iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VI **Month of publication:** June 2020

DOI: http://doi.org/10.22214/ijraset.2020.6112

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

Application of Visual Basic in Bioinformatics for the Structural Prediction of Protein

Aruna Aasirvatham
Bharathiar University

Abstract: ***Protein secondary structure prediction is a problem of central importance in predictive structural biology. It has been studied intensively for more than 30 years. However, the best models currently available to perform this task suffer from a fundamental drawback their prediction is only based on the content of a small segment of the sequence surrounding the residue the conformational state of which is to be determined when compared with the database searches it is good to use programming language (Visual Basic) The choufasman algorithm developed using visual basic for the secondary structure of protein plays a significant role. The protein structure prediction using choufasman in visual basic showed prediction accuracy of 56 to 60%.In these studies the protein catalase was selected and using the visual basic programming language the amino acid sequence for protein catalase it's secondary structure and choufasman algorithm with score was also analyzed.***
Keywords: Catalase, secondary structure prediction, visual basic, chou Fasman algorithm, SOPMA..

I. INTRODUCTION

A. What is Bioinformatics?

Bioinformatics is a field that develops database, methods and software tools for studying biological data, when the data sets are large and complex Bioinformatics involves the integration of computers, software tools, and databases. Bioinformatics approaches are often used for major initiatives that generate large data sets. The genomics and proteomics are the two important collection, of large-scale activities that use bioinformatics. Classification, storage, and analysis of biochemical and biological information using computers especially as applied to molecular genetics and genomics. Bioinformatics has become an important part of many areas of biology. It plays an important role in structural prediction of protein, genome analysis, DNA and RNA binding region, Homology modeling, and identification of drug compounds. It plays a role in the analysis of gene and protein expression and regulation.

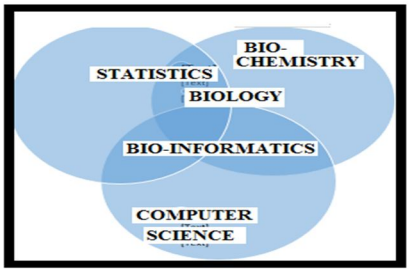


Figure1 Bio-informatics overview

B. Visual Basic

Visual Basic (VB) is a third generation event-driven programming language first released by Microsoft in 1991.Microsoft that provides a graphical user interface (GUI) which allows programmers to modify code by simply dragging and dropping objects and defining their behavior and appearance. Visual Basic can create executable (EXE files), ActiveX controls, or DLL files, but is primarily used to develop Windows applications and to interface database systems. Dialog boxes with less functionality can be used to provide pop-up capabilities. NET (VB.NET) is a general-purpose programming language that works in the NET environment.

Figure 2 Visual Basic

C. Structural Prediction Of Protein

Proteins are polymers of amino acids covalently linked through peptide bonds into a chain. They constitute a group of highly complex organic compounds found in all living cells and comprising the most abundant class of all biological molecules. Protein comprises approximately 50% of cellular dry weight. Hundreds of protein molecules have been isolated in pure, homogeneous form; many have been crystallized.

All contain carbon, hydrogen, and Oxygen, and nearly all contain sulfur as well. Some proteins also incorporate phosphorous, ion, zinc, and copper.

Proteins are large molecules with high molecular weights from about 10,000 for small ones of 50-100 amino acids to more than 1,000,000 for certain forms.

They are composed of varying amounts of the same 20 acids, which in the intact protein are united through covalent chemical linkages called peptide bonds.

Within and outside of cells, proteins serve a myriad of functions, including structural roles (cytoskeleton), as catalysts (enzymes transporter to ferry ions and molecules across membranes, and hormones.

A protein molecule that consists of but a single polypeptide chain is said to be monomeric; proteins made up of more than one polypeptide chain, as many of the large ones are called oligomer.

Based upon chemical composition, proteins are divided into two major classes: simple proteins, which are composed of only amino acids, and conjugated proteins which are composed of amino acids and additional organic and inorganic groupings, certain of which are called prosthetic groups. Conjugated proteins include glycoproteins, which contain carbohydrates; lipoproteins, which contain lipids; and nucleoproteins, which contain nucleic acid

Classified by biological function, proteins include the enzyme, which are responsible for catalyzing the thousands of chemical reactions of the living cell; keratin, elastin, and collagen, which are important types of structural, or support, proteins; hemoglobin and other gas transport proteins; ovalbumin, casein, and other nutrient molecules; antibodies, which are molecules of the immune system; protein hormones, which regulate metabolism : and proteins that perform mechanical work, such as actin and myosin, the contractile muscle proteins

D. Classes Of Protein Structure

1) *Class Alpha-* this comprises a bundle of alpha helices connected by loops on the surface of the proteins.
2) *Class Beta-* this comprises antiparallel beta sheets, usually two sheets in close contact forming a sandwich. Alternatively, a sheet can twist in to a barrel with the first and last strands touching. Examples are enzymes, transport proteins, antibodies and virus coat proteins such as neuraminidase.
3) *Class Alpha/Beta-* this comprises mainly parallel beta sheets with intervening alpha helices, but may also have mixed beta sheets. In addition to forming a sheet in some proteins in this class, in others parallel beta strands may form in to a barrel structure that is surrounded by alpha helices. This class of proteins includes many metabolic enzymes.
4) *Class Alpha+Beta-* this comprises mainly segregated alpha helices and antiparallel beta sheets.
5) *Multidomain (Alpha and Beta)-* Proteins comprise domains representing more than one of the above four classes.
6) *Membrane and cell surface-* Proteins and peptides excluding proteins of the immune system comprise this class.

E. Levels Of Protein Structure

Structural features of proteins are usually described at four levels of complexity

1) *Primary Structure:* The linear arrangement of amino acids in a protein and the location of covalent linkages such as disulfide bonds between amino acids.
2) *Secondary Structure:* The areas of folding or coiling within a protein; examples include alpha helices and pleated sheets, which are stabilized by hydrogen bonding.
3) *Tertiary structure:* The final three-dimensional structure of a protein, which results from a large number of non-covalent interaction between amino acids.
4) *Quaternary Structure:* Non-covalent interactions that bind multiple polypeptides into a single, larger protein. Hemoglobin has quaternary structure due to association of two alpha globin and two beta globin polyproteins.

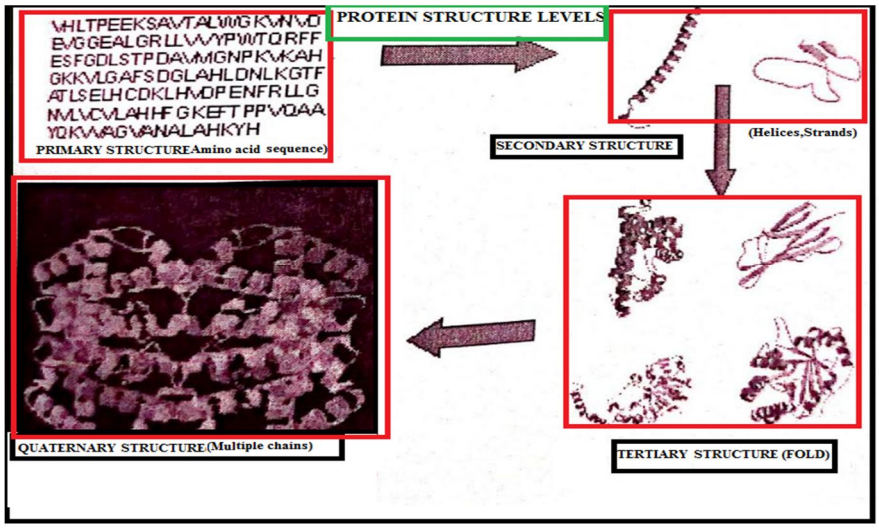


Figure3 protein structure types

II. REVIEW OF PRIMARY STRUCTURE OF PROTEIN AND TERMINOLOGY

Proteins are polymers of amino acids joined together by peptide bonds. The Primary structure of a protein can readily be deduced from the nucleotide sequence of the corresponding messenger RNA. Based on primary structure, many features of secondary structure can be predicted with the aid of computer programs. However, predicting protein tertiary structure remains a very tough problem, although some progress has been made

A. Assembly of Primary Structure

Synthesizing primary structure of a protein is encoded in genetic material of the chromosomes, which is double stranded DNA in most organisms though some viruses use single stranded DNA or RNA. In all cases the information is coded as sequences of 4 kinds of nucleotides on one strand of DNA - A, C, G, and T or of RNA which has the same the nucleotides except U which replaces T. The sequence of other strand of DNA is chemically complimentary to first i.e. A with T of other strand and G with C.

Specific region of DNA or RNA codes for primary structure of every protein synthesized by organism and the DNA, RNA segments on both sides of such regions are involved in regulation and expression of that genetic information, these regions of DNA or RNA are called Genes. Parts of a gene that code for primary structures of protein are well defined but regulatory parts are not. The information in nucleic acid sequence of gene is invariably read in one direction from 5' end to 3 end. So it is possible to describe gene structure in terms of upstream and downstream in coding sequence.

Some genes code only for stable RNA molecules such as ribosomal and transfer RNA. Initial steps in expression of genes are same as in transcription into RNA of genes that code for proteins but resulting RNA is not translated into protein.

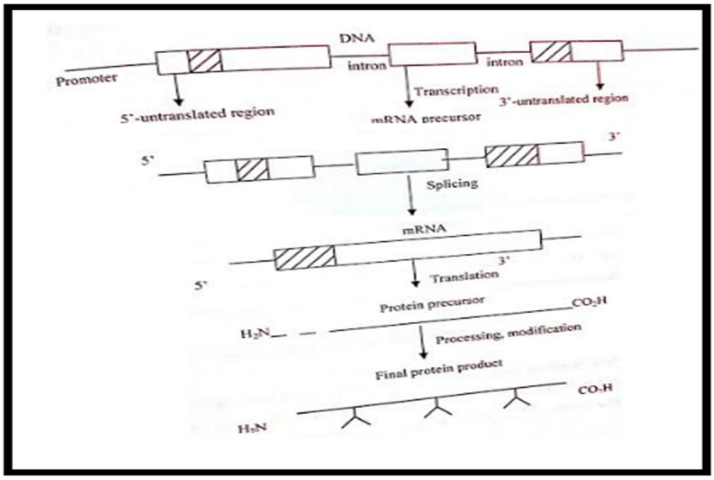
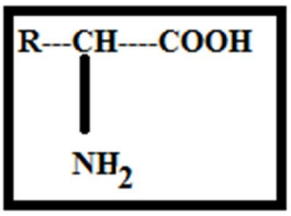


Figure 4 Assembly of protein structure

B. Amino Acids

Amino acids appear in all proteins in all forms of life (and viruses). Amino acids "condense" in the ribosome to form proteins Amino acids are the building blocks of protein. The general structure of an amino acid is shown below.



Each amino acid is a nitrogenous compound consisting of an acidic carboxyl, COOH) group and a basic amino, (- NH2) group. To the alpha carbon, a side chain (often denoted as - R) is attached. R can be simple as a hydrogen atom (H) or a methyl group CH2) or a more complex structure. The side chains vary with each amino acid, and these various side chains confer unique stereo chemical properties on each amino acid.

C. Secondary Structure

Protein structure plays an important role in its function. The protein structure may not be functional if it loses its structure. Primary structure is the amino acid sequence. Secondary structure is local interactions between stretches of a polypeptide chain and includes α-helix and β-pleated sheet structures. Stretches or strands of proteins or peptides have distinct, characteristic local structural conformations, or secondary structure, dependent on hydrogen bonding. The two main types of secondary structure are the α-helix and the ß-sheet. The α-helix is a right-handed coiled strand. In common there are three types of secondary structures in proteins, namely alpha helices, beta sheets, and turns. That which cannot be classified as one of the standard three classes is usually grouped into a category called "other" or "random coil".

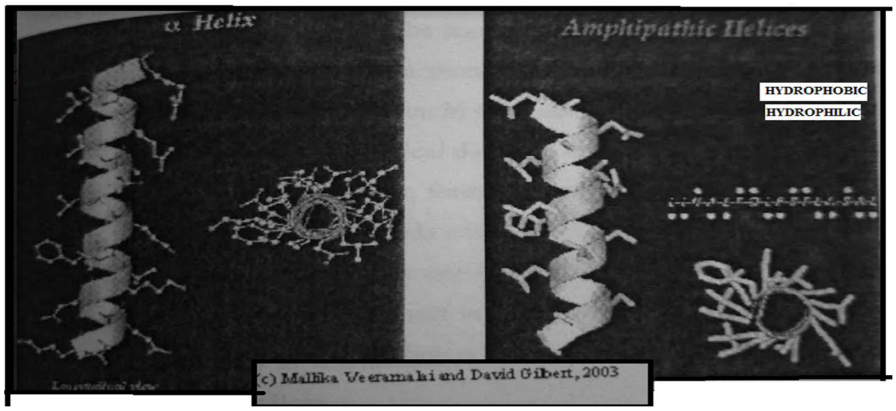


Figure5 Secondary Structure

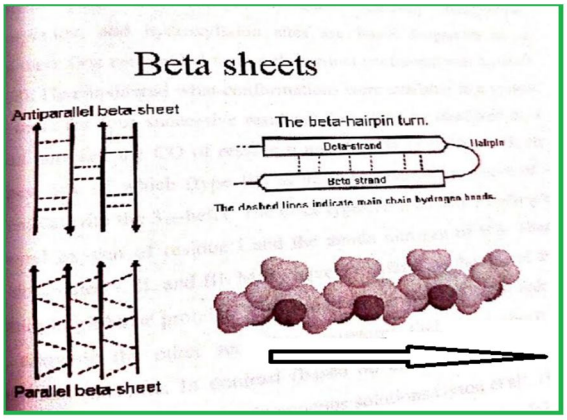


Figure 6 Beta-sheets

D. Tertiary Structure

Protein tertiary structure is the three dimensional shape of a protein. The tertiary structure will have a single polypeptide chain "backbone" with one or more protein secondary structures, the protein domains. Amino acid side chains may interact and bond in a number of ways (Wikipedia)

Tertiary structure is the next level of complexity in protein folding the tertiary structure of proteins deals with how the regional structures are put together in space (www.sciencedirect.com).

E. Quaternary Structure

The quaternary structure of a protein was usually determined by x-ray crystallography .It is the association of several protein chains and subunits into a closely packed arrangement. The protein subunits has its own primary, secondary, and tertiary structure. Protein subunits are held together by hydrogen bonds and van der Waals forces between nonpolar side chains.

III. USING VISUAL BASIC FOR PROTEIN STRUCTURE PREDICTION

A. Program 1(Amino acid Sequence, mRNA sequence prediction)

```
Private Sub MDIForm_Load ()
End Sub
---------------------------------------------------------------------------------------------------------------
Private Sub mnuAmino ClickO
IFlag "MAIN"
FrmAmino.Show
End Sub
---------------------------------------------------------------------------------------------------------------
Private Sub mnuBack_ Click ()
IFlag= "MAIN"
FrmBackProcess.Show
End Sub
Private Sub mnuExit Click ()
End
End Sub
---------------------------------------------------------------------------------------------------------------
Private Sub mnuMRna Click ()
IFlag= "MAIN"
FrmMrna.Show
End Sub
---------------------------------------------------------------------------------------------------------------
Private Sub mnuSec_Click ()
FrmSecondary.Show
End Sub
---------------------------------------------------------------------------------------------------------------
```

B. Program 2 (Secondary Structure Prediction)

```
Dim i as Integer
Dim mRNA as String
---------------------------------------------------------------------------------------------------------------
Private Sub Command1 Click ()
Unload Me
End Sub
---------------------------------------------------------------------------------------------------------------
Private Sub Form_Load ()
Me. Width 3735
```

```
Me.Height 4455
Me.Top=0
Me.Left= 0
I=0
MRna = "UCAG"
Randomize 3
End Sub
Private Sub Form Resize ()
On Error Resume Next
txtMrna.Height = Me.ScaleHeight - 500
TxtMrna. Width = Me.Scale Width
End Sub
---------------------------------------------------------------------------------------------------------------
Private Sub mnuCopy_Click ()
Clipboard.SetText (txtMrna.SelText)
End Sub
Private Sub mnuCut Click ()
Clipboard.SetText (txtMrna.SelText)
txtMrna.SelText = ""
End Sub
---------------------------------------------------------------------------------------------------------------
Private Sub mnuExit Click ()
Unload Me
End Sub
Private Sub MnuFindAmino_Click ()
Dim I as Integer
|= Len (txtMrna.Text)
IFlag = "MRNA"
If1< 100 Then
MsgBox "MRNA less than 100"
Else
FrmAmino.Show
End If
End Sub
---------------------------------------------------------------------------------------------------------------
Private Sub mnuNew _Click ()
0=!
TxtMrna. Text = w
End Sub
---------------------------------------------------------------------------------------------------------------
Private Sub mnuOpen_Click ()
On Error Resume Next
cd1.ShowOpen
txtMrna.LoadFile (cd1.FileName)
MnuFindAmino.Enabled True
End Sub
---------------------------------------------------------------------------------------------------------------
Private Sub mnuPaste_Click ()
txtMrna.SelStart = Clipboard.GetData ()
End Sub
```

```
---
Private Sub mnuRandom_Click ()
Dim I, j, limit As Integer
Dim s as String
TxtMrna. Text =
Limit = Input Box ("Input the limit")
For i = 1 to limit
|= Rnd * 3
I +1 = 1
s = Mid (mRNA, 1, 1)
txtMrna.Text = txtMrna.Text & s
Next
End Sub
---
Private Sub mnuSave Click ()
On Error Resume Next
cd1.ShowSave
txtMrna.SaveFile (cd1.FileName), rtfCFText
End Sub
---
Private Sub txtMrna Key Down (Key Code as Integer, Shift As
If Key Code 65 and Key Code >97 and Key Code 67 and Key code<>99 and
Key Code 85 and Key Code 117 and Key code<>71 And
Key Code 103 and Key Code<> 8 Then
Key Code= 0
End If
End Sub
```

```
Private sub txtMrna_KeyPress (KeyAscii As integer)

End sub
```

```
Private Sub txtMrna_LostFocus ()

End sub
```

C. Program 3(Back Translation Process)

```
Dim rs As Recordset
Dim frame (5) As String
Dim strAmino As String
Public get Amino as String
Dim LOSOURCESTR as String
Dim i as Integer
---
Private Function FindStartCodon (mRNA as String) As String
Dim j, I as Integer
Dim STR, StrStartCodon As String
For j 1 to Len (mRNA)
STR = Mid (mRNA, j, 3)
If str "AUG" Then
```

```
StrStartCodon = str
= True
Exit For
End If
Next j
1 = Len (mRNA) -j-2
If1>= 0 Then
StrStartCodon = StrStartCodon & Mid (mRNA, j+3, 1)
StrStartCodon = StrStartCodon & Mid (mRNA, 1,j-1)
FindStartCodon= StrStartCodon
Else
    MsgBox "No start Codon found"
    Unload Me
End If

End Function
```

```
Private Sub SubFindFilling (By Val SOURCESTR as String, By Val j as Integer, By Val
No as Integer
For i=j to Len (SOURCESTR) Step 3
s= Mid (SOURCESTR, i, 3)
If Len(s) 3 Then
Set rs = db.OpenRecordset ("Select Amino from tblmRNA where Codon=" &s
                If rs.RecordCount > 0 Then
                Select Case rs. Fields (0)
                Case "M"
                 Frame (no) = frame (no) & rs. Fields (0)
                 StrAmino = strAmino & "M"
                  Text show. Text = txtshow.Text & "M"
                  Case "*", "$",
                  Frame (no) = frame (no) & rs. Fields (0)
                  StrAmino = strAmino & rs. Fields (0)
                  Case Else
                  Frame (no) = frame (no) & rs. Fields (0)
                  StrAmino = strAmino & rs. Fields (0)
                  Text show. Text text show.Text & rs. Fields (0)
     End Select
End If

End If
PB.Value = PB.Value + 1
Next i
            Text show.Text text show.Text & Chr (13) & Chr (10)
            StrAmino = strAmino & ";"
End Sub
```

```
Private Sub Filling (SOURCESTR as String)
Dim revstr, desstr As String
Dim i, j, k as Integer
```

```
Dim s as String
For j =0 To 5
Frame (j) =””
Next
strAmino=” "

PB.Value 0
txtshow.Text = “”

For j =1 To 3
Call SubFindFilling (SOURCESTR, j, j - 1)
Next

For j Len (SOURCESTR) To 1 Step -1
    Revstr = revstr & Mid (SOURCESTR, j, 1)
Next j

For j 1 To 3
Call SubFindFilling (revstr, j, j+ 2)

Next j

PB.Visible= False
End Sub
```

```
Private Sub mnuSave_Click ()
cd1.ShowSave
Open cdl.FileName For Output As #2
Print #2, strAmino
Close #2
End Sub

Private Sub mnuSec_Click ()
Get Amino = strAmino
FrmSecondary.Show
End Sub
%3D

Private Sub txtshow_Change ()
End Sub
```

```
Dim frame (5) As String
Dim Amino Acid as String
Dim iApa parti1 As Long
A Count Partl.
A Fcount Partil, A BCount Partil As Long
Dim iFApa_ part2, iBApa part2 As Long
Dim Alpha Range (100, 1) As Long
Dim iBpa_ part! As Long
```

```
Dim B _Count Parti, B_ FCount Partll, B _ BCount Partll as Long
Dim Beta Range (100, 1) As Long
Dim Alpha Helices, Beta Sheets as Long
Dim Alpha Helices no, Beta Sheets no As Integer
Dim Beta Turns As Double
Dim Beta Stop Codon as Integer
Dim From Where As Long
Dim Where Up To As Long
Dim ARange_ Count As Long
Dim BRange_ Count As Long
```

```
Private Sub FindAlphaHelicesPartl ()
Dim IPA value (5), c, i, j, k, pos As Long
Dim countless 100 As Long
Dim rs As Recordset
Dim flag As Boolean
Dim strAnAmino As String
From Where = 1
While from Where < Len (Amino Acid)
StrAnAmino- Mid (Amino Acid, From Where, 1)
Flag True
Set rs = db.OpenRecordset ("Select pa from Amino Acid where Amino acid=" &
strAnAmino & ")
If rs.RecordCount >0 Then
I+ sod = sod
ipa_value (j) = rs. Fields (0)
Count less100 =0
For k 0 To 5
If ipa value (k) <= 100 Or From Where <= 5 Then count_less100 = count l-
If count less100 > 2 Then
Exit For
End If
    Next k
If flag true then
     Call ForwardFindAlphalHelicesPartllI (Alpha_part1, i)
From Where i
Exit For
Else
If> 3 Then
For k 0 To 2
ipa_value (k) = ipa_value (k + 1)
Next k
j=2
End If
End if
J=j+1

End if
    Next i
```

```
J=0
For i=0 To 3
ipa_value (i) = 0
Next i
For i = (Alpha_partl - 6) To WhereUpTo Step -1
StrAnAmino = Mid (Amino Acid, i, 1)
Flag = True
Set rs = db.OpenRecordset ("Select pa from Amino Acid where Amino acid=" &
strAnAmino & "")
If rs.RecordCount > 0 Then
ipa_value (j) = rs. Fields (0)
For k = 0 To 3
If IPA value (k) >= 100 or ipa_value (k) = 0 Then
Flag = False
Exit For
Next k

If flag = True Then
Call BackwardFindAlphaHelicesPartlI1 (Alpha partl, i)
Exit For

Else
If j>=3 Then
For k =0 To 2
ipa_value (k) ipa value (k + 1)
Next k
J=2
End if
        End if
I=j+1
       End if
             Next i
End Sub
```

```
Flag = False
Exit For
End if
Next k
       If flag= True Then
       IApa part1 From Where
       FindAlphaHelicesPartll (From Where)
       Fork 0 To 5
       ipa_value (k) = 0
Next k
j3=1
Else
If j>= 5 Then
For k =0 To 4
ipa_value (k) = ipa_value (k + 1)
Next k
```

```
J=4
End if
End if

J=j+1
J=j+1
From Where = From Where + I
Wend
End Sub
```

```
Private Sub FindAlphalHelicesPartII (Alpha_partl As Long)
Dim ipa_value (3), c, i, j, count As Long
Dim strAnAmino As String
For i= Alpha_part1 +1 to Len (Amino Acid)
StrAnAmino = Mid (Amino Acid, i, 1)
Flag = True
Set rs = db.OpenRecordset ("Select pa from Amino Acid where Amino acid="'&StrAnAmino&"")

If rs.RecordCount>0 Then
ipa_value(j) = rs.Fields(0)
For k =0 To 3
If ipa_value(k) >= 100 Or ipa_value(k) 0 Then
flag = False
```

```
Private Sub ForwardFindAlphalHelicesPartIII (Alpha_partl as Long, By Val up to As Long)
Dim Sigma_Pa, Sigma_Pb, Region_ Count, c as Long
Dim strAnAmino As String
Dim rs As Recordset
Sigma_Pa = 0
Sigma_Pb = 0

On Error GoTo Solve
                For i = Alpha_partI +1 to up to
                STR An Amino = Mid (Amino Acid, i, 1)
Set rs = db.OpenRecordset ("Select pa, pb from Amino Acid where Amino acid=" &
strAnAmino & ")
                Sigma_Pa = Sigma_Pa + rs. Fields (0)
                 Sigma_Pb = Sigma_Pb+ rs. Fields (1)
Next i
Region_Count = up to - Alpha_ part1
If Region_Count >5 and Sigma_Pa > Sigma_Pb Then
Alpha_Helices = Alpha_Helices + Region_Count
Alpha Range (ARange_Count, 1) = up to
End if

 Exit Sub

Solve:
MsgBox "Error in Forward Alpha"
```

```
End Sub
```

```
Private Sub Form _Load0
Me.Left =3800
Me.Top=0
Me.Height=4425
Me.Width 8150
IF i Flag="MAIN" Then
MsgBox "Select File and Open save Amino Sequence"
MnuAmino. Enabled False
Else
MsgBox "Select Find Menu and Amino Acid Sequence"
End if

End Sub
```

```
Private Sub Form _Resize ()
On Error Resume Next
txtshow.Height= Me.ScaleHeight-500
txtshow.Width =Me.Scale Width
End Sub
```

```
Private Sub mnuAmino Click ()
PB.Visible True
MnuSec.Enabled True
MnuBack. Enabled True
LOSOURCESTR FrmMrna.txtMrna.Text
LOSOURCESTR FindStartCodon (LOSOURCESTR)
If LOSOURCESTR "" Then
PB.Max Len (LOSOURCESTR) * 2
Filling (LOSOURCESTR)
End if
End Sub
```

```
Private Sub mnuBack_Click ()
IFlag = "AMINO"
get Amino = ""
Get amino strAmino
FrmPB.Show
FrmBackProcess.Show
End Sub
```

```
Private Sub mnuCopy_Click ()
Clipboard.SetText (txtshow.Sel Text)
End Sub
Clipboard.SetText (Ixtshow.SelText)
```

```
Private Sub mnuCut ClickO
Clipboard.SetText (txtshow.SelText)
```

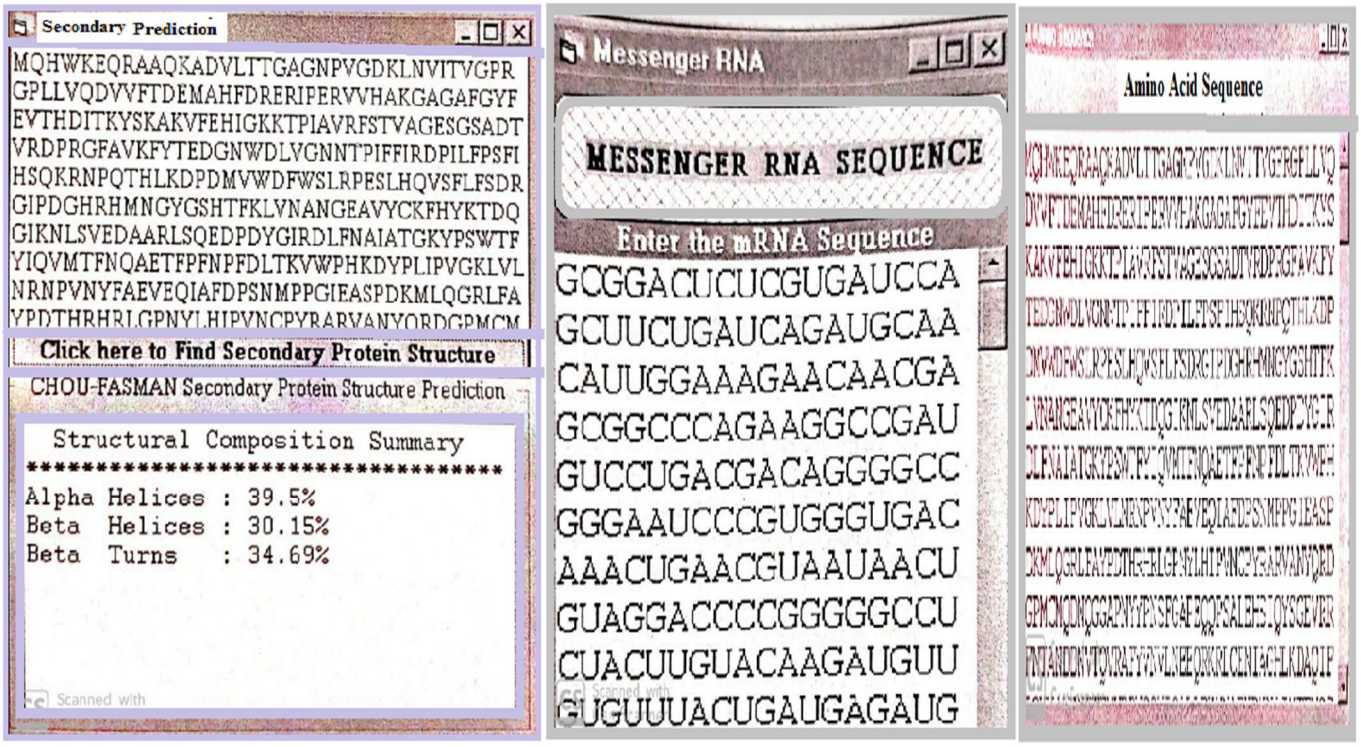
```
Extshow.SelText
End Sub
---------------------------------------------------------------------------------------------------------------------
Private Sub mnuExit Click ()
Unload Me
End Sub
---------------------------------------------------------------------------------------------------------------------
Private Sub mnuOpen Click ()
MnuAmino. Enabled = False
MnuBack. Enabled = True
mnuSec.Enabled = True
Dim LinestrAmino, char Amino as String
Dim no, k as Integer
On Error Resume Next
cd1.ShowOpen
txtshow.Text = ""
txtshow.LoadFile (cd1.FileName)
On Error Resume Next
Open cd1.FileName For Input As #1
Do Until EOF (1)
Line Input #1, LinestrAmino
StrAmino = strAmino & LinestrAmino
StrAmino = strAmino & ":"
Loop
Close #1
End Sub
---------------------------------------------------------------------------------------------------------------------
Private Sub mnuPaste_Click ()
txtshow.SelStart =Clipboard.GetText ()
End Sub
---------------------------------------------------------------------------------------------------------------------
Exit For
End If
Next k
If flag =True Then
Call ForwardFindAlphalHelicesPartllI (Alpha_part1, i)
From Where i
Exit For
Else
If> 3 Then
For k 0 To 2
ipa_value (k) = ipa_value (k + 1)
Next k
j=4
End If
End if
J=j+1
      End if
From Where=From Where+1
For i=0 To 3
```

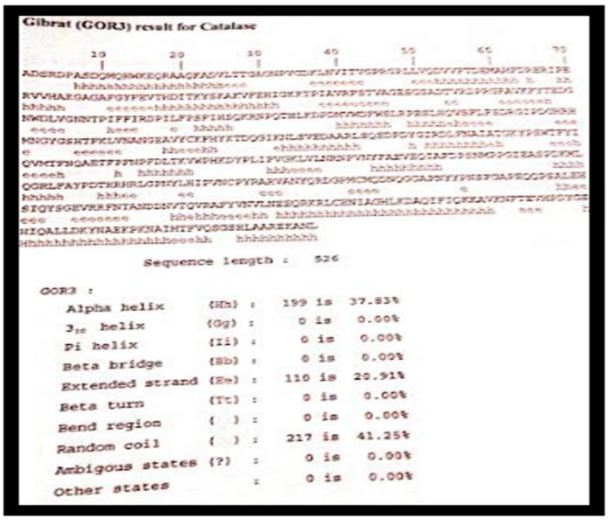
```
End
ipa_value (i) = 0
Next i
For i = (Alpha_partl - 6) To WhereUpTo Step -1
StrAnAmino = Mid (Amino Acid, i, 1)
Flag = True
Set rs = db.OpenRecordset ("Select pa from Amino Acid where Amino acid=" &
strAnAmino & "")
If rs.RecordCount > 0 Then
ipa_value (j) = rs. Fields (0)
For k = 0 To 3
If ipa value (k) >= 100 Or ipa_value (k) = 0 Then
Flag = False
Exit For
Next k
J=0
If flag = True Then
Call BackwardFindAlphaHelicesPartlI1 (Alpha partl, i)
Exit For
```

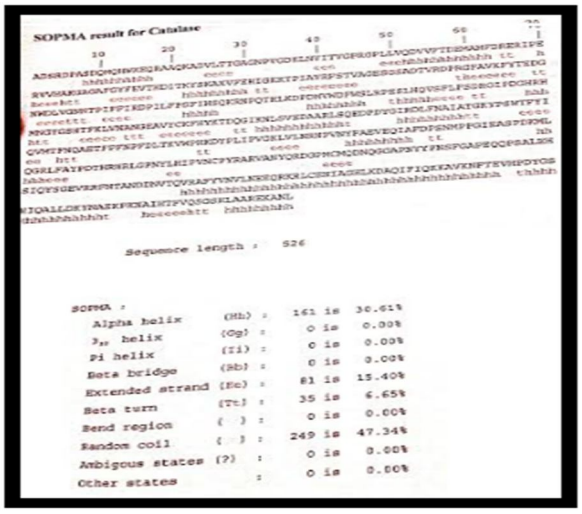
IV. RESULTS AND DISCUSSION.

An amino acid sequence catalase was taken from Swissprot and the E-score value was noted to be negative from BLAST.Then the sequence was saved in visual basic program and noted for the secondary structure prediction of catalase and their back translation.The below result shows the amino acid sequence of catalase,secondary structure prediction alpha,beta, and turn content of catalase.,back translation process and sequence conversion was also found using visual basic computer programming.

Recently a new method called the self-optimized prediction method(SOPMA)has been described to improve the success rate in the prediction of the secondary structure of protein.In this paper we report improvemnts brought about by predicting all the sequence of a set of aligned proteins belonging to the same family.



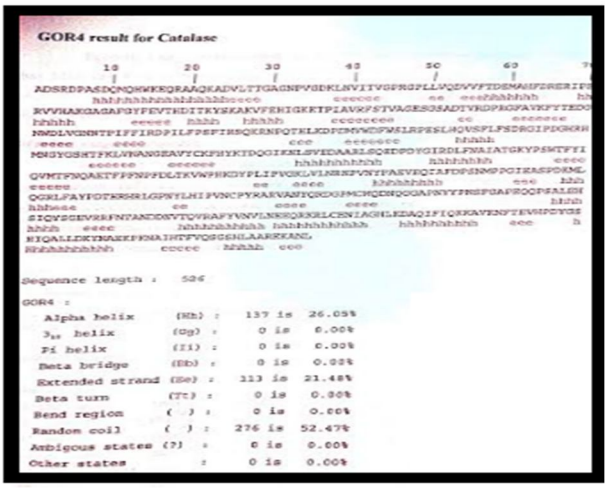


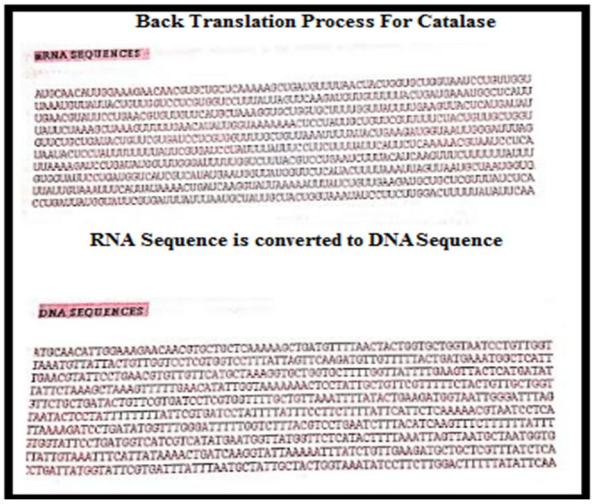


Secondary Structure rates from Chou Fasman method,GOR and SOPMA

	Chou fasman method	GOR Method			SOPMA Method
		GOR1	GOR3	GOR4	
Alpha helices	39.5%	34.60%	34.60%	37.83%	29.66%
Beta sheet	30.15%	0.00%	37.83%	20.91%	16.16%
Beta turn	34.69%	13.69%	13.88%	0.00%	7.60%

The results of secondary structure prediction of proteins from chou fasman algorithm using the visual basic method were compared with the GOR and SOPMA methods to show the percentage similarity between the three methods.



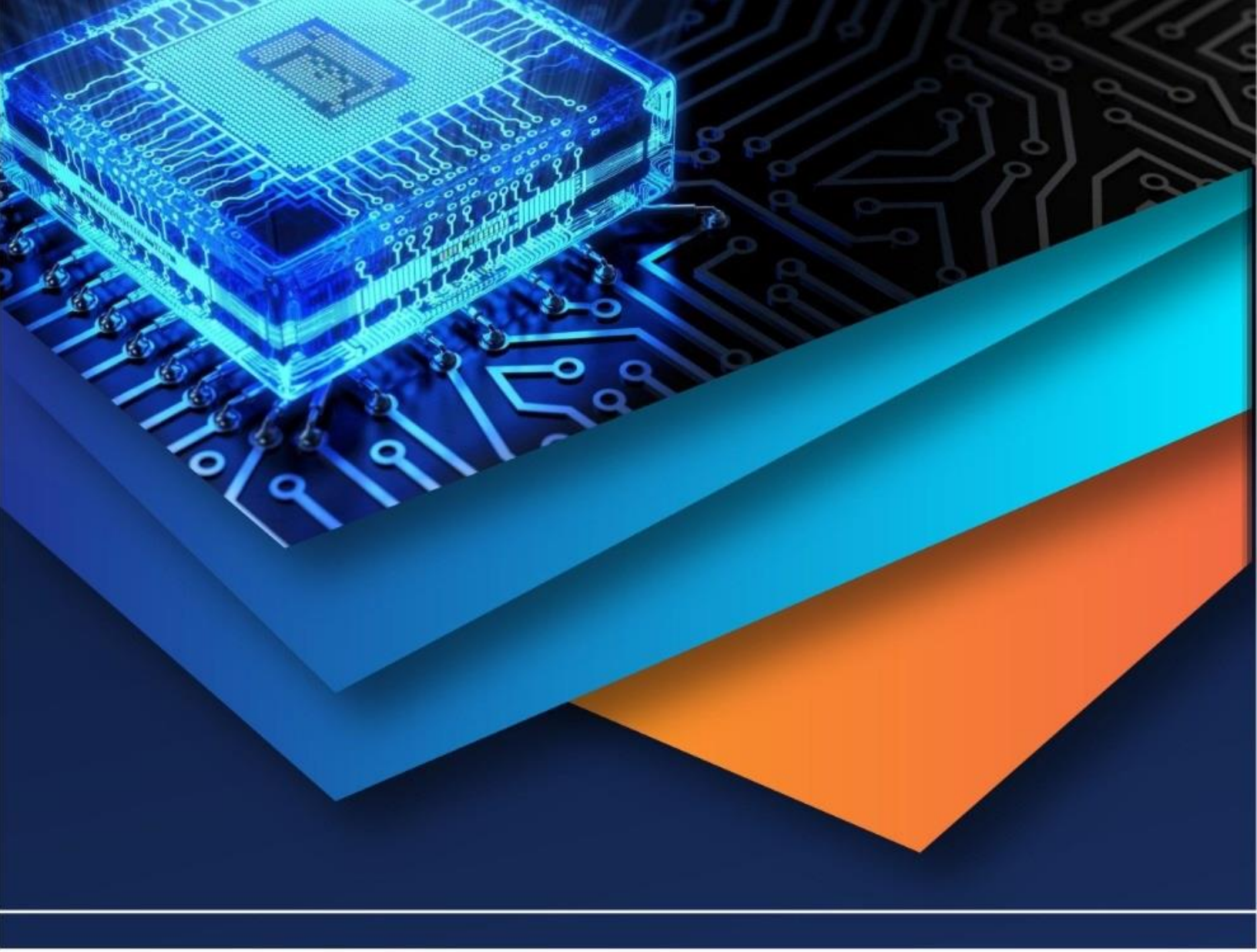


V. CONCLUSION

The secondary structure of protein catalase was predicted using visual basic programming language. The method of using programming language has been improved and gives the same result like database. Its recognition rate is already equivalent to those of the current best prediction methods. Different colors can also be used to differentiate the model. The regions corresponding to the alpha, beta, and turns can be identified on the basis of sequence length. The individual amino acid corresponding to the secondary structure states can be noted.

REFERENCES

[1] www.sciencediect.com
[2] www.wikipedia.com
[3] Atiya khanun, irfan,A. khan, Emerging trends in bioinformatics,185-201.
[4] Benard M.E, Moret, Oliver Gascuel, Algorithm in bioinformatics, 12-25



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY