iterables:
args = tuple(it)
      if function is None:
         yield tuple(args)
      else:
         yield function(*args)
distances = imap(hamming_dist, *itertools.combinations(ls,2))
for dist in distances:
print(dist)
from Bio import AlignIO
import itertools
aln = AlignIO.read(...)
for r1, r2 in itertools.combinations(aln, 2):
print("\n".join([r1.id, str(r1.seq), r2.id, str(r2.seq), str(hamming_distance(str(r1.seq), str(r2.seq)))]))
```

The algorithm scans the sequence of the amino acids in the protein at the initial stage and after certain mutations, the changes happen in the sequence based on the values provided by the system. The stability decreases if the value of the protein is less than 0 or otherwise the stability increases if the value larger than 0. The variations are verified from the obtained sequences and it checks the string length after each mutation to identify the possibility for bipolar disorder.

3. Results and discussions

The data set was collected from 100 protein sequences with the invariants from UniProt. The frequency occurrences triplets are shown in Table 1. Since the previous studies concluded the deviation

in W, Y and C amino acid frequency would cause bipolar disorder, we gave importance to these amino acids to validate our algorithm. The algorithm was validated by comparing the obtained output with the existing tools such as ProtParam and Protein Information Resources and the sample input sequence was shown in Figure 2.

MWRVRKRGYFGIWSFPLIIAAVCAQSVNDPSNMSLVKETVDRLLKGYDIRLRPDFGGPP
VAVGMNIDIASIDMVSEVNMDYTLTMYFQQAWRDKRLSYNVIPLNLTLDNRVADQLWVP
DTYFLNDKKSFVHGVTVKNRMIRLHPDGTVLYGLRITTTAACMMDLRRYPLDEQNCTLEIE
SYGYTTDDIEFYWRGDDNAVTGVTKIELPQFSIVDYKLITKKVVFSTGSYPRLSLSFKLKRNI
GYFILQTYMPSILITILSWVSFWINYDASAARVALGITTVLTMTTINTHLRETLPKIPYVKAID
MYLMGCFVFVFMALLEYALVNYIFFGRGPQRQKKAAEKAASANNEKMRLDVNKIFYKDIK
QNGTQYRSLWDPTGNLSPTRRTTNYDFSLYTMDPHENILLSTLEIKNEMATSEAVMGLGDP
RSTMLAYDASSIQYRKAGLPRHSFGRNALERHVAQKKSRLRRRASQLKITIPDLTDVNAIDR
WSRIFFPVVFSFFNIVYWLYYVN

Figure 2. Sample input sequence

Table 1. Frequency occurrence of various triplets in protein sequences

Protein	Amino acid composition (%)		
	Trp (W)	Tyr (Y)	Cys (C)
P16615	1.5%	1.8%	2.8%
Q9NRI5	1.2%	0.8%	2.6%
P42263	1.6%	4.6%	1.1%
P43005	0.8%	2.1%	1.1%
P47870	2.0%	5.7%	0.8%
P35462	1.2%	4.2%	4.0%

The experiments are sorted out the 100 protein sequences of the protein with ID: P12883 (length: 1935 amino acids). This was based on the highest frequency occurrences of triplets and with the availability of sufficient natural variants.

Table 2. Frequency and composition level of P12883

protein ID	P12883		
Total Length	1935		
Amino Acid Name	**Single Code Representation**	**Frequency**	**Composition %**
Glycine (Gly)	G	66	3.4
Alanine (Ala)	A	147	7.6
Leucine (Leu)	L	180	9.3
Methionine (Met)	M	51	2.6
Phenylalanine(Phe)	F	53	2.7
Tryptophan (Trp)	W	10	0.5
Lysine (Lys)	K	153	7.9
Glutamine (Gln)	Q	109	5.6
Glutaminc Acid (Glu)	E	171	8.8
Serine (Ser)	S	82	4.2
Proline (Pro)	P	29	1.5
Valine (Val)	V	75	3.9
Isoleucine (Ile)	I	77	4.0
Cysteine (Cys)	C	14	0.7

Tyrosine (Tyr)	Y	40	2.1
Histidine (His)	H	32	1.7
Arginine (Arg)	R	105	5.4
Asparagine (Asn)	N	86	4.4
Asparatic Acid (Asp)	D	88	4.5
Threonine (Thr)	T	83	4.3

Table 2 provides the output received from preprocessing with a single code representation and also gives the frequency of amino acids with composition percentage. Figure 3 shows the same in pictorial representation. Based on ITSRepeats, the SRR was identified with its frequency based on the composition as shown in Table 2.

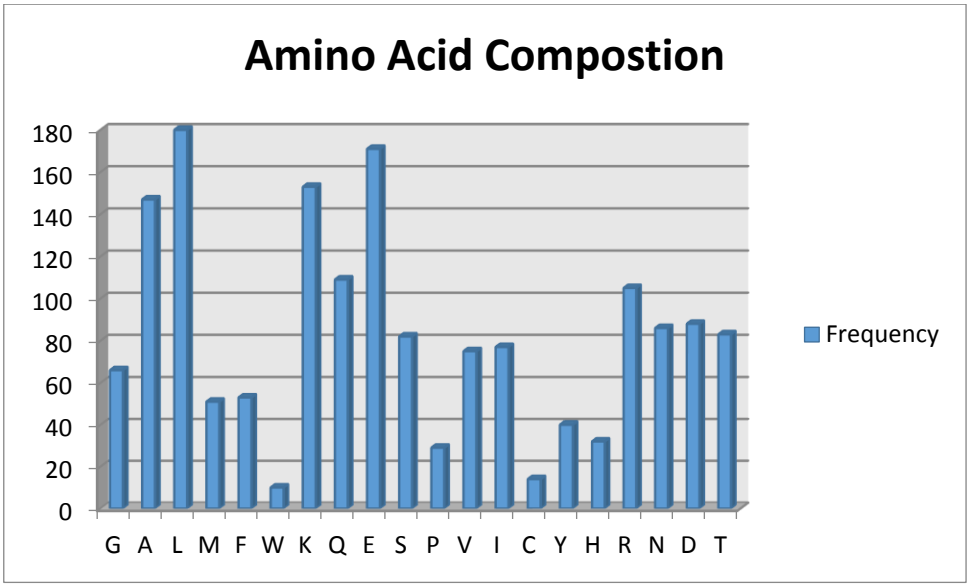

Figure 3. Composition of Amino acids for P12883

Table 3. Frequency and composition level of P12883 based on SRR

protein ID	P12883	
Total Length	1935	
Single Residue Repeats	**Frequency**	**Composition**
AA	5	0.26
AAA	2	0.10
AAAA	1	0.05
EE	34	1.76
EEE	5	0.26
KK	18	0.93
KKK	4	0.21
LL	17	0.88
DD	8	0.41
QQ	7	0.36
SS	7	0.36
II	6	0.31
RR	5	0.26
VV	4	0.21

GG	3	0.16
PP	2	0.10
NN	2	0.10
FF	2	0.10
MM	1	0.05
HH	1	0.05
TT	1	0.05

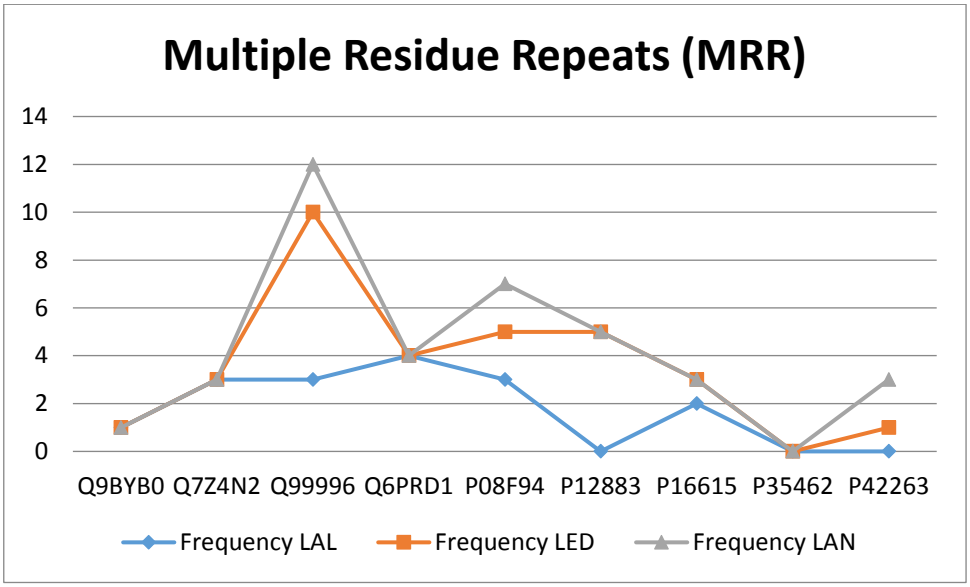

Figure 4. Frequency of amino acids based on MRR for triplets

After executing the SRR, the next level of similarity composition was identified using a hamming distance similarity measure. The frequency based on MRR for triplets was derived and the frequency of triplet occurrence was identified as shown in Figure 4. Based on the occurrences of triplets, the amino acid which causes mental disorder was recognized by applying any classification algorithm. Table 4 represents the protein with its amino acid which causes mental disorder with its composition ratio.

Table 4. Recognition of amino acid caused mental disorder

*Protein ID	Amino Acid Sequence Total Length	Frequency					Mental Disorder Caused Amino Acid Composition Ratio	Essential Amino Acid	Non-Essential Amino Acid
		W	Y	C	M	F			
Q9BYB0	1731	9	25	8	14	48	6.0	20.2	79.8
Q7Z4N2	1603	22	53	21	48	49	12.0	39.3	60.7
Q99996	3907	14	57	38	89	85	7.2	37.1	62.9
Q6PRD1	2367	46	30	56	37	45	9.0	29.1	70.9
P08F94	4074	64	95	96	68	148	11.6	39.1	60.9
P12883	1935	10	40	14	51	53	8.7	36.9	63.1
P16615	1042	14	17	29	32	41	12.8	42.9	57.1
P35462	400	5	13	14	6	15	13.3	40.0	60.0
P42263	894	10	41	10	26	43	14.5	39.1	60.9

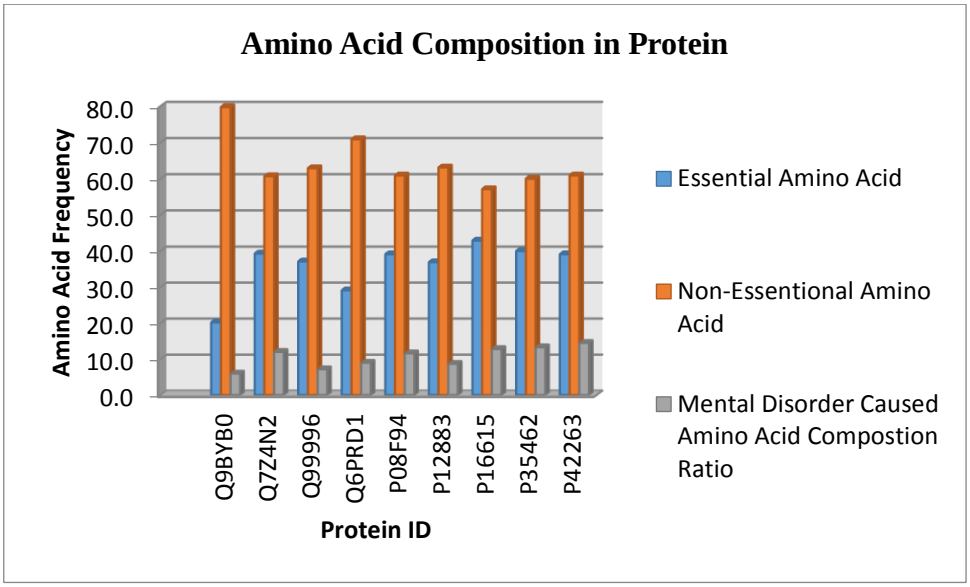

Figure 5. Amino acids are implicated in mental disorder / identification of amino acid caused mental disorder

For finding similarities in the amino acid sequences of the two proteins using a computer adaptable method was developed and found it can possibly determine the important similarity between the proteins [27]. An enhancement in identifying the common molecular subsequences by finding a pair of segments one from each other with long sequence elements thus having no other segment pairs with high similarities. In this study, the similarity measure used allows for insertion and deletion of arbitrary length [28]. Markov models provide a theoretical basis for a wide range of applications and have proven to be effective in practice for machine recognition of speech and other important applications. The theoretical aspects of this type of modeling and illustrate its application in selected problems in machine recognition of speech were discussed [29]. The methods for analyzing regularly spaced patterns of amino acids in proteins, applied to the α1 chain of collagen. Fourier methods and a pattern detection method for general shape are used. The study finds statistically significant periodicities in the distribution of hydrophobic and charged groups in collagen, with a possible formation of salt bridges across the fibril. It also finds no sign of any regular pattern of amino acids over the triple helix [30]. An automated algorithm that can identify and delimit protein sequence fragments that display similarity. The algorithm is able to tolerate mutations, insertions and deletions in the repeats and the sequence spans between repeats can be of different lengths [31]. The purpose of the study is to highlight the majority of applications of neural computing in pattern recognition and the importance of using a principled approach for successful implementation [32]. The Long Short-Term Memory (LSTM) is a fast model-based recurrent neural network for protein homology detection. It extracts indicative patterns for the positive class and uses correlations between all detected patterns for classification [33]. To identify alpha-rod repeats in approximately 0.4% of proteins in eukaryotic genomes and demonstrate its utility in directing experimental work to demarcate three alpha-rods in huntingtin, a protein mutated in Huntington's disease [34]. The development of a new tool, based on the SIMPLE algorithm that facilitates the quantification of the amount of simple sequence in proteins and determines the type of short motifs that show clustering above a certain threshold. The tool can be modified to study simple sequence content at various levels, from highly organized tandem structures to complex combinations of repeats [35]. The problem of identifying low-complexity regions (LCRs) in a protein sequence using new complexity measures based on a given scoring matrix, such as BLOSUM 62, which considers the order of amino acids in the sequence and the sequence length. Thus developed a novel graph-based algorithm called GBA for identification of LCRs in a protein sequence [36]. Methods for detecting errors in circuits, caused by faulty contacts, open circuits, or external disturbances, have been an ongoing focus in the field. Attention

is paid to identifying and addressing these errors to ensure the proper functioning of the circuit [37]. The spectrum of mental health disorders includes mood disorders, anxiety disorders such as post-traumatic stress disorder (PTSD), panic disorders, eating disorders, attention deficit disorder/attention deficit hyperactivity disorder (ADD/ ADHD), and autism. Major depressions, bipolar disorder, schizophrenia, and obsessive-compulsive disorders (OCDs) are the four mental illnesses most often associated with disabilities [38]. Amino acids are organic compounds that comprise amino groups, carboxylic acids and unique carbon structures. Humans use 21 different amino acids; nine of them are essential (histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, valine [39]. Alanine, arginine, asparagine, aspartate, glutamate, glutamine, glycine, proline and serine are synthesized from glucose, whereas phenylalanine and methionine produce tyrosine and cysteine, respectively, from their metabolism [40]. Branch chain amino acids (leucine, isoleucine, valine) play an important role in protein synthesis and turnover in the human body and comprise 20% of amino acids in muscle proteins. Branched chained amino acids (BCAA) catabolism is the cause of Maple Syrup Urine Disease (MSUD) [41]. The protein translation process begins with the amino acid methionine, which is the most common starting codon. In addition to tyrosinemia type I (associated with liver disease) [42]. Serotonin, dopamine, noradrenaline, and GABA deficiencies have been linked to depression for some time now [43-46]. Researchers have reported that the amino acids tryptophan, tyrosine, phenylalanine, and methionine help treat depression and several mood disorders [47-50]. Patients with major depression may also suffer from bipolar disorders, which include manic-depressive episodes, uncontrollable mania, hypomania, and mixed episodes (manic and depressive episodes) [51]. Anemia fatty acid deficiencies and omega-3 fatty acid deficiencies are among the biochemical abnormalities observed in people with bipolar disorder. A calming effect is known to be induced when taurine is ingested by the brain as an amino acid made in the liver from cysteine. Reduced red-cell folate levels were found in hospitalized patients with DSM-III-R diagnosis of mania, compared to the control group. The manic patients and controls were matched by socio-economic status, indicating that the reduced red-cell folate is associated with the illness and not due to reduced absorption or dietary deficiency of folate. Findings suggest that reduced red-cell folate occurs in both phases of bipolar disorders [52]. Bipolar patients can experience manic episodes if they are deficient in this amino acid. The central nervous system of schizophrenic patients has been found to produce less serotonin [53]. Tryptophan levels are lower in patients with schizophrenia based on plasma amino acid concentrations [54-55]. As part of this research, we also examined the role of the serotonergic system and the relationship between it and the NMDA receptor by determining the central availability of the precursor amino acid for 5-HT, tryptophan, and other large neutral amino acids, e.g., tyrosine, phenylalanine, valine, isoleucine, and leucine. Glutamate is one of the most abundant neurotransmitters within our brain, contributing to synaptic plasticity, learning, memory, and modulating limbic system function [56]. Serotonin deficiency has been linked to anxiety disorders such as generalized anxiety, panic disorder, specific phobias, and obsessive-compulsive disorder [57]. Several serotonergic system changes have been associated with anxiety, including polymorphisms in the serotonin transporter gene (SERT) [58].

4. Conclusions

It is concluded that a python based algorithm, ITRRepeats with intelligent rules have been developed to find out the single and triplet amino acids repeat counts such as LAL, LED, LAN and GAS in proteins. Further, it is found that no previous research study is available on these triplet repeats. Hence, it is obvious that the present research is the first contribution to these triplets which also considered hamming distance similarity measure for increasing the efficiency in analyzing the patterns between the sequences of triplets. This may play a major role in bipolar disorder. The amino-terminal cap residues are also found to contain a higher amount of natural variants, which have been linked to both increased and decreased stability of proteins. The hydrophobicity affects the flexibility, and the charged property may be responsible for altered function during intra- and inter-molecular interactions in neurological related proteins. Also, E → V, Y →C increases the stability and I → T decreases the stability. It is also concluded

that hydrophobic and charged duplets repeats might be responsible for bipolar related neurological disorders and that the said web server MPaaRep will be very useful for the users to identify the repeats with ease and quickly. The new triplets such as LAL, LED, LAN and GAS identification would give certainly new insights while designing novel proteins and drug discovery.

Symbols and abbreviations

SRR – Single Residue Repeats; MRR – Multiple Residue Repeats; SL – Sequence Length; ΔΔG – free energy change in kcal/mole; pH - Power of Hydrogen; T – Temperature; A – Alanine; R – Arginine; N – Asparagine; D - Aspartic acid; C – Cysteine; Q – Glutamine; E – Glutamic acid; G – Glycine; H – Histidine; I – Isoleucine; L – Leucine; K – Lysine; M – Methionine; F – Phenylalanine; P – Proline; S – Serine; T – Threonine; W – Tryptophan; Y – Tyrosine; V – Valine

References

1.LANDER, ERIC S., "Initial sequencing and analysis of the human genome." (2001).
2.Richards RI, Sutherland GR. 1992, Dynamic mutations: a new class ofmutations causing human disease. Cell 70:709-712.
3.FONDON, J. W., III et al. (2008), Simple sequence repeats: genetic modulators of brain function and behavior. Trends Neurosci. 31, 328-334.
4.KASHI, Y., KING, D.G. (2006), Simple sequence repeats as advantageous mutators in evolution. Trends Genet. 22, 253259.
5.VINCES, M.D. et al. (2009), Unstable tandem repeats in promoters confertranscriptional evolvability. Science 324, 1213-1216)
6.GREEN, HOWARD, NORMAN WANG. Codon reiteration and the evolution of proteins " *Proceedings of the National Academy of Sciences* 91.10 (1994), 4298-4302.
7.VAN DELLEN A, GROTE HE, HANNAN AJ. 2005. Gene-environment interactions, neuronal dysfunction and pathological plasticity in Huntington'sdisease. ClinExpPharmacolPhysiol 32:1007-1019.
8.GATCHEL JR, ZOGHBI HY. 2005. Diseases of unstable repeat expansion:mechanisms and common principles. Nat Rev Genet 6:743-755.
9.VAN DELLEN A, GROTE HE, HANNAN AJ. 2005. Gene-environment interactions, neuronal dysfunction and pathological plasticity in Huntington'sdisease. ClinExpPharmacolPhysiol 32:1007-1019)
10.DAVIES SW, TURMAINE M, COZENS BA, DIFIGLIA M, SHARP AH., et. al. 1997. Formation of neuronal intranuclear inclusions underlies theneurological dysfunction in mice transgenic for the HD mutation. Cell 90:537-548
11.ROSS CA. 1997. Intranuclear neuronal inclusions: a common pathogenic mechanism for glutamine-repeat neurodegenerative diseases? Neuron19:1147-1150
12.BROWN LY, BROWN SA. 2004, Alanine tracts: the expanding story of human illness and trinucleotide repeats. Trends Genet 20:51-58).
13.CALADO A, TOME FM, BRAIS B, ROULEAU GA, KUHN U, et al. 2000, Nuclear inclusions in oculopharyngeal muscular dystrophy consist of poly(A) binding protein 2 aggregates which sequester poly(A) RNA. Hum MolGenet 9:2321-2328
14.FAN X, DION P, LAGANIERE J, BRAIS B, ROULEAU GA. 2001, Oligomerization of polyalanine expanded PABPN1 facilitates nuclear protein aggregation that is associated with cell death. Hum Mol Genet 10:2341-2351
15.DORSMAN JC, PEPERS B, LANGENBERG D, KERKDIJK H, IJSZENGA M., et al.2002, Strong aggregation and increased toxicity of polyleucine over polyglutamine stretches in mammalian cells. Hum Mol Genet 11:1487-1496
16.BEAR MF, HUBER KM, WARREN ST. 2004, The mGluR theory of fragile X mental retardation. Trends Neurosci 27:370-377
17.HUNTLEY, MELANIE A., G. BRIAN GOLDING. "Neurological proteins are not enriched for repetitive sequences." *Genetics* 166.3 (2004): 1141-1154.

18.PERUTZ, M. F., et al. "Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid β-peptide of amyloid plaques." *Proceedings of the National Academy of Sciences* 99.8 (2002), 5596-5600.
19.HUNTLEY, MELANIE, G. BRIAN GOLDING, "Evolution of simple sequence in proteins." *Journal of molecular evolution* 51.2 (2000): 131-140).
20.PERUTZ, MAX F., A. H. WINDLE, "Cause of neural death in neurodegenerative diseases attributable to expansion of glutamine repeats." *Nature* 412.6843 (2001), 143-144
21.KARLIN, SAMUEL, CHRIS BURGE, "Trinucleotide repeats and long homopeptides in genes and proteins associated with nervous system disease and development." *Proceedings of the National Academy of Sciences* 93.4 (1996), 1560-1565.
22.LI, SHI-HUA, et al. "Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms." *Genomics* 16.3 (1993), 572-579
23.KOIDE, R., et al. "Unstable expansion of CAG repeat in hereditary dentatorubral–pallidoluysian atrophy (DRPLA)." *Nature genetics* 6.1 (1994), 9-13.
24.SIMON, MICHELLE, JOHN M. HANCOCK, "Tandem and cryptic amino acid repeats accumulate in disordered regions of proteins." *Genome Biology* 10.6 (2009), 1-16
25.DUNKER, A. KEITH, et al. "Intrinsic disorder and protein function." *Biochemistry* 41.21 (2002), 6573-6582.
26.ORR, HARRY T., HUDA Y. ZOGHBI, "Trinucleotide repeat disorders." *Annu. Rev. Neurosci.* 30 (2007), 575-621.
27.NEEDLEMAN, SAUL B., CHRISTIAN D. WUNSCH, "A general method applicable to the search for similarities in the amino acid sequence of two proteins." *Journal of molecular biology* 48.3 (1970), 443-453.
28.SMITH, TEMPLE F., MICHAEL S. WATERMAN, "Identification of common molecular subsequences." *Journal of molecular biology* 147.1 (1981), 195-197.
29.RABINER, LAWRENCE R., "A tutorial on hidden Markov models and selected applications in speech recognition." *Proceedings of the IEEE* 77.2 (1989), 257-286.
30.MCLACHLAN, A. D., Analysis of periodic patterns in amino acid sequences:colagen. *Biopolymers: Original Research on Biomolecules* 16.6 (1977), 1271-1297
31.HERINGA, JAAP, PATRICK ARGOS, "A method to recognize distant repeats in protein sequences." *Proteins: Structure, Function, and Bioinformatics* 17.4 (1993), 391-411
32.BISHOP, CHRISTOPHER M., *Neural networks for pattern recognition*. Oxford university press, 1995.
33.HOCHREITER, SEPP, MARTIN HEUSEL, KLAUS OBERMAYER, "Fast model-based protein homology detection without alignment." *Bioinformatics* 23.14 (2007), 1728-1736.
34.PALIDWOR, GARETH A., et al., "Detection of alpha-rod protein repeats using a neural network and application to huntingtin." *PLoS computational biology* 5.3 (2009), e1000304.
35.Alba MM, Laskowski RA, Hancock JM., Detecting cryptically simple protein sequences using the SIMPLE algorithm. Bioinfrormatics, 2002, 18: 672-8
36.LI X, KAHVECI T., A Novel algorithm for identifying low-complexity regions in a protein sequence. Bioinformatics 2006, 22:2980-7
37.HAMMING RW., Error detecting and error correcting codes. Bell SystTechJ, 1950:270.
38.MURRAY CJL, LOPEZ AD., The Global Burden of Disease. *WorldHealth Organization*1996:270. American Psychiatric A: Diagnostic and Statistical Manual of Mental Disorders. *Fourth edition, text revision Washington DC* 2000).
39.ALIU, ERMAL, SHIBANI KANUNGO, GEORGIANNE L. ARNOLD, "Amino acid disorders." *Annals of translational medicine* 6.24 (2018)
40.LIEBERMAN, MICHAEL, ALLAN D. MARKS, *Marks' basic medical biochemistry: a clinical approach*. Lippincott Williams & Wilkins, 2009

41.CHUANG DT, SHIH VE, MAX WYNN R., Maple Syrup Urine Disease (Branched-Chain Ketoaciduria). In: Beaudet AL, Vogelstein B, Gibson KM, et al. editors.
42.RUSH AJ., The varied clinical presentations of major depressivedisorder.*The Journal of clinical psychiatry* 2007, 68(8 Suppl):4-10.
43.***National Institute of Mental Health: Depression. *National Institute ofMental Health, National Institutes of Health* 2000, US Department of Health and Human Services, Bethesda (MD) [Reprinted September 2002.
44.RUSH AJ., The varied clinical presentations of major depressivedisorder.*The Journal of clinical psychiatry* 2007, 68(8 Suppl):4-10.
45.STOCKMEIER CA., Neurobiology of serotonin in depression andsuicide.*Ann N Y AcadSci* 1997, 836:220-232.
46.VANPRAAG HM., Depression, suicide and the metabolism of serotonin in the brain. *J Affect Disord* 1982, 4(4):275-290 .
47.LEONARD BE., The role of noradrenaline in depression: a review. J Psychopharmacol, 1997, 11(4 Suppl):S39-S47) .
48.PETTY F., GABA and mood disorders: a brief review andhypothesis. *J Affect Disord* 1995, 34(4):275-281.
49.MCLEAN A, RUBINSZTEIN JS, ROBBINS TW, SAHAKIAN BJ., The effects oftyrosine depletion in normal healthy volunteers: implications for unipolar depression. Psychopharmacology, 2004, 171(3), 286-297
50.AGNOLI A, ANDREOLI V, CASACCHIA M, CERBO R., Effect of s-adenosyll-methionine (SAMe) upon depressive symptoms.*J PsychiatrRes* 1976, 13(1):43-54.
51.RIHMER Z, GONDA X, RIHMER A., Creativity and mental illness.*Psychiatr Hung* 2006, 21(4):288-294.
52.HASANAH CI, KHAN UA, MUSALMAH M, RAZALI SM., Reduced red-cellfolate in mania.*J Affect Disord* 1997, 46:95-99.
53.Van DER HEIJDEN DFMMA, TUINIER, S., SIJBEN AES, KAHN R.S., VERHOEVENWMA, Amino acids in schizophrenia: evidence for lower tryptophan availability during treatment with atypical antipsychotics? Journal of Neural Transmission, 2005, 112(4), 577-585
54.RAO, M.L., GROSS, G., STREBEL, B., BRAUNIG, P., HUBER, G., KLOSTERKÖTTER, J., 1990, Serum amino acids, central monoamines, andhormones in drug-naive, drug-free, and neuroleptic-treated schizophrenic patients and healthy subjects. Psychiatry Res., 34, 243 – 257.
55.TORTORELLA, A., MONTELEONE, P., FABRAZZO, M., VIGGIANO, A., DE LUCA, L., MAJ, M., 2001. Plasma concentrations of amino acids in chronic schizophrenics treated with clozapine. Neuro-psychobiology 44, 167 – 171
56.PRICE DL, BRADY ST, SIEGEL GJ, ALBERS RW., *Basic neurochemistry: principlesof molecular, cellular, and medical neurobiology*, Amsterdam: Elsevier Ltd (2012).
57.RESSLER KJ, NEMEROFF CB., Role of serotonergic and noradrenergic systems in thepathophysiology of depression and anxiety disorders. *Depress Anxiety* (2000), 12Suppl 1:2-19.
doi: 10.1002/1520-6394(2000)12:1+<2::AID-DA2>3.3.CO;2-W).
58.HOMBERG JR, LESCH KP., Looking on the bright side of serotonin transportergene variation. *Biol Psychiatry* (2011), 69(6):513-9. doi: 10.1016/j. biopsych.2010.09.024).

Manuscript received: 14.02.2022