ALGORITHM DESIGN TO IDENTIFY WHETHER A SEQUENCE (AMINO ACID/NUCLEIC ACID) WILL BIND TO ANOTHER SEQUENCE (PREFERABLY NUCLEIC ACID)

By-

ANAND KUMAR SHARMA (041516)

NIKHIL PARASHARA (041532)

SUNNYL KUMAR (041538)

MAY-2008

Submitted in partial fulfillment of the Degree of Bachelor of Technology

DEPARTMENT OF BIOTECHNOLOGY & BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY

WAKNAGHAT, SOLAN, H.P., INDIA

TABLE OF CONTENTS

CERTIFICATE

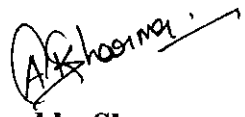
This is to certify that the work entitled, "*Algorithm design to identify whether a sequence (amino acid/nucleic acid) will bind to another sequence (preferably nucleic acid)*" submitted by **Nikhil Parashara** (041532), **Sunnyl Kumar** (041538) and **Anand Kumar Sharma** (041516) in partial fulfillment for the award of Degree Of Bachelor Of Technology in Bioinformatics of Jaypee University of Information Technology has been carried out Under my supervision. This work has not been submitted partially or wholly to any other university or institute for the award of this or any other degree or diploma.

C. D. Poduri

(Project Coordinator)

ACKNOWLEDGMENT

We would like to extend our heartfelt thanks and reverence to our project teacher ***Dr.Chetan Datta Poduri*** for his continuous support in our project. He has been an inspiration for us in being innovative and liberated. He showed us different ways to approach a research problem and the need to be persistent for accomplishing any goal. He has always been there to discuss our ideas to proofread and asked thought provoking questions helping us think through our problems.

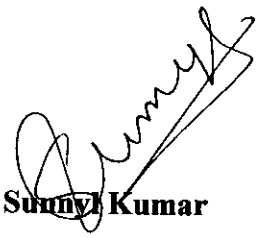


Anand kr.Sharma

(041516)

Nikhil Parashara

(041532)



Sunny Kumar

(041538)

ABSTRACT

In this thesis report, algorithms are presented to determine whether an input sequence (either Nucleic Acid/Protein) bind to another input sequence (be it protein or nucleic acid). In other words, presented here are a set of programs, written in C, C++, to determine whether a pair of input sequences (be it protein or nucleic acid) interact with each other or not.

INTRODUCTION

Interaction between nucleic acids and proteins play a central role in molecular biology. Expression of genetic information is controlled predominantly through the interaction of regulatory proteins with DNA. Numerous protein-DNA complexes are involved in DNA packaging, repair, recombination and replication. Such complexes also perform important functions in RNA splicing, storage and transport.

Proteins and nucleic acids do not operate within the biological system as independent entities. Protein: nucleic acid interactions, (i.e., protein: RNA and protein: DNA interactions), are involved in several processes essential to normal cell function. As with protein: protein interactions, disruption of protein: nucleic acid interactions lead to serious and often catastrophic consequences within the system.

Protein: nucleic acid interactions are integrated into several key cellular processes. These processes include transcription, translation, and regulation of gene expression, recognition, replication, recombination, repair, nucleic acid packaging and the formation of cellular machinery, such as ribosome. The role of DNA as the genetic repository of information requires interaction with proteins for the extraction of this information for timely utilization within the cell.

Protein and nucleic acids are particularly prominent among the molecules essential in life. Their importance stems from the remarkable diversity of their functional roles. This importance stems from the remarkable diversity of their functional roles. This diversity can be illustrated by listing a few of the major groups within each of these molecular families. Proteins are molecules that act to build the structural elements of organisms and to provide the energy necessary for life processes.

Enzymes are proteins that catalyze that degrade foodstuffs to simple, assimilable compounds; the biosynthetic enzymes that build complex molecules from simpler compounds; and muscle proteins that build complex molecules from simpler compounds; and muscle proteins that produce mechanical work from chemical reactions. Transport proteins such as hemoglobin facilitate the movement of molecular oxygen and other essential compounds to their sites of utilization.

Antibodies are proteins that bind to and neutralize foreign materials that may be harmful to an organism. Other proteins are responsible for maintaining the structures of cells, organs, and organisms, while still others play essential roles in genetic expression, nerve organisms, while still others play essential roles in genetic expressions, nerve conduction and all other biological processes.

Nucleic acids are the molecules that carry the information necessary for protein synthesis; they can be considered the 'blueprints' that contain the design of the living organism. In both prokaryotes and eukaryotes, the genetic information of heredity is carried from one generation to the next in DNA. While various types of RNA's play vital roles in the translation of the DNA sequence of each gene into the amino acid sequence of the corresponding protein. The regulation of the expression of different genes, which is vital to the control of development, growth, repair, and reproduction, involves a wide range of interaction between proteins and nucleic acids.

Chapter I

DNA-DNA Interactions

Significance

DNA-DNA interactions, represented by double stranded DNA, are at present believed by the scientific community to be naturally occurring in genomes of all living organisms, except some viruses. In addition, DNA makes an excellent system for benchmarking combinatorial AFM because:

1. the interactions are highly selective
2. the interaction forces may be made arbitrarily strong (by changing the number of base pairs)
3. large libraries of short oligonucleotides are readily available at a low cost
4. It is basically focused on solution conditions like pH, ionic strength, buffer, additive nature and concentration.
5. It also plays important role in recombination of homologous gene.

(A) ALGORITHM FOR DNA-DNA INTERACTION:

```
#include<iostream.h>
#include<stdio.h>
#include<string.h>
#include<conio.h>
#include<process.h>

Void main()
{clrscr();
Char seq1 [100], seq2 [100], bond [100];
int len1,len2,i;
cout<<"Enter Sequence 1\n";
gets(seq1);
len1 = strlen(seq1);
for(i=0;i<len1;i++)
{
if(seq1[i] =='A' || seq1[i]=='T' || seq1[i]=='G' || seq1[i]=='C' || seq1[i]=='N')
{}
else
{
cout<<"\error in sequence";
exit(0);
}}
cout<<"\nEnter Sequence 2\n";
gets(seq2);
len2 = strlen(seq2);
for(i=0;i<len2;i++)
{
if(seq2[i] =='A' || seq2[i]=='T' || seq2[i]=='G' || seq2[i]=='C' || seq2[i]=='N')
{}
else
{
```

```
cout<<"\error in sequence";
exit(0);
}}
i=0;
while(seq1[i]!='\0' && seq2[i]!='\0')
{
if(seq1[i]=='A'&& seq2[i]=='T')
{
bond[i]='|';
}
else if(seq1[i]=='T' && seq2[i]=='A')
{
bond[i]='|';
}
else if(seq1[i]=='G' && seq2[i]=='C')
{
bond[i]= '/';
}
else if(seq1[i]=='C' && seq2[i]=='G')
{
bond[i]='/';
}
else
{
bond[i]=' ';
}
i++;
}
bond[i]='\0';
cout<<"\n";
cout<<seq1;
cout<<"\n"<<bond;
cout<<"\n"<<seq2;
getch();
}
```

Chapter II

RNA-RNA Interactions

Significance

RNA plays several roles in biology:

1. Genetic material (instead of DNA) in some viruses (HIV, influenza, etc.);
2. Protein synthesis: through mRNA, tRNA, rRNA;
3. Regulatory RNA: through "antisense RNA" (snoRNA, snRNA, gRNA, stRNA, micro RNAs) rRNA modification, RNA editing, mRNA splicing, developmental regulation, plasmid copy-number regulation, and gene knock-out.

Bonds in interactions involved:-

internal bonding: base pairings where both bases are on the same RNA,

external bonding: base pairings where bases are on different RNA strands.

(B) ALGORITHM FOR RNA-RNA INTERACTION:

```
#include<iostream.h>
#include<stdio.h>
#include<string.h>
#include<conio.h>
#include<process.h>
void main()
{clrscr();
char seq1[100],seq2[100],bond[100];
int len1,len2,i;
cout<<"Enter Sequence 1\n";
gets(seq1);
len1 = strlen(seq1);
for(i=0;i<len1;i++)
{
if(seq1[i] =='A' || seq1[i]=='U' || seq1[i]=='G' || seq1[i]=='C' || seq1[i]=='N')
{}
else
{
cout<<"\nError in sequence";
exit(0);
}}
cout<<"\nEnter Sequence 2\n";
gets(seq2);
len2 = strlen(seq2);
for(i=0;i<len2;i++)
{
if(seq2[i] =='A' || seq2[i]=='U' || seq2[i]=='G' || seq2[i]=='C' || seq2[i]=='N')
{}
else
{
cout<<"\nError in sequence";
exit(0);
```

```
}}
i=0;
while(seq1[i]!='\0' && seq2[i]!='\0')
{
if(seq1[i]=='A'&& seq2[i]=='U')
{
bond[i]='|';
}
else if(seq1[i]=='U' && seq2[i]=='A')
{
bond[i]='|';
}
else if(seq1[i]=='G' && seq2[i]=='C')
{
bond[i]= '/';
}
else if(seq1[i]=='C' && seq2[i]=='G')
{
bond[i]='/';
}
else
{
bond[i]=' ';
}
i++;
}
bond[i]='\0';
cout<<"\n";
cout<<seq1;
cout<<"\n"<<bond;
cout<<"\n"<<seq2;
getch();
}
```

***Interaction among same pair of Nucleic acids (DNA-DNA & RNA-RNA) are governed by Watson & Crick's finding that proves 3H-bonds can exist b/w G&C & 2H-bonds b/w A&T (in case of DNA) & 2H-bonds b/w A&U (in case of RNA).*

***algorithm for studying DNA/DNA & RNA/RNA interaction has double bond b/w A &T shown by '|' &triple bond existing b/w G & C shown by '\'.*

Chapter III

DNA-Protein Interaction

Significance

The physiological role of DNA-binding proteins is determined by the affinity and specificity of the DNA-protein interaction. These properties depend upon the precise interactions between amino acids in the DNA-binding protein and nucleotides in the DNA-binding site. Protein-protein interactions are required for efficient DNA-protein interactions as well. Gene expression is often regulated by proteins that activate or repress transcription by binding to short, specific DNA sequences. Any genetic approach for characterizing DNA-protein interactions that regulate gene expression requires the isolation of mutants that affect regulation. The DNA-protein interaction can be dissected using in vitro biochemical approaches or using in vivo genetic approaches. Even when DNA-protein interactions have been characterized in vitro, genetic analysis of the interaction is needed to confirm that the specific DNA-protein contacts identified in vitro are necessary and sufficient for DNA-binding under physiological conditions.

(C) ALGORITHM FOR DNA-Protein INTERACTION:

```
#include<iostream.h>
#include<string.h>
#include<fstream.h>
void main()
{
char n[20],p[20];
char a[1000],*nseq,*pseq;
char a1[1000];
 int i;
cout<<"enter nucleotide file name";
cin>>n;
cout<<"enter protein name";
cin>>p;
int length;
length=0;
fstream F1,F2;
F1.open(n,ios::in);
F2.open(p,ios::in);
a1[0]='\0';
while(F1>>a)
{
strncpy(a1,a,(strlen(a)-1));
length=length+(strlen(a)-1);
}
 //cout<<length;
 nseq=new char[length+1];
nseq[0]='\0';
F1.close();
F1.open(n,ios::in);
while(F1>>a)
{
strncpy(a1,a,(strlen(a)-1));
```

```
a1[strlen(a)]='\0';
strcat(nseq,a1);
 }
//cout<<nseq;
 length=0;
 a1[0]='\0';
while(F2>>a)
{
strncpy(a1,a,(strlen(a)-1));
length=length+(strlen(a)-1);
}
 pseq=new char[length+1];
pseq[0]='\0';
F2.close();
F2.open(p,ios::in);
while(F2>>a)
{
strncpy(a1,a,(strlen(a)-1));
a1[strlen(a)]='\0';
strcat(pseq,a1);
 }
//verification of the  HOX  gene interaction
int count,count1;
count=0;count1=0;
int j;
fstream F3;
F3.open("2.txt",ios::in);
char s1[100],s2[100],s3[100];
int number1[100],number2[100];
char **m1,**m2;
int c1,c2,e1,e11,e12,e2,i1,i2;
int q[25];
while(F3>>s1>>s2>>s3)
{
        cout<<s1<<" "<<s2<<" "<<s3<<"\n";
```

```
e1=0;
for( i1=0;i1<strlen(s1);i1++)
{
if(s1[i1] == '|' )
{
e1=e1+1;
}
}
e1=e1+1;
m1=new char*[e1];
e11=0;
e12=0;
        for( i1=0;i1<strlen(s1);i1++)
{
if(s1[i1] == '|' )
{
m1[e12]=new char[e11];
for(i2=0;i2<e11;i2++)
{
m1[e12][i2]=q[i2];
}
m1[e12][e11]='\0';
number1[e12]=e11;
e12=e12+1;
e11=0;
}
else
{
q[e11]=s1[i1];
e11=e11+1;
}
}
m1[e12]=new char[e11];
for(i2=0;i2<e11;i2++)
{
```

```
 m1[e12][i2]=q[i2];
 }
 m1[e12][e11]='\0';
  number1[e12]=e11;
 e12=e12+1;
cout<<"\n########################################\n";
for(i1=0;i1<e1;i1++)
{
for(i2=0;i2<number1[i1];i2++)
{
cout<<m1[i1][i2]<<" ";
}
cout<<"\n";
}
e2=0;
 for( i1=0;i1<strlen(s2);i1++)
 {
 if(s2[i1] == '|' )
 {
 e2=e2+1;
 }
 }
e2=e2+1;
m2=new char*[e2];
 e11=0;
 e12=0;
        for( i1=0;i1<strlen(s2);i1++)
{
if(s2[i1] == '|' )
{
m2[e12]=new char[e11];
for(i2=0;i2<e11;i2++)
{
m2[e12][i2]=q[i2];
}
```

```
m2[e12][e11]='\0';
number2[e12]=e11;
e12=e12+1;
e11=0;
}
else
{
q[e11]=s2[i1];
e11=e11+1;
}
}
m2[e12]=new char[e11];
for(i2=0;i2<e11;i2++)
{
m2[e12][i2]=q[i2];
}
m2[e12][e11]='\0';
number2[e12]=e11;
e12=e12+1;
cout<<"\n########################################\n";
for(i1=0;i1<e2;i1++)
{
for(i2=0;i2<number2[i1];i2++)
{
cout<<m2[i1][i2]<<" ";
}
cout<<"\n";
}
int ref1=0,ref2=0,i3,i4,i5;
int k1=0,k2=0;
//cout<<nseq;
for(i3=0;i3<(strlen(nseq)-e1);i3++)
{
k1=0;
for(i4=i3;i4<(i3+e1);i4++)
```

```
{
for(i5=0;i5<number1[i4-i3];i5++)
{
//cout<<i4<<" "<<i5;
//cout<<" "<<m1[i4-i3][i5]<<" "<<nseq[i4]<<"\n";
if(m1[i4-i3][i5]==nseq[i4] )
{
k1=k1+1;
i5=number1[i4-i3];
}
//cout<<"x";
}
}
if(k1 == e1)
{
ref1=1;
}
}

for(i3=0;i3<(strlen(pseq)-e2);i3++)
{
k1=0;
for(i4=i3;i4<(i3+e2);i4++)
{
for(i5=0;i5<number2[i4-i3];i5++)
{
if(m2[i4-i3][i5]==pseq[i4] )
{
k1=k1+1;
i5=number2[i4-i3];
}
}
}
if(k1 == e2)
{
```

```
ref2=1;
}
}
if(ref1 == 1 && ref2 == 1 )
{
cout<<s3;
break;
}
 }
}
```

Chapter IV

RNA-Protein Interaction

Significance

RNA, in its varied forms, interacts with protein to carry out fundamental roles in the cell. Understanding the contributions of various RNAs to the control of translation (protein synthesis on the ribosome) in the cell forms an important theme within the structural biology and biophysics group. The understanding of molecular recognition events, including those involved in assembly of macromolecular complexes consisting of both protein and RNA are a particular challenge for in vitro biochemical, biophysical and structural study. The research group interests are centered on the role of RNA-protein interactions on the post-transcriptional regulation of gene expression. It not only provides a paradigm of a novel mechanism of regulation, but also offers important indications on how pre-RNA splicing can be regulated by RNA structures and proteins.

The following algorithm is essentially the same as the DNA-Protein interaction except for the observation that the nucleotide 'U' replaces nucleotide 'T'.

DNA-DNA Interaction (Sec. Str. Pred.)

(E) ALGORITHM FOR DNA-DNA INTERACTION (Sec.Str. Pred.):

```
#include<iostream.h>
#include<stdio.h>
#include<string.h>
#include<conio.h>
#include<process.h>
void main()
{clrscr();
char seq1[100],seq2[100];
int len1,len2,i;
cout<<"Enter Sequence 1\n";
gets(seq1);
len1 = strlen(seq1);
for(i=0;i<len1;i++)
{
if(seq1[i] =='A' || seq1[i]=='T' || seq1[i]=='G' || seq1[i]=='C' || seq1[i]=='N')
{}
else
{
cout<<"\nError in sequence";
exit(0);
}}
cout<<"\nEnter Sequence 2\n";
gets(seq2);
len2 = strlen(seq2);
for(i=0;i<len2;i++)
{
```

```
if(seq2[i] =='A' || seq2[i]=='T' || seq2[i]=='G' || seq2[i]=='C' || seq2[i]=='N')
{}
else
{
cout<<"\nError in sequence";
exit(0);
}}
char temp[100],temp1[100],inv[100];
int j,k,a,m,n,l,total=0;
if(len1<=6)
{
exit(0);
}
else
{
i=0;
a=i+6;
while( a < len1)
{
k=0;
for(j=i;j<(i+6);j++)
{
temp[k]=seq1[j];
k++;
}
temp[k]='\0';
cout<<"\n"<<temp;
for(j=0;j<6;j++)
{
if(temp[j]=='A')
{inv[j]='T';
}
else if(temp[j]=='T')
{inv[j]='A';
}
```

```
else if(temp[j]=='G')
{inv[j]='C';
}
else if(temp[j]=='C')
{inv[j]='G';
}
}
inv[j++]='\0';
m = i + 1;
n = m + 6;
while(n <= len1)
{
l=0;
for(j=m;j<(m+6);j++)
{
temp1[l]=seq1[j];
l++;
}
temp1[l]='\0';
cout<<"\n"<<temp1;
if(strcmp(inv,temp1)==0)
{
cout<<"\n"<<temp<<" found an inverse with "<<temp1;
total++;
}
m++;
n=m+6;
}
i++;
a=i+6;
}
}
if(total==0)
{
cout<<"\nNo Secondary structures can be formed by Sequence 1";
```

```
}
else
{
cout<<"\n Total number of matches found \t"<<total;
cout<<"\n Secondary structures can be formed BY Sequence 1";
}
total=0;
if(len2<=6)
{
exit(0);
}
else
{
i=0;
a=i+6;
while( a < len2)
{
k=0;
for(j=i;j<(i+6);j++)
{
temp[k]=seq2[j];
k++;
}
temp[k]='\0';
cout<<"\n"<<temp;
for(j=0;j<6;j++)
{
if(temp[j]=='A')
{inv[j]='T';
}
else if(temp[j]=='T')
{inv[j]='A';
}
else if(temp[j]=='G')
{inv[j]='C';
```

```
}
else if(temp[j]=='C')
{inv[j]='G';
}
}
inv[j++]='\0';

m = i + 1;
n = m + 6;
while(n <= len2)
{
l=0;
for(j=m;j<(m+6);j++)
{
temp1[l]=seq2[j];
l++;
}
temp1[l]='\0';
cout<<"\n"<<temp1;
if(strcmp(inv,temp1)==0)
{
cout<<"\n"<<temp<<" found an inverse with "<<temp1;
total++;
}
m++;
n=m+6;
}
i++;
a=i+6;
}
}
if(total==0)
{
cout<<"\nNo Secondary structures can be formed by Sequence 2";
}
```

```
else
{
cout<<"\n Total number of matches found \t"<<total;
cout<<"\n Secondary structures can be formed BY Sequence 2";
}
getch();
}
```

Chapter VI

RNA-RNA Interaction (Sec. Str. Pred.)

Significance

Single-stranded RNA molecules can form local secondary structures through the interactions of complementary segments. These secondary structure elements may influence many cellular processes, including mRNA stability and localization, transcription, RNA processing, and translation. Functionally important RNA secondary structures can be found in untranslated regions (UTRs), introns, and coding sequences. For example, in eukaryotes, stem-loop structures in 5' UTRs may prevent association of the 40S ribosomal subunit with the mRNA, inhibiting translation initiation. Similarly, secondary structure elements in bacterial 5' UTRs can reduce the rate of mRNA degradation through the inhibition of nuclease activity. Introns and coding regions may also contain important structural elements. For example, mutations predicted to disrupt a stem-loop structure in intron 10 of the human tau gene cause higher levels of inclusion of exon 10 and are linked to debilitating neurodegenerative conditions. Additionally, a stem-loop in the coding sequence of the yeast ASH1 gene can localize ASH1 mRNA to the bud tip. It is widely believed that secondary structure in ORFs can interfere with translation giving rise to the expectation that RNA structure should generally be avoided in coding regions.

(F) ALGORITHM FOR RNA-RNA INTERACTION (Sec.Str. Pred.):

```
#include<iostream.h>
#include<stdio.h>
#include<string.h>
#include<conio.h>
#include<process.h>
void main()
{clrscr();
char seq1[100],seq2[100];
int len1,len2,i;
cout<<"Enter Sequence 1\n";
gets(seq1);
len1 = strlen(seq1);
for(i=0;i<len1;i++)
{
if(seq1[i] =='A' || seq1[i]=='U' || seq1[i]=='G' || seq1[i]=='C' || seq1[i]=='N')
{}
else
{
cout<<"\nError in sequence";
exit(0);
}}
cout<<"\nEnter Sequence 2\n";
gets(seq2);
len2 = strlen(seq2);
for(i=0;i<len2;i++)
{
if(seq2[i] =='A' || seq2[i]=='U' || seq2[i]=='G' || seq2[i]=='C' || seq2[i]=='N')
{}
else
{
cout<<"\nError in sequence";
exit(0);
```

```
}}
char temp[100],temp1[100],inv[100];
int j,k,a,m,n,l,total=0;

if(len1<=6)
{
exit(0);
}
else
{
i=0;
a=i+6;
while( a < len1)
{
k=0;
for(j=i;j<(i+6);j++)
{
temp[k]=seq1[j];
k++;
}
temp[k]='\0';
cout<<"\n"<<temp;
for(j=0;j<6;j++)
{
if(temp[j]=='A')
{inv[j]='U';
}
else if(temp[j]=='U')
{inv[j]='A';
}
else if(temp[j]=='G')
{inv[j]='C';
}
else if(temp[j]=='C')
{inv[j]='G';
```

```
}
}
inv[j++]='\0';
m = i + 1;
n = m + 6;
while(n <= len1)
{
l=0;
for(j=m;j<(m+6);j++)
{
temp1[l]=seq1[j];
l++;
}
temp1[l]='\0';
cout<<"\n"<<temp1;
if(strcmp(inv,temp1)==0)
{
cout<<"\n"<<temp<<" found an inverse with "<<temp1;
total++;
}
m++;
n=m+6;
}
i++;
a=i+6;
}
}
if(total==0)
{
cout<<"\nNo Secondary structures can be formed by Sequence 1";
}
else
{
cout<<"\n Total number of matches found \t"<<total;
cout<<"\n Secondary structures can be formed BY Sequence 1";
```

```
}
total=0;
if(len2<=6)
{
exit(0);
}
else
{
i=0;
a=i+6;
while( a < len2)
{
k=0;
for(j=i;j<(i+6);j++)
{
temp[k]=seq2[j];
k++;
}
temp[k]='\0';
cout<<"\n"<<temp;
for(j=0;j<6;j++)
{
if(temp[j]=='A')
{inv[j]='U';
}
else if(temp[j]=='U')
{inv[j]='A';
}
else if(temp[j]=='G')
{inv[j]='C';
}
else if(temp[j]=='C')
{inv[j]='G';
}
}
```

```
inv[j++]='\0';

m = i + 1;
n = m + 6;
while(n <= len2)
{
l=0;
for(j=m;j<(m+6);j++)
{
temp1[l]=seq2[j];
l++;
}
temp1[l]='\0';
cout<<"\n"<<temp1;
if(strcmp(inv,temp1)==0)
{
cout<<"\n"<<temp<<" found an inverse with "<<temp1;
total++;
}
m++;
n=m+6;
}
i++;
a=i+6;
}
}
if(total==0)
{
cout<<"\nNo Secondary structures can be formed by Sequence 2";
}
else
{
cout<<"\n Total number of matches found \t"<<total;
cout<<"\n Secondary structures can be formed BY Sequence 2";
}
```

getch();}

Chapter VII

DNA-RNA Interaction

Significance

Significance of DNA-RNA interactions lies in the observation that these biomolecular interactions are demonstrated to represent, in part, the molecular biology of some of the viruses, *viz.*, Retroviruses (Rich, 2006). Furthermore, with the discovery of cellular reverse transcriptases, any work on DNA-RNA interactions provides vital insights into the biology of cells and hence assumes utmost significance (Kantor *et al.*, 1979). Experiments were carried out with polyribonucleotides of different composition , and their interactions could generally be explained in terms of the ability of the bases to form at least two hydrogen bonds in the center of the molecule.

(G) ALGORITHM FOR DNA-RNA INTERACTION:

```
#include<iostream.h>
#include<stdio.h>
#include<string.h>
#include<conio.h>
#include<process.h>
void main()
{clrscr();
char seq1[100],seq2[100],bond[100];
int len1,len2,i;
cout<<"Enter Sequence 1\n";
gets(seq1);
len1 = strlen(seq1);
for(i=0;i<len1;i++)
{
if(seq1[i] =='A' || seq1[i]=='T' || seq1[i]=='G' || seq1[i]=='C' || seq1[i]=='N' || seq1[i] =='U')
{}
else
{
cout<<"\nError in sequence";
exit(0);
}}
cout<<"\nEnter Sequence 2\n";
gets(seq2);
len2 = strlen(seq2);
for(i=0;i<len2;i++)
{
if(seq2[i] =='A' || seq2[i]=='T' || seq2[i]=='G' || seq2[i]=='C' || seq2[i]=='N' || seq1[i] =='U')
```

```
{}
else
{
cout<<"\nError in sequence";
exit(0);
}}
i=0;
while(seq1[i]!='\0' && seq2[i]!='\0')
{
if(seq1[i]=='A'&& seq2[i]=='T')
{
bond[i]='|';
}
else if(seq1[i]=='T' && seq2[i]=='A')
{
bond[i]='|';
}
else if(seq1[i]=='A' && seq2[i]=='U')
{
bond[i]='|';
}
else if(seq1[i]=='U' && seq2[i]=='A')
{
bond[i]='|';
}
else if(seq1[i]=='G' && seq2[i]=='C')
{
bond[i]= '/';
}
else if(seq1[i]=='C' && seq2[i]=='G')
{
bond[i]='/';
}
else
{
```

```
bond[i]=' ';
}
i++;}
bond[i]='\0';
cout<<"\n";
cout<<seq1;
cout<<"\n"<<bond;
cout<<"\n"<<seq2;
getch();
}
```

Chapter VIII

Protein-Protein Interaction

Significance

Protein–protein complexes that dissociate and associate readily, often depending on the physiological condition or environment, play an important role in many biological processes. From sequence alignments we find that the interface residues of the weak transient homodimers are generally more conserved than surface residues, consistent with being constrained to maintain the protein–protein interaction during evolution. Protein families that include members with different oligomeric states or structures are identified, and found to exhibit a lower sequence conservation at the interface. The results are discussed in terms of the physiological function and evolution of protein–protein interactions.

(H) ALGORITHM FOR Protein-Protein INTERACTION:

```
#include<iostream.h>
#include<conio.h>
#include<stdio.h>
#include<string.h>
#include<stdlib.h>
void main()
{
clrscr();
char seq1[100],seq2[100];
int len1,len2,i,a=0,b=0;
cout<<"Enter Sequence 1\n";
gets(seq1);
len1 = strlen(seq1);
for(i=0;i<len1;i++)
{
if(seq1[i] =='G' || seq1[i]=='A' || seq1[i]=='P' || seq1[i]=='V' || seq1[i]=='L' || seq1[i]=='I' ||
seq1[i]=='M' || seq1[i]=='F' || seq1[i]=='Y' || seq1[i]=='W' || seq1[i]=='S' || seq1[i]=='T' || seq1[i]=='C'
|| seq1[i]=='N' || seq1[i]=='Q' || seq1[i]=='k' || seq1[i]=='H' || seq1[i]=='R' || seq1[i]=='D' ||
seq1[i]=='E')
{}
else
{
cout<<"\nError in sequence";
a=1;
}}
for(i = 6; i<len1;i=i+7)
{
if(seq1[i]!='L')
{
```

```
cout<<"Interaction is not possible";
a=1;
}
}
cout<<"\nEnter Sequence 2\n";
gets(seq2);
len2 = strlen(seq2);
for(i=0;i<len2;i++)
{
if(seq2[i] =='G' || seq2[i]=='A' || seq2[i]=='P' || seq2[i]=='V' || seq2[i]=='L' || seq2[i]=='I' ||
seq2[i]=='M' || seq2[i]=='F' || seq2[i]=='Y' || seq2[i]=='W' || seq2[i]=='S' || seq2[i]=='T' || seq2[i]=='C'
|| seq2[i]=='N' || seq2[i]=='Q' || seq2[i]=='k' || seq2[i]=='H' || seq2[i]=='R' || seq2[i]=='D' ||
seq2[i]=='E')
{}
else
{
cout<<"\nError in sequence";
b=1;
}
}
/*
if(len1 <= 9 || len2 <= 9)
{
cout<<"Interaction is not possible";
}
*/
if(seq1[3]=='T' && seq1[9]=='E' && seq1[14]=='E' && seq1[16]=='E' && seq1[17]=='K' &&
seq1[23]=='E' && seq1[30]=='E' && seq1[31]=='k')
{}
else
{
cout<<"Interaction is not possible";
a=1;
}
if(seq2[4]=='R' && seq2[8]=='L' && seq2[11]=='K' && seq2[15]=='L' && seq2[16]=='K' &&
seq2[18]=='Q' && seq2[19]=='N' && seq2[22]=='L' && seq2[29]=='L' && seq2[30]=='R' &&
seq2[32]=='Q' && seq2[36]=='L' && seq2[37]=='K')
```

```
{}
else
{
cout<<"Interaction is not possible";
b=1;
}
if(a==0 && b==1)
{
cout<<"Sequence 1 can interact with some other sequence but not with Sequence 2";
}
if(a==1 && b==0)
{
cout<<"Sequence 2 can interact with some other sequence but not with Sequence 1";
}
if(a==1 && b==1)
{
cout<<"Both sequence can't interact";
}
if(a==0 && b==0)
{
cout<<"Sequence 1 can interact with  Sequence 2";
getch();
}
```

Chapter IX

DNA-DNA Interaction (Seq. inf. Not given)

(I) ALGORITHM FOR DNA-DNA INTERACTION (Seq.Info. not given):

```
#include<iostream.h>
#include<stdio.h>
#include<string.h>
#include<conio.h>
#include<process.h>
void main()
{clrscr();
char seq1[100],seq2[100],bond[100];
char seq3[100],seq4[100],bond1[100],bond2[100];
int len1,len2,i;
float ph;
cout<<"Enter Sequence 1\n";
gets(seq1);
len1 = strlen(seq1);
for(i=0;i<len1;i++)
{
if(seq1[i] =='A' || seq1[i]=='T' || seq1[i]=='G' || seq1[i]=='C' || seq1[i]=='N')
{}
else
{
cout<<"\nError in sequence";
exit(0);
}}
cout<<"\nEnter Sequence 2\n";
```

```
gets(seq2);
len2 = strlen(seq2);
for(i=0;i<len2;i++)
{
if(seq2[i] =='A' || seq2[i]=='T' || seq2[i]=='G' || seq2[i]=='C' || seq2[i]=='N')
{}
else
{
cout<<"\nError in sequence";
exit(0);
}}
cout<<"\n Enter the Ph value\t";
cin>>ph;
i=0;
while(seq1[i]!='\0' && seq2[i]!='\0')
{
if(seq1[i]=='A'&& seq2[i]=='T')
{
bond[i]='|';
}
else if(seq1[i]=='T' && seq2[i]=='A')
{
bond[i]='|';
}
else if(seq1[i]=='G' && seq2[i]=='C')
{
bond[i]= '/';
}
else if(seq1[i]=='C' && seq2[i]=='G')
{
bond[i]='/';
}
else
{
bond[i]=' ';
```

```
}
i++;
}
bond[i]='\0';

i=0;
while(seq1[i]!='\0')
{
if(seq1[i]=='A')
{ seq3[i]='S';
        if(ph>6.8)
        {
        bond1[i]='!';
        }
        else
        {
        bond1[i]='I';
        }
}
else
{
seq3[i]=' ';
bond1[i]=' ';
}
i++;
}
seq3[i]='\0';
bond1[i]='\0';
i=0;
while(seq2[i]!='\0')
{
if(seq2[i]=='A')
{ seq4[i]='S';
        if(ph>6.8)
        {
```

```
        bond2[i]='!';
        }
        else
        {
        bond2[i]='I';
        }
}
else
{
seq4[i]=' ';
bond2[i]=' ';
}
i++;
}
seq4[i]='\0';
bond2[i]='\0';
cout<<"\n";
cout<<"\n"<<seq3<<"\n";
cout<<bond1<<"\n"<<seq1;
cout<<"\n"<<bond;
cout<<"\n"<<seq2;
cout<<"\n"<<bond2<<"\n"<<seq4;
getch();}
```

CONCLUDING REMARKS

- These programs form a starting point upon which a database can be constructed and linked to the programs, a task that can be taken up by the subsequent batches.

- Additionally, as more numbers of motifs are yet to be discovered, a provision for incorporating the same by the users is another possible task that can be taken up by the subsequent workers.

- These algorithms can be used as a single module to check whether an interaction is possible or not between two input sequences.

BIBLIOGRAPHY

- Alexander Rich. *Discovery of the Hybrid Helix and the First DNA-RNA Hybridization. J. Biol. Chem., Vol. 281, Issue 12, 7693-7696, March 24, 2006s*

- J A Kantor, YH Lee, JG Chirikjian, and WF Feller. *DNA polymerase with characteristics of reverse transcriptase purified from human milk.* Science, Vol 204, Issue 4392, 511-513

- Mengeritsky G, Smith TF. *Recognition of characteristic patterns in sets of functionally equivalent DNA sequences.* Comput Appl Biosci. 1987 Sep

- *Encyclopedia of human genome.*

- *Biophysical Chemistry by* Cantor & Schemmel .