Modelinig and Simulation of Amino Acide

Basil Younis Thanoon Mathmathics Dept College of Computer Sciences & Mathematics Mosul, Iraq

ABSTRACT

The study of amino acids was one of the important issues in bioinformatics, and the prediction of the secondary structure of proteins was one of the important steps in the knowledge of the structure and function of the protein. In this research, an algorithm to generate amino acids is suggested using simulation. A software program is build using MATLAB according to the proposed algorithm for the purpose of conducting simulation experiments. Fourteen simulation experiments were performed to generate sequences of different sizes of amino acids of fourteen protein, some of them is private of mitochondria diseases and some other were taken from other types of proteins. Comparisons are performed between the data generated by the proposed algorithm with real data available in international global centers in genetic engineering databases. Percentages of successfulness of similarities and identity between successive cases with those generated by the simulation program were calculated. The practical application of the proposed algorithm indicated that this algorithm gives encouraging results than the similarities proportion between generated data with real data are sometimes exceeds 90%.

Keywords

modeling, simulation, bioinformatics, amino acids, prediction, secondary structure of proteins.

1. INTRODUCTION

The process of identification of the amino acids and genes analysis considered a modernness important concept and finds a great interest by the searchers.

The huge development in the bimolecular and biochemical science and analysis and study of the anatomical map which illustrates the genes or the hereditary formants carry in human inside the human cells.

In addition to the use of the nuclear acid DNA in achieving trust of people because it is the most important biological formant and the most accurate one made the need so important to read the chain of the nuclear acid DNA and identify it.

The proteins perform a lot of different functions and in general it consists of 20 different amino acids combined together by peptide bond to determine the proteins by linear sequence from these amino acids ,and this sequence of amino acid determine the final corps and function of the protein.

In general the protein molecules have different levels of composition contains the initial composition level and the secondary composition level and the tertiary .[2]composition level and sometimes the 4th level in some proteins

The reading of the amino acids sequence considered as the most important concept of biological statistics because the

Fatima Mahmood Hasan Zamzwm Mathmathics Dept College of Computer Sciences & Mathematics Mosul, Iraq

most diseases in human should investigated

for their hereditary bases and causes to make it easy for the doctors to treat them, that's why it becomes so important to determine the basic proteins composition ,and also the determination of the first step to build the knowledge and opens the door to design treatments that suits the humans hereditary specs and to understand the elderly health problems. Not all the amino acids leans to get in or involve in the secondary composition and this is because some amino acids are more leaning to get in Alpha _helices composition, in other side some amino acids leans to become Beta - strand and some leans to weakens or destroy the helixes because of the clash of the side chains and in other side some of the amino acids are almost suitable to all the composition Table(1) declines the leaning of the amino acid to get in secondary composition and it is noticed that the favoring of amino acids to get in one of the compositions is not complete [5], [6]

Table (1):Inclination amino acid to enter in to secondary structures.

amino acid	α- helixes	β- stran ds	Turn
Alanine	1.29	0.90	0.78
Cysteine	1.11	0.74	0.80
Leucine	1.30	1.02	0.59
Methionine	1.47	0.97	0.39
Glutamic Acid	1.44	0.75	1.00
Glutamine	1.27	0.80	0.97
Histidine	1.22	1.08	1.69
Lysine	1.23	0.77	0.96
Valine	0.91	1.49	0.47
Isoleucine	0.97	1.45	0.51
Phenylalanine	1.07	1.32	0.58
Tyrosine	0.72	1.25	1.05
Tryptophan	0.99	1.14	0.75
Threonine	0.82	1.21	1.03
Glycine	0.56	0.92	1.64
Serine	0.82	0.95	1.33
Aspartic Acid	1.04	0.72	1.41

Asparagine	0.90	0.76	1.28
Proline	0.52	0.64	1.91
Arginine	0.96	0.99	0.88

2. PREDICTION OF THE SECONDARY PROTEIN COMPOSITION

The process of prediction of secondary structure of a protein composition from its primary sequence one considered as a special importance in biochemistry . and the prediction of the protein corps according to the sequence of the amino acids is considered as one of the main unsolved problems in bimolecular science . in addition to that the detection of the proteins composition is so important to understand the keys of different functions of these proteins. [11]

The detection of the main (basic) functions of the proteins and their structural composition became one of the great challenges in drug designing field

And with the increase of differences between the amount of knowledge that published by the genome projects and the number of discovered proteins with the function and compositions makes the dependence on computerized informational tools inevitable . The process of creating proteins takes place first by the use of the DNA hereditary material composing genes from the cells as a map for its building, then the second stage of preparing protein will take place in the cell during the protein molecules to certain composition to do the function which is responsible for in the cell, And the type of the taken secondary composition by the protein molecule considered as a key to this preparing process which done through molecular folding by away that determined by its first composition till the final normal composition which will do its function [2]. There are a lot of method s for detection of proteins secondary composition and these methods differ from each other by their accurate intuition to the 2ndory composition and also they differ from generation to generation The enhancements were taken upon the old methods were basically depend on considering the special relationship with protein rolling theory and the inter ship that occurs between the proteins and by these enhancements the special statistics of each residue has returned depending a large number of compositionally detected proteins .

In addition to input of a lot of possible residues inter ship and the most famous method that used in detecting the secondary protein composition which has been suggested by the two researchers Chou and Fasman in the seventies

The chou and fasman methods determine the possibility of Alpha- helices composition and Beta- strands taken by the amino acids depending on the x-ray analysis results of a lot of well-known second compositioning proteins.

The possibility of taking Alpha- helices composition by a residue depends on the possibility of existence of this residue within the Alph- helices composition regions that could be calculated by this equation [5]:

$$f_{\alpha} = \frac{No. of x residue in \alpha regions}{Total No. of x residues}$$

And this gives an example of different residues that exist in Alpha – helices region and usually composition parameter calculated by this equation [6]:

(1)

As f_{α} the probability represent different residues that fall within the snail body area-Alpha, And is usually calculated body spiral parameter of the formula [10]:

$$P_{\alpha} = \frac{f_{\alpha}}{\langle f_{\alpha} \rangle}$$

The $\langle f_a \rangle$ represents the average of the possibility of the residues that exists in the Alpha- helices region and in same way the possibility of existence of this residue in Beta –strands composition can be calculated by this equation [5].

 $f_{\beta} = \frac{No. of x residue in \beta regions}{Total No. of x residues}$

(3)

(2)

The parameter of Beta–strand composition can be calculated by this equation [5]

$$P_{\beta} = \frac{f_{\beta}}{\langle f_{\beta} \rangle} \tag{4}$$

3. SIMULATION AMINO ACIDS

The process of secondary protein composition detection depends basically on the knowledge of the initial protein composition which consist of 20 amino acid, and there are a lot of bases that available in the internet contains proteins and secondary compositions of these proteins and these proteins that exist in these data bases depends on the usage of x-ray and an expensive magnetic resonance imaging

That's why an simulation method has been created to initial the reality and generate an amino acid sequence which represent the initial protein composition and in different sizes and by usage of table (1) which represent the lean of the amino acids to be involved in the secondary protein composition

In this research there is an algorithm suggested to generate a sequence on the available in formation in the previous schedules .For generation of a sequence of amino acids of certain size N we suggest this algorithm :

The algorithm(1): generation of a sequence of amino acids by using simulation.

`Step(1): Entrance of the wanted sequence size in the
simulationNStep(2): Entrance of table data that represent the lean of the
amino acid to be involved in the secondary composition
Table(1).

Step(3): formation of amino acid probability matrix P, and the probability law of total has been used for this purpose(*), see[4]:

International Journal of Computer Applications (0975 – 8887) Volume 141 – No.13, May 2016

P(Ala)=P(Ala/Alfa)P(Alfa)+P(Ala/Beta)P(Beta)+P(Ala/Rot)P(Rot)P(Cvs)=P(Cvs/Alfa)P(Alfa)+P(Cvs/Beta)P(Beta)+P(Cvs/Rot)P(Rot) P(Leu)=P(Leu/Alfa)P(Alfa)+P(Leu/Beta)P(Beta)+P(Leu/Rot)P(Rot)P(Met)=P(Met/Alfa)P(Alfa)+ P(Met/Beta)P(Beta)+P(Met/Rot)P(Rot) P(Glu)=P(Glu/Alfa)P(Alfa)+P(Glu/Beta)P(Beta)+P(Glu/Rot)P(Rot)P(Gln)=P(Gln/Alfa)P(Alfa)+P(Gln/Beta)P(Beta)+P(Gln/Rot)P(Rot)P(His)=P(His/Alfa)P(Alfa)+P(His/Beta)P(Beta)+P(His/Rot)P(Rot)P(Iys)=P(Iys/Alfa)P(Alfa)+P(Iys/Beta)P(Beta)+P(Iys/Rot)P(Rot)P(Val)=P(Val/Alfa)P(Alfa)+P(Val/Beta)P(Beta)+P(Val/Rot)P(Rot)P(Ile)=P(Ile/Alfa)P(Alfa)+ P(Ile/Beta)P(Beta)+P(Ile/Rot)P(Rot) P(Phe)=P(Phe/Alfa)P(Alfa)+P(Phe/Beta)P(Beta)+P(Phe/Rot)P(Rot)P(Tyr)=P(Tyr/Alfa)P(Alfa)+P(Tyr/Beta)P(Beta)+P(Tyr/Rot)P(Rot)P(Trp)=P(Trp/Alfa)P(Alfa)+ P(Trp/Beta)P(Beta)+P(Trp/Rot)P(Rot) P(Thr)=P(Thr/Alfa)P(Alfa)+P(Thr/Beta)P(Beta)+P(Thr/Rot)P(Rot)P(Gly)=P(Gly/Alfa)P(Alfa)+ P(Gly/Beta)P(Beta)+P(Gly/Rot)P(Rot) P(Ser)=P(Ser/Alfa)P(Alfa)+ P(Ser/Beta)P(Beta)+P(Ser/Rot)P(Rot) P(Asp)=P(Asp/Alfa)P(Alfa)+P(Asp/Beta)P(Beta)+P(Asp/Rot)P(Rot)P(Asn)=P(Asn/Alfa)P(Alfa)+ P(Asn/Beta)P(Beta)+P(Asn/Rot)P(Rot) P(Pro)=P(Pro/Alfa)P(Alfa)+ P(Pro/Beta)P(Beta)+P(Pro/Rot)P(Rot) P(Arg)=P(Arg/Alfa)P(Alfa)+ P(Arg/Beta)P(Beta)+P(Arg/Rot)P(Rot)

(*) the Probability: Law of Total If, A1, A2, ... Ak are dual negative events and general in sample space S, so for any other event B in S so:

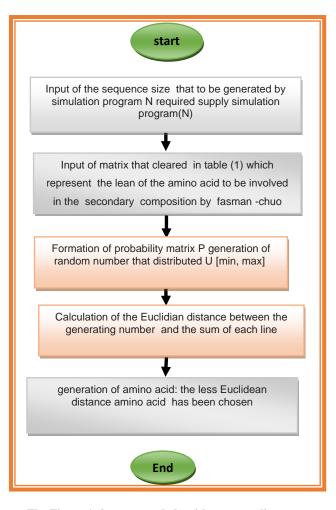
P(B)=P(B|A1)P(A1)+P(B|A2)P(A2)+...+P(B|Ak)P(Ak)

Step 4: generation of random number that follow uniform distribution in the interval [min, max] which is U [min, max] that (min) represents the least value of the sum of each line of the matrix P, which represent the lean of the amino acid to be involved in the secondary composition and (max) represent the largest value of the sum of each line of the matrix p which represent the lean of amino acid to be involved in the secondary composition.

Step5: Calculate the Euclidean distance between the generated number and the sum of each line of the matrix that's obvious in step(2).

Step6:generation of amino acid: the amino acid that has the less Euclidean distance between the sum of each line of the matrix P and the random number that generated will be chosed . the next form clears the stream line chart of algorithm (1)

The following figure shows the flow chart of the algorithm(1)



The Figure 1:the suggested algorithm streamline to generate a sequence of amino acid by using simulation

There is a computer program which has been built by using MATLAB language according to the previous algorithm to do an experiments for simulation .look at the un published ph .Dthesis of the other researcher [3].

4. ALIGNMENT

The pair wise sequence alignment represent its most easiest form that the main target from aligning a pair of sequence is to find the best pair of sequence which has the most number of similar amino acid and one sequence represents the querying sequence that want to search for sequence similar in its database and of known composition and function ,and the process of composition of sequence considered as one of the most important analysis method in biological information, and it is the 1st step in the path of analysis of new sequence structure and function and in which the search for the matched amino acid residue symbol in the related sequence or DNA bases symbol take place.

The pair wise sequence alignment process considered as a main base in the searching in databases for the multiple sequence alignment and this alignment process gives the dual sequence under standing relationship detection possibility that if the two sequences are shared in incorporeal degree of similarity or identify that its so far to this identity to be random and this mean that these two sequences could be descended from combined developmental sources (root or radically matching)

Homologous [10] but also there is a possibility of

mismatching in some regions resulting from amino acids changing processes (in protein form situation) and here the weighting of the possibility of descending of the two sequences from one source (origin) is possible. but they became so far that the radical relation which are not capable to be characterized on the relay level and this appears obviously clear in the fact that the nucleotide in DNA be in two situation either similar or different in comparison with three situation that happens in the amino acids they are either different or similar or different but not matching that's mean that some amino acids are similar in chemical composition like serine and thereonin that each one contain groups of hydroxyl (-OH) and also Lucien and isolucien have similar chemical characters and glutamic and aspartic both are of acidic interaction

The programs outputs differs in the multiply alignment sequences show and most of them depends on colures (in protein form situation)that the colur became specs for each group of amino acids depending on the physiochemical characters but the nitrogen bases have fixed color in most programs of according to the space of the amino acids that have been divided in to many groups which are look also at the next shape [7]:

1.polar amino acids that with green colures are N, Q, C,Y,T,S,G.

2.basic amino acids that with blue colure H, R, K. 3.acidic amino acids that with red colure are E, D. 4. water reluctant amino acid that with black colure are of large number W, P, I, L, V, A, M, F.



Figure 2: amino acid groups.

It is Also possible to consider the amino acid H as a polar amino acid and according to amino acid characters which b divided in to groups that are similar to the previous groups (Homology)

5. RADICAL MATCHING AND SIMILARITY AND MATCHING OF SEQUENCES

The sequence homology considered a one of the most important bases in sequences analysis when two sequences descended from one origin so it is possible to say that they have sequences homology or they have shared grandfather in versus the sequence similarity represents the alignment ratio and the amino acids residue matching which are similar in their physiochemical characters like size charge and water reluctance there for it was important to distinguish between the term radical matching (homology) and the other related terms like sequence similarity and also sequence identity[9].

That means that the residues are themselves in the two sequences and not from the group and the fact there is a confusion in using of these terms for example : the radical matching clears the relationship of the shared granddads that taken from the sequences comparison that are highly degree similar . but the term similar is a direct result from the observations resulted from sequences alignment and it is possible to determine the quantity of sequences similarity by using percentage . But the sequences radical matching is a qualitative state (case): for example it is possible to say that the two sequences shared by 40% of similarity and its wrong to say that the two sequences shared by a 40% of radical matching : because there is either a radical matching between them or there is not .

In general if the sequence similarity level is too high so it is possible to conclude that there is a shared developmental relationship in spite of that sometimes this relationship is not always clear and the answer depends on the type of the below studying and on the length of the sequences.

It is obvious from all what mentioned above that the nucleotide sequences consist of 4 letters which are the number of nitrogenous bases and then the sequences that have no connection between them have the possibility of matching 14 = 0.25 = 25%, at least as a result of random matching, on the other side the protein sequence have 25% that consist of 20 amino acids which have no matching possibility relationship reach to 1/20= 0.05=5% of matching as a result of random coincidence and in case of gaps usage the percentage increased to reach up to 10-20%

The sequence length considered as an important factor : so the short sequences have a larger chance of matching that caused by random coincidence and that's why it was necessary to put a cut off to the short sequences during identification matching relationships in comparison with long sequences.

For example if a 100 amino acid length sequences aligned so the similarity of 30% or more denotes that the two sequences have convergent similarity and the next shape clears the statistic symbolizing of similar degrees :

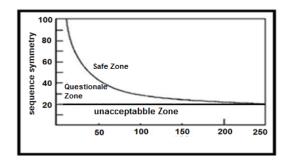


Figure 3: similarity degrees statistic representation.

The matching level of 20-30% indicates that the similarity relationship situated within uncertain region which is called (twilight zone) in which the little similarity intersects with sequences intersections as a result of random relationships but when the similarity percentage is less than 20% so a high percent of sequences appear to be not related sequences and appear symbolized in the shape as called (midnight zone)which are undependable results.

This symbolization couldn't be considered as an accurate measurement to determine the relationship between the sequences especially in the relationship between the sequences especially in the twilight zone and that's why it needs more strong statistic methods to determine the similarity relationship to determine the statistic in corporeal which will be explaining later.

The matching and similarity are used synonymously in the nucleotide sequences and the matching of two sequences refers to the percentage of the same amino acids presences in the two sequences .but the similarity refers to the percentage of amino acids alignment in the same group . which has similar physiochemical characters which are more liable to be replaced by each other without a big

Influence. there are two methods to calculate the sequences matching and similarity [1]

The first method :This method include the use of the longest sequence so if a and b are sequences and the length of each one is La and Lb, respectively and Ls represents the number of the character similar aligned amino acids in the two sequence so the sequences a ,b similarity percentage calculated by this equation:

$$S = [(Ls \times 2) / (La + Lb)] \times 100\%.$$
 (5)

And also could be calculated according to this standard equation :

$$\mathbf{S} = (\mathbf{Ls} / \mathbf{La}) \times 100. \tag{6}$$

But the sequencing similarity calculated in similar method as below:

$$I = [(Li \times 2) / (La + Lb)] \times 100\%.$$
 (7)

Li considered as number of the matched amino acids. It also can be calculated according to this standard equation:

$$I = (Li / La) \times 100.$$
 (8)

The second method: This method calculation depends on the percentage of each matching and similarity of the amino acid numbers by using this form

 $S = [(Ls \times 2) / (La + Lb)] \times 100\%$ (9)

I(S) = Li(S) / La % (10)

That La represents the shorter sequence from the two sequences number study .

6. DATA AND DATA SIMULATION 6.1 Data Description

The data that are used this research are 14 proteins . some of them are special for mitochondrial diseases and especially patients of gram ping chromosome to the rotina sudden spasmodic epilepsy, and the other has been taken from another types of proteins of different sizes of prion . look at the unpublished thesis of 2nd searcher and these information are available in data bases in international centers specialized in genetic engineering and gene function study like NCBI, Gen Bank, MBL

6.2 Amino Acids Simulation Experiments

A new program has been built according to suggested algorithm generates an amino acids sequences which represents the initial composition of any protein to be generated .look at the unpublished thesis of the 2nd searcher for details .and in this program the dependence on the available information in the previous table (1) has been done which are a scientific experiments depend on x-ray of proteins group of gaining the lean of the amino acids to be involved in the 2ndory composition a different sized sequences have been generated which are equals to a certain disease specialized proteins size exists in international data bases. and a data similarity of matching percentage has been found which are generated from the simulation program and the factious existent sequence in these data bases . as mentioned below a summary of 14 simulation experiment that carried on 14 international known

First experiment : when choosing a sequence size of N=1075 for mitochondrial brain disuse disease protein simulation which consist of 1075 amino acids the program gives this sequence.

GAGNGRMASDLSRAGPVERDIEQAIIALKKGAYLLKYRLSND ETVLIWFSSNDETVLIWFSGNEEISGORTPIFORYSGORTPIYPR PEKEYQSFSLIYSERSLRSLDVICKDKDEAEVWFTGLKALISHC HQRNRRTESRSDGTPSEANSPRTYTRRSSPLHSPFSSNDSLQKD **GSNHLRIHSPFESPPKNGLDKAFSDMALYAVPPKGFYPSDSATI** SVHSGGSDSMHGHMRGMGMDAFRVSMSSAVSSSSHGSGHDD GDALGDVFIWGEGIGEGVLGGGNRRVGSSFDIKMDSLLPKAL ESTIVLDVQNIACGGQHAVLVTKQGESFSWGEESEGRLGHGV DSNIQQPKLIDALNTTNIELVACGEFHSCAVTLSGDLYTWGKG DFGVLGHGNEVSHWVPKRVNFLLEGIHVSSIACGPYHTAVVT SAGQLFTFGDGTFGVLGHGDKKSVFIPREVDSLKGLRTVRAA CGVWHTAAVVEVMVGSSSSSNCSSGKLFTWGDGDKGRLGHG NKEPKLVPTCVAALVEPNFCOVACGHSLTVALTTSGHVYTMG SPVYGQLGNSHADGKTPNRVEGKLHKSFVEEIACGAYHVAVL TSRTEVYTWGKGSNGRLGHGDVDDRNSPTLVESLKDKQVKSI ACGTNFTAAVCIHRWASGMDOSMCSGCROPFSFKRKRHNCY NCGLVFCHSCTSKKSLKACMAPNPNKPYRVCDKCFNKLKKTE EKHLKLSHVSSRRGSINTPIFQRYPRPEKEYQSFSLIYSESLMES MRQVDSRHKKNKKYGGIGHCLSPIPSGSSQGMALNIAKSFNPV FGNTPGLYGTHMNGGMPTLFFTHPNATMYFVANPTQMPGGN SASLAGTVGFNFFPGGFLNQFDTMGDSVKLRSQVTRKAQLQE VELERTTKQLKEALAIGFGGTSSFQNFMLLGTTRLGNGGTLTE RLPVGSAGSARTVTQGVGGFPAALLMFFANILNQANSQESEPS EITTPMFSNGSNYFNGQVNFSLQGLIALTGGMFLQNRPPYITLT GPAGGARYLIDLTIALYGLG

A- when a generated sequence compared with a real protein sequence of (1075) amino acids size a following results may be obtained : that the letter (I) refers (red colour) do residua matching (amino acids) between dates which generated from simulation program with the real data that exist in the international center.

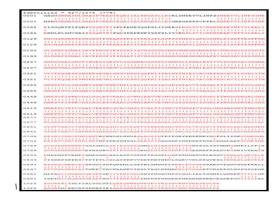


Figure (4): Sequence alignment in sequence generated the first experiment

And for a simulation programs efficacy checking a percentage of matching between areal sequences cases and chose generated from a simulation program calculated and as it is clear from the previous figure so the percentage matching between areal sequences cases with the generated sequences cases from simulation program reach to 77% of cases .And through the existence of such a match between the 1st sequence composition it is possible to predict the position and corp of 2nd protein composition which generated from a simulation program with 77% of success

B. also the similarity percentage between areal sequences cases and the generated sequences cases from a simulation program calculated and the figure (5a) clear areal sequence . but the figure (5b) clears a generated sequence . that the special colures of each group of amino acids depends on physiochemical characters (properties).

			10		20		30	4	0	50	60
			- T		цĨ		чŤ			il	Turn
	1	MSRNGRM	ASD	LERAGPVI	ERD	IEQAIIAL	KK	GAYLLKYGRE	GEPEFCPF	RL SNDE	TVLIUF
н.	61	SGNEEKH				RTPIFORY		PEKEYQSFSI		AI CKDK	
	121	FTGLKAL		CHORNER		RSDGTPSE		SPRTYTERS			DGSNHL
	181	RIHSPFE		KNGLDKA		MALYAVPP		FYPSDSATIS			GHGHDA
	241	FRVSMSS		SSSHGSG		GDALGDVF		GEGIGEGVLO			DSLLPK
	301	ALESTIV		QNIACGG		VLVTROGE		SWGEESEGRI			DALNTT
		NIELVAC		HSCAVTL		LYTWCKCD		VLGHGNEVSI			VSSIAC
	421 481	GPYHTAV		AGOLFTF		TEGVLCHC		KSVFIPREVI			WHTAAV
	481	VEVHVGS TVALTTS		SNCSSGKI VTMGSPV		UGD GD KGR		HGNKE PKLVI PNEVE GKLHI			ACGHSL
	601	EVYTUGE		GRLGHGD		RNSPTLVE		KDKOVKSIA			SGMDOS
	661	MCSGCRO		FERENN		CGLVFCHS		SEESLEACMA			NELEET
		METDPSS		LSRRGST		SDPIDKDD		DSRSDGOLA			RENERY
	781	EFNSSRV		PSGSSOR		NIAKSFNP		GASEKFFSAS			SPRPSP
	841	PRSTTPT		SGLATPK		DDTKRTND		SOEVVKLRSO			
	901	OLKEALA		EETTRCK		EVIKSLTA		KDMAERLPV			GSSPGR
	961	IDPFNIL		NSOESEPI		TTPMFSNG		TPAFGNGEAT			PGVYIT
	021							ADNRGRVYE			
		t	fig	ure(5	ja)	area	l s	equen	ce		
_			fig		a)		l s	-			
		10 1	fig	ure(5	(a)	area	l s	equen	ce		60
1	GAGN							40		SNDETV	60 l
		10 l	LSR	20	IEQ.	30 	GAY	40 LLKYRLS N	50	SNDETW	LIVE
61	SGNE	10 IGRMASD	LSRI	20 AGPVERD	IEQ. RTP	30 l AIIAL <mark>KK</mark>	GAY EYQ	40 LLKYRLS N SFSLIYS E	50 DETVLIWFS		LIUF
61 21	SGNE FTGL	IO IGRMASD EISGQR	LSRI TPII	20 AGPVERD FORYSGO	IEQ. RTP RSD	30 AIIALKK IYPRPEK	GAY EYQ SPR	40 LLKYRLS N SFSLIYS E TYTRRSS P	50 DETVLIWFS RSLRSLDVI	CKDKDE	LIWF
61 21 81 41	SGNE FTGL RIHS FRVS	IO GRMASD EISGOR KALISH PFESPP MSSAVS	LSRI TPII CHQI KNGI	20 AGPVERD FORYSGO RNRRTES	IEQ. RTP RSD HAL	30 AIIALKK IYPRPEK GTPSEAN	GAY EYQ SPR FYP	40 LLKYRLS N SFSLIYS E TYTRRSS P SDSATIS V	50 DETVLIWFS RSLRSLDVI LHSPFSSND	CKDKDE	LIWF AEVW SNHL IGMDA
61 21 81 41 01	SGNE FTGL RIHS FRVS	IO IGRMASD EISGOR KALISH SPFESPP	LSRJ TPII CHQI KNGI SSSI QNIJ	20 AGPVERD FORYSGO RNRRTES LDKAFSD HGSGHDD ACGGQHA	IEQ. RTP RSD HAL GDA	30 AIIALKK IYPRPEK GTPSEAN YAVPPKG	GAY EYQ SPR FYP GEG	40 LLKYRLS N SFSLIVS E TYTRRSS P SDSATIS V IGEGVLG G	50 DETVLIWFS RSLRSLDVI LHSPFSSND MSGGSDSNH	CKDKDE SLQKDO GHMRGE DIKMDS PKLIDA	LIWF AEVU SNHL IGMDA LLPK LNTT
61 21 81 41 01 61	SGNE FTGL RIHS FRVS ALES NIEL	IGRMASD EISGQR KALISH PFESPP MSSAVS TIVLDV VACGEF	LSRJ TPII CHQI KNGI SSSI QNIJ HSCJ	20 AGPVERD FORVSGO RNRRTES LDKAFSD HGSGHDD ACGGOHA AVTLSGD	IEQ. RTP RSD HAL GDA VLV LYT	30 AIIALKK IYPRPEK GTPSEAN YAVPPKG LGDVFIW TKQGESF WGKGDFG	GAY EYQ SPR FYP GEG SUG VLG	40 LLKYRLS N SFSLIYS E TYTRRSS P SDSATIS V IGEGVLG G EESEGRL G HGNEVSH W	50 DETVLIUFS RSLRSLDVI LHSPFSSND HSGGSDSMH GNRRVGSSF HGVDSNIQQ VPKRVNFLL	CKDKDE SLQKDC GHMRGE DIKMDS PKLIDA EGIHVS	LIWF AEVU SNHL GMDA LLPK LLPK SIAC
61 21 81 41 01 61 21	SGNE FTGL RIHS FRVS ALES NIEL GPYH	IGRMASD EISGQR KALISH PFESPP MSSAVS TIVLDV VACGEF TAVVTS	LSRJ TPII CHQI SSSI QNIJ HSCJ	20 AGPVERD FORVSGO RNRRTES LDKAFSD HGSGHDD ACGGQHA AVTLSGD LFTFGDG	IEQ. RTP RSD HAL GDA VLV LYT TFG	30 AIIALKK IYPRPEK GTPSEAN YAVPPKG LGDVFIW TKQGESF WGKGDFG VLGHGDK	GAY EYQ SPR FYP GEG SWG VLG KSV	40 LLKYRLS N SFSLIYS E TYTRRSS P SDSATTS V IGEGVLG G EESEGRL G HENEVSH W FIPREVD S	50 DETVLIUFS RSLRSLDVI LHSPFSSND HSGGSDSMH GNRRVGSSF HGVDSNIQQ VFRVNFLL LKGLRTVRA	CRDRDE SLORDO GHMRGH DIRMDS PRLIDA EGIHVS ACGVUE	LIWF AEVW SNHL GMDA LLPK LNTT SIAC (TAAV
61 21 81 41 61 21 61	SGNE FTGL RIHS FRVS ALES NIEL GPYH VEVM	LO CERMASD EISGOR KALISH PFESPP MSSAVS TIVLDV VACGEF TTAVVTS	LSRJ TPII CHQI SSSI QNIJ HSCJ AGQI SNC:	20 AGPVERD FORYSGO RNRRTES LDKAFSD HGSGHDD ACGGQHA ACTLSGD LFTFGDG SSGKLFT	IEQ. RTP RSD HAL GDA VLV LYT TFG WGD	30 AIIALKK IYPRPEK GTPSEAN YAVPPKG LGDVFIW TKQGESF WLGKGDFG VLGHGDK GDKGRLG	GAY EYQ SPR FYP GEG SUG VLG KSV HGN	40 LLKYRLS N SFSLIVS E TYTRRSS P IGEGVLG G EESEGRL G HGNEVSH W KEPKLVP T	50 DETULIUFS RSLRSLDVI LHSPFSSND HGVDSNIQQ VPRRVNFLL LKGLRTVRA CVAALVEPN	CKDRDE SLQKDC GHMRGE DIKMDS PKLIDA EGIHVS ACGVWE FCQVAC	LIUF AEVU SNHL GMDA LLPK LLPK SIAC TAAV GHSL
61 21 41 01 61 21 61 21 81 41	SGNE FTGL RIHS FRVS ALES NIEL GPYH VEVM TVAL	IO IGRMASD EISGOR KALISH PFESPP MSSAVS TIVLDV VACGEF TAVVTS IVGSSSS .TTSGHV	LSRJ TPII CHQI SSSI QNIJ HSCJ AGQI SNC: YTH	20 AGPVERD FORYSGO RNRTES HGSGHDD ACGGQHA AVTLSGD LFTFGDG SSGKLFT SSFVYGQ	IEQ. RTP RSD HAL GDA VLV LYT TFG UGD LGN	30 AIIALKK IYPRPEK GTPSEAN LGDVFIW TKQGESF UGKGDFG VLGHGDK VLGHGDK SHADGKT	GAY EYQ SPR FYP GEG SWG VLG KSV HGN PNR	40 LLKYRLS N SFSLIVS E TYTRRSS P SDSATLS V IGEGVLG G HGNEVSH W FIPREVD S VESKLVP T VEGKLMK S	50 DETVLIMFS RSLRSLDVI LHSPFSSND HSGGSDSMH GNRRVGSSF HGVDSNIQQ VPKRVNFLL LKGLRTVRA LKGLRTVRA FVEEIACGA	CKDRDE SLQRDG GHMRGE DIRMDS PRLIDA EGIHVS ACGVUE FCQVAC YHVAVI	LIWF AEVW SNHL IGMDA LLPK LLPK SIAC TAAV GHSL TSRT
61 21 81 01 61 21 81 61 81 81 81 01	SGNE FTGL RIHS FRVS ALES NIEL GPYH VEVM TVAL EVYT	LO EISGOR KALISH MSSAVS TIVLDV VACGEF TAVVTS VACSSS TTSGHV WGKGSN	LSRJ TPII CHQI SSSI QNIJ HSCJ AGQI SNC: YTRI GRLI	20 AGPVERD FORYSGO RNRRTES LDKAFSD HGSGHDD ACGGQHA AVTLSGD LFTFGDG SSGKLFT SSFVFGQ GHGDVDD	IEQ. RTP RSD HAL GDA VLV LYT TFG UGD LGN RNS	30 AIIALKK IYPRPEK GTPSEAN YAVPPKG UGKGDFG UGKGDFG VLGHGDK GDKGRLG GDKGRLG SHADGKT PTLVESL	GAY EYQ SPR GEG SWG VLG KSY HGN PNP KDR	40 LLKYPLS N SFSLIYS E SDSATIS V IGEGVLG G EESEGRL G EESEGRL S KEPKLVP T FIPREVD S KEPKLVP T VEGKLHK S QVKSIAC G	50 DETVLIUFS RSLRSLDVI LHSPFSSND HSGGSDSNH GNRRVGSSF HGVDSNIQQ VPRRVNFLL LKGLRTVRA CVAALVEPN FVEFIACGA TNFTAAVCI	CKDRDE SLQRDG GHMRGE DIRMDS PKLIDA EGIHVS ACGVUE FCQVAC YHVAVI HRWASG	LIWF AEVW SNHL GMDA LLPK LLPK LNTT SIAC TAAV GHSL TSRT MDQS
61 21 81 41 61 21 61 21 81 81 81 01 61	SGNE FTGL RIHS FRVS ALES NIEL GPYH VEVM TVAL EVYT MCSG	IO IGENASD EISGOR KALISH PFESPP MISSAVS VACGEF ITAVVTS VACGEF ITAVVTS VGCGSN UGCGSN CROPFS	LSRJ TPII CHQI SSSI QNIJ HSCJ AGQI SNC: VTRI GRLI FKRI	20 AGPVERD FORVSGO HGSGHDD ACGGGHA ACGGGHA ACTLSGD LFTFGDG SSGKLFT SSPVYGQ SHGDVDD KRHNCYN	IEQ. RTP RSD HAL GDA VLV LYT TFG UGD LGN RNS CGL	30 AIIALKK IYPRPEK GCTPSEAN YAVPPKG LGDVFIW TKQGESF WLGHGDK GDKGRLG SHADGKT SHADGKT PTLVESL VFCHSCT	GAY EYQ SPR FYP GEG SUG VLG KSV HGN PNR KDR SKK	40 LLEYPLS N SFSLIVS E TYTRRSS P SDSATIS V IGEGVIG G EESEGRL G HGNEVSH W KEPKLVP T VEGKLHK S QVRSIAC G SLKACMA P	50 DETVLINFS RSLRSLDVI LHSPFSSND HSGGSDSMH HGVDSNIQQ VPRCVNFLL LKCLRTVR LKCLRTVR CVAALVEPN FVEEIACGA TNFTAAVCI NNTKPYRVC	CKDRDE SLORDO GHMRGE DIRMDS PRLIDA EGIHVS ACGVVH FCQVAC YHVAVI HRWASG DKCFNH	LIWF AEVU SNHL GMDA LLNTT SIAC GHSL TSRT MDQS LLKKT
61 21 81 41 61 21 61 21 81 41 01 61 21	SGNE FTGL RIHS FRVS ALES NIEL GPYH VEVM TVAL EVYT MCSG EEKH	IGRMASD EISGQR KALISH PFESPP MSSAVS TIVLDV VACGEF TTAVVTS VGSSSS TTSGHV WGKGSN CCROPFS LLKLSHV	LSRJ TPII CHQI SSSI QNIJ HSCJ SNC: VTRI GRLI FKRI SSRI	20 AGPVERD QRYSGQ NRRTES LGSGHDD ACGGUHA AVTLSGD SIGKLFT SSPVYGQ SHGDVGQ SHGDVGV NRGSINTP	IEQ. RTP RSD GDA VLV LYT TFG UGD LGN RNS CGL IFQ	30 AIIALKK GTPSEAN YAVPPKG LGDVFIW TKQGESF WGKGDFG VLGHGDK SHADGKT PTLVESL SHADGKT PTLVESL RYPRPEK	GAY EYQ SPR FYP GEG SVG VLG KSV HGN PNR KDK SKK EYQ	40 LLKYRKS N SFSLIYS E TYTRRSS P IGEGVLG G EESEGRL G EESEGRL G FIPREVD S KEPKLVP T VEGKLMK S QVKSIAC G SLKACMA P SJKACMA P	DETVLIVES RSLRSLDVI LHSPFSSND GNRRVGSSF HGVDSNIQQ VPKRVNFLL LKGLRTVRA CVAALVEPN TNFTAAVCI NPHKPYRVC SLHESHRQV	CKDRDE SLORDO GHMRGE DIRMDS PRLIDA EGIHVS ACGVVH FCQVAC YHVAVI HRWASC DKCFNE DSRHKE	LIUF AEVU SNHL GMDA LLPK LLNTT SIAC TTAAV TTAAV TTAAV TSRT MDQS LLKKT NKKY
61 21 81 41 61 21 81 81 81 81 61 21 81 81	SGNE FTGL RIHS FRVS ALES NIEL GPYH VEVM TVAL EVYT NCSG EEKH GGIG	LO GRMASD EISGQR KALISH PFESPP MSSAVS TIVLDV VACGEF TAVVTS VGSSSS TTSGHV WGKGSN UGCSN UCROPFS UCROPFS UCROPFS	LSRI TPII CHQI SSSI QNII HSCI AGQI SNC: YTRI GRL0 FKRI SSRI PSG:	20 AGPVERD FORYSGO NMRRTES LDKAFSD IGSGHDD ACCGQHA AVTLSGD LFTFGDG SSGKLFT SSFVFGQ HGDVDD HGDVDD HGDVDD SSGGNAL	IEQ RTP RSD HAL GDA VLV LYT TFG UGD LGN RNS CGL IFQ NIA	30 AITALKK GTPSEAN YAVPPKG LGDVFIW TKGGESF WGKGDFG WLGHGDK GDKGRLG SHADGKT PTLVESL VYCHSCT RYPRPEK KSFNPVF	GAY EYQ SPR FYP GEG SVG VLG KSV HGN PNP KDR SKR EYQ GNT	40 LLKYRLS N SFSLIYS E TYTRRSS P SDSATIS V IGEGVIG G EESEGRL G HGNEVSH W KEPKLVF T VEGKLHK S QVKSLAC G SLKACMA P SFSLIYS E SFSLIYS E	50 DETVLIVES RSERSLDVI LHSPFSSND HSGGSDSNH USGGSDSNIQQ VPKRVNFLL LKGLFTWRA CVAALVEPN TNFTAAVCI NPTKPYRVC SLNESNRQV SLNESNRQV	CKDRDE SLOKDO GHMRGE DIKMDS PKLIDA EGIHVS ACGVWH FCQVAC YHVAVI HRWASO DKCFNF DSRHKF THPNAT	LIUF AEVU SNHL GMDA LLPK LLNTT SIAC TTAAV GHSL TTSRT MDQS LLRKT INKKY MYFV
61 21 81 41 61 21 81 61 21 81 61 21 81 81 81 41	SGNE FTGL RIHS FRVS ALES NIEL GPYH VEVM TVAL EVYT MCSG EEKH GGIG ANPT	LO GRMASD EISGOR KKALISH PFESPP INSSAVS TIVLDV VACGEF ITAVVTS IVGKSSS ITTSGHV UGKGSN ICROPFS ILKLSHV HCLSPI	LSRJ TPII CHQI SSSI QNIJ HSCJ AGQI SNCJ SRCJ FKRI SSRI PSG: SASI	20 AGGVERD GGRYSGQ NRRTES DKAFSD dGSGHDA ACGGQHA ACGGQHA ACGGQHA ACGGQHA CSGHLFT SSPUGQ SSGKLFT SSPUGQ SSGGNAL ACGTUGF	IEQ. RTP RSD HAL GDA VLV LYT TFG UGD LGN RNS CGL IFQ. NIA	30 AIIALKK GTPSEAN YAVPPKG LGDVFIW UKKGESF WICKGFG GDKGRLG GDKGRLG SNADGKT PTLVESL VFCHSCT RYPPPEK KSFNPYF PGGFLNO	GAY EYQ SPR FYP GEG SVG VLG KSV HGN PNP KDR SKR EYQ GNT FDT	40 LLKYPRIS N SFSLIYS E TYTRRSS P SDSATTS V ICECVIC G EESEGRIC G FIPREVDS N VECKLWF T VECKLWF T VECKLWF T VECKLWF S SIKACMA P SFSLIYS E PGLYGTH M MGDSYKL R	50 DETVLIUFS RSLRSLDVI LHSPFSSND HGVDSNIQQ VFRRVNFLL LKCLRTVRA FVEEIACGA TNFTAAVCI NFWEYRVC SLMSSNRQV NGGNFLFF	CKDRDE SLOKDO GHMRGE DIRMDS PRLIDA EGIHVS ACGVUE FCQVAC YHVAVI HRWASC DKCFNE DSRKKE DSRKKE QEVELE	LIUF AEVU SNHL CGMDA CLLPK LLNTT SIAC TTAAV GHSL TSRT MDQS LKKT NKKY NKKY WYFV RTTK
61 21 41 61 21 61 21 81 61 21 81 61 21 81 81 81 81 81 81 81 81 81 81 81 81 81	SGNE FTGL RIHS FRVS ALES NIEL GPYH VEVM TVAL EVYT NCSG EEKH GGIG ANPT QLKE	IGRMASD IGRMASD EISGQR KALISH IFFESPP IMSSAVS VACGEF ITAVDY VACGEF ITAVDS VCGSSSS ITTSGHY UCGSSSS ITTSGHY UCGSSSS ITTSGHY UCGSSSS ITTSGHY UCGSSSS ITTSGHY UCGSSSS ITTSGHY UCGSSSS ITTSGHY ICCOPFS ICCO	LISRA TPII CHOI SSSI QNIA HSCJI SNCI SSCI SSRI PSGI SASI GGTI	20 AGPVERD FORYSGO RNRTES LDKAFSD HGSGHDL LFTFGDG SSGKLFT SSPVFGO HGDVDD RGSINTP SSGGMAL LAGTVGF	IEQ. RTP RSD MAL GDA VLV LYT TFG UGD LGN RNS CGL IFQ. NIA NFF LLG	30 AITALKK IYPRPEK GTPSEAN YAVPREG LGDVFIW TKOGESF WLGHGDK GDKGRLG SHADGKT PTLVESL VFCHSCT RYPRPEK KSFNPWF PGGFLNQ TTFLGENG	GAY EYQ SPR FYP GEG SUG VLG KSV HGN PNP RDR SKR EYQ GNT FDT GTL	40 LLEXPELS N SFSLIYS E SDSATIS Y IGEGVIG G EESEGRI G HGNEVSH W KEPKLVP T VEGKLHK S SLKACMA P OVRSIAC G SLKACMA P PGLYGTH M MGDSVKL R HGDSVKL R	50 DETVLIUFS RSLFSLDVI HSGGSDSMH HSGGSDSMH HGVDSNIQQ VPRRVNFLL LKCLFTVRA CVAALVEPN FVEIACGA LKCLFTVRA CVAALVEPN SLMESHRQV NGGMPTLFF SQVTPRAQL	CRDRDE SLOKDO GHMRGE DIENDS PRLIDA EGIHVS ACGVUE FCOVAC YHVAVI HRWASG DRCFNE DSRHKP DSRHKP GVGGFF	LIUF AEVW SNHL CGMDA CLLPK LLNTT SIAC CTAAV GHSL LKKT MDQS LKKT MVFW MYFW WYFW AALL
61 21 41 01 61 21 81 41 01 61 21 81 41 21 81 41 01 81 41 01 61	SGNE FTGL RIHS FRVS ALES NIEL GPYH VEVM TVAL EVYT MCSG EEKH GGIG ANPT QLKE MFFA	LO CRIASD EISGOR KALISH PFESPF WSSAVS TIVLDV VACGEF TTAVUTS UKCSSS UKLSST UKLSPI OHPGGN SALATGF	LISRA TFII CHOI SSSI ONIA HSCA AGOI SNC: SSRI FKRI SSRI FKRI SSRI FSG SSRI SSRI SSRI SSRI SSRI SSRI SSRI SS	20 PORYSOU PORYSOU NURRES LDKAFSD LGKAFSD LGSGHDD ACGGUAA AVTLSGD SSGVAG SSGVAD RCSINFY SSGVAD LAGTVGF SSFQNFM SSFQNFM	III RTP RSD HAL GDA VLVT TFG UGD LGN RNS CGL IFQ NIA NFF LLG TTP:	30 ALIALKK IYPRPEK GTPSEAN YAVPPKG ULGDVFIW UKGKDFG VLGHGDK SHADGKT VFLWESL VFCHSCT RYPRPEK KSFNPVF PGGFLNQ TFRLGHG MFSNGSN	GAY EYQ SPR FYP GEG SUG VLG KSV HGN PNP KDK SKK EYQ GNT FDT GTL YFN	40 LLEXPELS N SFSLIYS E SDSATIS Y IGEGVIG G EESEGRI G HGNEVSH W KEPKLVP T VEGKLHK S SLKACMA P OVRSIAC G SLKACMA P PGLYGTH M MGDSVKL R HGDSVKL R	80 DETULIUFS RSL&SLDVI LISPFSSND HSOGDDMH HSOGDDMH HSOGDDMH CREPTGSE HONDNIOQ VPKRVNFL UGCHTTVA CALVEPN FYEILACG SLMESTRNOV NGCMTFLFF GOVTRKADC LIALTGOM	CKDRDE SLOKDC GHMRCE DIKNDS FRLDA EGIHYS ACGVUE FCOVAC YHVAVI HRWASC DKCFNE DSRHKE THPNAT QEVELE GVGGFF FLONRE	LIUF AEVW SNHL CGMDA CLLPK LLNTT SIAC CTAAV GHSL LKKT MDQS LKKT MVFW MYFW WYFW AALL

figure(5b) generated sequences for the simulation program

And according to the previous symbols so Ls refers to the number of the amino acids that aligned and of similar properties on the other side La, Lb refers to the sum of the length of each sequence separately as a sequence and the computer gave these result :

1075La = Lb =

SimlarityPerc = 81%

As the concluded sequences similarity percentage is 81% so this is an evidence that both have physiochemical properties that similar by 81% and by this there well be a shared evolutional relationship between them which may leads to a generated sequence to be replaced by the real one with success percentage of 81%

While considering the amino acid H as a polar amino acids the computer gave these results .

Ls = 9041075La = Lb =SimlarityPerc = 84%

The similarity percentage when considering amino acid H as a polar amino acid increases to be 84%.

2ndexperiment :When closing a sequence size N=818 in order to simulation a mitochondrial brain tissue disease protein which consist of 818 amino acids so the following sequence had been given by the program .

NTGTTTSHAMPHVGYFLLLMALLTACGGGSSSTDAPSEDTYP VGGTLTGLEQGHTVTLQLNGANDLTLDHTANDNPYSFAVKLP HGSAYEVTLPAEPSQHHCTIHNATGTVDGAVMNVDVECTTTF HAHLNLGTGPPHFTGYTPWAWGRNGYGRLGLGDTDDRDVPE QVGNFGFMGGAVNGGPNTGTYATYGTPALGLNTGAVGPNMS GDRDEPEQVGVDNDWIALSAGAMGTTGAKADGTLWAWGNF AYYGGLTHGANHHYTPTNLNLGGLPFFSNSAGRRHSLNMFGL PSHFFWGDNEYGQLGLGDTDERLTNGGASFPLFGLGTRGDQQ MPTGFRGGFGPWSWGYMTSGQLGLGDTADRNAPEQVGADT DWDLVSNFTSLVPNVKADGTLMSLLGPPGGLFLGGDTVSRDA PVPIASGIDTVAGRNHTVAVNPDGTSGFGGDNEYGQLGLGVM RYIVTPSQTVTGSYWAAVNGGSFNHSLGLRGQGTLWAWGNN NGMFGFSMDTDDRATPEQTVPNIDWAAVNAHSYHTLAVKID GTLWAWGRNSSGQLGLSDTNDRHTPERVGRDTDWATVSVG QSGLTGVKPDGTLWAWGWNHGTLGGYFFGPNDLSPEQVGGE MGSGNGCLGLHHTVAVKTDGTLWAFNFTARGQLGLGGGFLVG QTMALTFNPLMQGAVSASGYRTLAVKADGTLWAWGENNNG QLGLGDTDNRRGTAQVGNDADWAAVSTGLFHTLAFKEDDTL WAWGRNHDGQLGLGDTDNRDAPVQVGNETDWSTVSCGNGL YGGTMSDGTLRAWGLNTSGQLGQRTMWFDQNGLYSPC

A - when a generated sequence compared with a real protein sequence of (818) amino acids size following results has obtained :



Figure (6) : Sequence alignment in sequence generated the 2nd experiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 71% of cases .and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 71%.

B- regarding the similarity percentage between the real sequences cases and the sequences cases that generated from simulation program the figure (7a)clears the real sequence

.but the figure (7b)clears the generated sequence and the computer gave these results. regarding the figure(2):

	10	20	30	40	50	60
1	MTGTTTSHRL		ALLTACGGGS		YPVGGTLTGL	
61	NGANDLTLDH	TANDNPYSFA	VKLPHGSAVE	VTLPAEPSQH	HCTIHNATGT	VDGAVMNVDV
121	ECTAFPVLAA	GDDHTVAVKA	DGTLWAWGRN	GYGRLGLGDT	DDRDVPEQVG	TDSDWTAVSA
181	GRDHTVAVKA	NGTLWAWGRS	GRIGLGDSGD	RDEPEQVGVD	NDWIALSAGG	SHTVAVKADG
241	TLUAUGRNNA	GQLGLGDTNH	HYTPTQVGTD	TDUAKKISAG	RRHSLAVKAD	GSLUAWGDNE
301	YEQLELEDTD	ERHAPEQVGS	NTDWAAISAG	RRHSLAVKAD	GSLWSWGYND	SEQLELEDTA
361	DRNAPEQVGA	DTDWDLVSAG	GEHSLAVKAD	GTLNVUGLNH	LGQLGLGDTV	SRDAPVPIAS
421	GIALVAGRNH	TVAVNPDGTL	RAWGDNEYGQ	LGLGVMRYIV	TPSQTVTGSA	UAAVNGGSFN
481	HSLGLRDDGT	LUANGRNNYG	QLGLGDTDDR	ATPEQVGSHT	DUAAVNAASY	HTLAVKNDGT
541	LWAUGRNSSG	QLGLSDTNDR	HTPERVGSDT	DWATVSVGQS	HTSAVKPDGT	LWAUGWNHYG
601	QLGLGDTSDR	LSPEQVGGES	NULAVSTGLH	HTVAVKTDGT	LUANGUNVRG	QLGLGHTSDR
661	HAPVQVGSDT	DUAAVSASGY	RTLAVKADGT	LWAUGENNNG	QLGLGDTDNR	NTPAQVGNDA
721	DWAAVSTGLF	HTLAFKEDDT	LUAUGRNHDG	QLGLGDTDNR	DAPVQVGNET	DWSTVSGGGY
781	HTLATMSDGT	LRANGLNTSG	QLGQRTNUFD	QPQEVPGW		

figure(7a) real sequence

	10				50	
	hummer	hummer	hummer			h
1	NTGTTTSHAM	PHVGYFLPGS	VQSTACGGGS	SSTDAPSEDT	YPVGGTLTGL	EQGHTVTLQL
61	NGANDLTLDH	TANDNPYSFA	VKLPHGSAVE	VTLPAEPSQH	HCTIHNATGT	VDGAVMNVDV
121	ECTTTFHAHL	NLGTGPPHFT	GYTPPGQPRN	GYGRLGNFGG	HVVDVPEQVG	NFGFMGGAVN
181	GGPNTGTYAT	YGTPALGLNT	GAVGPNMSGD	RDEPEQVGVD	NDVIALSAGA	MGTTGAILFM
241	HNYNPANFAY	YGGLTHGANH	HYTPTNLNLG	GLPFFSNSAG	RRHSLNNFGL	PSHFFWGDNE
301	YNNLGGLGMF	NGLINGGASF	PLFGLGTRGD	QOMPTOFROG	FGPUSNGYMT	SGQLGLGDTA
361	DRNAPEQVGA	DTDWDLVSNF	TSLVPNVKAD	GTLNSLLGPP	GGLFLGGDTV	SRDAPVPIAS
421	GIDTVAGRNH	TVAVNPDGTS	GFGGDNEYGQ	LGLGVMRVIV	TPSQTVTGSY	WAAVNGGSFN
481	HSLGLRGQGT	LUAUGNNNGM	FGFSHDTDDR	ATPEQTVPNI	DWAAVNAHSY	HTLAVKIAGE
541	CVLVTNNSSG	QLGLSDTNDR	HTPERVGRDT	DWATVSVGQS	GLTGVKPDGT	LUAUGUNHGT
601	LGGYFFGPND	LSPEQVGGEN	GSGNGCLGLH	HTVAVKTDGT	LWAFNFTARG	QLGLGGFLVG
661	QTHALTFNPL	NUGAVSASGY	RTLAVKADGT	LWAUGENNNG	QLGLGDTDNR	RGTGPFTNDA
721	DUAAVSTGLF	HTLAFKEDDT	LUAUGRNHDG	QLGLGDTDNR	DAPVOVGNET	DUSTVSCONG
781	LYGGTMSDGT	LTPGGLNTSG	QLGQRTHWFD	QNGLYSPC		

Figure(7b) generated sequences for the simulation program

Ls= 640

La = Lb = 818

SimlarityPerc = 78%

As the concluded percentage of two sequence similarity is 78%.so this is an evidence that both have a similar physiochemical properties of about 78% and by this there will

be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of 78%.

But when considering an amino acid H as a polar amino acid the computer gave these results:

Ls= 661

La = Lb = 818

SimlarityPerc = 81%

The similarity percentage when considering an amino acid H as a polar amino acid increase to be 81%.

3rd experiment : when choosing a sequence of N =421 size in order to simulate a mitochondrial brain tissue disease protein which consist of 421 amino acid the program gave those results:

PGPKRIAKRRSPPADAIPKSKKVKVSHRSHSTEPGLVLTLGQG DVGQLGLGENSHYRKKPALVSIPEDVVQAEAGGMNGLYTFTF LTVYSFGCNDEGALGRDTSVEGSEMVPGKVELQEKVVQVSA GDSHTAALTDDGRVFLWGNLFSFGGGIGLLEPMKKSMVPVQ VQLDVPVVKVASGNDHLVMLTADGDLYTLGCGEQGQPHGV NTPGHLNRGGRQGLERLLVKCVMLLKSRGSRHVRFQDAFCG AYFFTDGLLGFFHLVPFTTLCMPNCTGVPFTLFIPQNLTSFKNS TKSNSFANGVQHHTVCMDSEGKAAYLGRAEYGRLGLGEGAE EKSIPTLISRRPAVSSVACGASVGYAVTKDGRVFAWGMGTNY YQLGTGQDEQTSPVEMSSGKQLENRVSVSSGGQHTVLLLVKD KEQS

A- when a generated sequence compared with a protein real sequence of 421 amino acid size following results had been gotten:

	titles = 306/421 (73%)
001	PG PKRIAKRRSPPADAIPKSKKVKVSHRSHSTEPGLVLTLGQGDVGQLGLGEN SHY RKKPALVS
001	MSPKRIAKRSPPADAIPKSKKVKVSHRSHSTEPGLVLTLGOGDVGOLGLGENVMERKKPALVS
001	HAPPENTARE APPADATE FARE AND A TEACHING AND A TEACH
065	I PEDVVQAEAGGMNGLYTFTFLTVYSFGCNDEGALGRDTSVEGSEMVPGKVELQEKVVQVSAGD
065	I PEDVVOAEAGMHTVCLSKSGOVYSFGCNDEGALGBDTSVEGSEMVPGKVELOEKVVOVSAGD
065	I FEDVVQAEAGGMHTVCLSRSGQVYSFGCNDEGALGRDTSVEGSEMVFGRVELQERVVQVSAGD
129	SHTAALTDDGRVFLWGNLFSFGGGIGLLEPMNKSMVPVOVOLDVPVVKVASGNDHLVMLTADGD
129	SHTAALTDDGRVFLWGSFRDNNGVIGLLEPMRKSMVPVQVQLDVPVVKVASGNDHLVMLTADGD
193	LYTLGCGEOGOPHGVNTPGHLNRGGROGLERLLVKCVMLLKSRGSRHVRFODAFCGAYFFTDGL
	111111111111 I I I I I I I I I I I I I
193	LYTLGCGEQGQLGRVPELFANRGGRQGLERLLVPKCVMLKSRGSRGHVRFQDAFCGAYFTFAIS
257	LGFFHLVFFTTLCMPNCTGVFFTLFIPONLTSFKNSTKSNSFANGVOHHTVCMDSEGKAAYLGR
207	
257	HEGHVYGFGLSNYHOLGTPGTESCFIPONLTSFKNSTKSWYGFSGGOHHTVCMDSEGKAYSLGB
321	AEYGRLGLGEGAEEKSIFTLISRRPAVSSVACGASVGYAVTKDGRVFAWGMGTNYYQLGTGQDE
321	AFYGRLGLGEGAFEKSIPTLISRLPAVSSVACGASVGYAVTKDGRVFANGMGTNYOLGTGODED
385	QTSPVEMSSGRQLENRVSVSSGGQNTVLLLVKDREQS
385	AWSPVEMMGROLENBYVLSVSSGGOHTVLLVRDKEOS

Figure (8): Sequence alignment in sequence generated the 3^{rd} experiment

And as it is clear from the previous figure the matching percentage between the real sequence cases and the generated sequence cases from the simulation program reach up to 73% of cases . and through the presence of such a matching between the primary protein composition so it is possible to predict the position and corp. of the 2ndory protein predict the position and corp of the 2ndory protein composition which generated from simulation . program with success percentage 73% of cases.

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from simulation program the figure(9a) clears the real sequences while the figure (9b) clear the generated sequences for the simulation sequences

-						
	10	20	30	40	50	60
	Juuruul	human	hunner	hunni	human	human
1	MSPKRIAKRR	SPPADAIPKS	KKVKVSHRSH	STEPGLVLTL	GQGDVGQLGL	GENVMERKKP
61	ALVSIPEDVV	QAEAGGMHTV	CLSKSGQVYS	FGCNDEGALG	RDTSVEGSEM	VPGKVELQEK
121	VVQVSAGDSH	TAALTDD GRV	FLWGSFRDNN	GVIGLLEPMK	KSMVPVQVQL	DVPVVKVASG
181	NDHLVMLTAD	GDLYTLGCGE	QGQLGRVPEL	FANRGGRQGL	ERLLVPKCVM	LKSRGSRGHV
241	RFQDAFCGAY	FTFAISHEGH	VYGFGLSNYH	QLGTPGTESC	FIPQNLTSFK	NSTKSWVGFS
301	GGQHHTVCMD	SEGKAYSLGR	AEYGRLGLGE	GAEEKSIPTL	ISRLPAVSSV	ACGASVGYAV
361	TKDGRVFAUG	MGTNYQLGTG	QDEDAWSPVE	MMGKQLENRV	VLSVSSGGQH	TVLLVKDKEQ
421	S					

figure(9a) real sequence

	Jummer	30	40	50	60
GTPFMPGRR	SPPADAIPKS	KKWKWSHRSH	STEPGLVLTL	GQGDVGQLGL	GENSHYRKKP
LVSIPEDVV	QAEAGGMNGL	YTFTFLTVYS	FGCNDEGALG	RDTSVEGSEM	VPGKVELQEK
VQVSAGDSH	TAALTDDGRV	FLUGNLFSFG	GGIGLLEPMK	KSMVPVQVQL	DVPVVKVASG
DHLVMLTAD	GDLYTLGCGE	QGQPHGVNTP	GHLNRGGRQG	LERLLVPKCV	MLKSRGSRHV
FQDAFCGAY	FFTDGLLGFF	HLVPFTTLCM	PNCTGVPFTL	FIPQNLTSFK	NSTRSNSFAN
VOHHTVCMD	SEGKAAYLGR	AEYGRLGLGE	GAEEKSIPTL	ISRRPAVSSV	ACGASVGYAV
KD GRVFAVG	MGTNYYQLGT	GQDEQTNNNG	SSSGKQLENR	VSVSSGGQHT	VLLLV <mark>KD</mark> KEQ
5					
	LVSIPEDVV VQVSAGDSH DHLVNLTAD FQDAFCGAY VQHHTVCND	LVSIPEDVV QAEAGGMNGL VQVSAGDSH TAALTDDGRV DHLVHLTAD GDLYTLGCGE FQDAFCGAY FFTDGLLGFF VQHHTVCHD SEGKAAYLGR	LVSIPEDVV QAEAGGMUGL YTFFLTVYS VQVSACDSH TAALTDDGRV FLWGNLFSFG DHLVMLTAD GDLYTLGCGE QGQPHGVNTP FQDAFCGAY FFTDGLLGFF HLVPFTTLCM VQHHTVCMD SEGKAAVLGR AEVGRLGLGE	LVSIPEDVV QAEAGGHNGL YTFTFLTVYS FGCNDEGALG YQVSAGDSH TAALTDDGRV FLWGHLSFFG GGTGLEPHK DHLVHLTAD GDLYTLGCGE QGOPHGVNTP GHLNRGGRQG FDQAFGCAY FFTDGLIGFF HLVFTTLCH PNCTVPFTL YQHHTVCHD SEGKAAYLGR AEYGRLGLGE GAEEKSIPTL	CTFFREGER SPADAIRSS KKWKYSHSM STEPCLVIL GOGOVGOLG LVSIFEDVV QAEAGGMNGL YTFFLTVYS FGCNDEGALG RDTSVEGSEM VOYSAGDSH TAALTDDGRY FUGULIESFG GGULEFMK KSWFVVQUD DHLVMLTAD GDLYTLGGGE QGQPHGVNTP GHLNRGGRQG LERLLVPKCV FQDAFCGAY FFTOGLLGFF HLVPFTTLGK PRCTGVFFTL FIPUNLTSK VQHTVCHD SECKAJUGR AFVGELGEG CAEEXSTPI LSRFAVSSV KDGRVFANG MGTNYYQLGT GQDEQTNNNG SSSGKQLENR VSVSSGQHT

figure (9b) generated sequences for the simulation program

and the computer gave these results. regarding the figure(2): .

Ls = 334

La = Lb = 421

SimlarityPerc = 79%

As the concluded sequences similarity percentage is 79% so this is an evidence that both have physiochemical properties that similar by 79% and by this there well be a shared evolutional relationship between them which may leads to a generated sequence to be replaced by the real one with success percentage of 79%

While considering the amino acid H as a polar amino acids the computer gave these results .

Ls =343

La = Lb = 421

SimlarityPerc = 81%

The similarity percentage when considering amino acid H as a polar amino acid increases to be 81%.

4thexperiment :When closing a sequence size N=1006 in order to simulation a mitochondrial brain tissue disease protein which consist of 1006 amino acids so the following sequence had been given by the program .

GVMPFANFLVPRDRTDEQAILALKKGAQLLKCRRRGNPKFCP FKLSMDEKYLIWYSGEEEROLRLSSVITIVRGOITPNFOKOAOS DRKEQSFSLIYANGEHTLDLLGGTCNTFNMNQGGNLQFGFGG LAFVGFLQGMHNGSCQMAGGRGPYNGLNKQNLGLLEETPDV TPFSTGATHLLGSTNVFGCLCFGLGGGHDTTGSDALGPVSSYY ETDYDFRNSGGFGANSGGDGFFSSQRFAASPPLSIITQPVTRSN VLKDIMIWMFALGLIDGSKNQNVTGSPKLLESATMFDVQNLY GGAKHAALVTRQGEVFCWGNGNSGTYYPFGSPGFPMRVGGT SLEDVAVRSVARQGFLTGPNGLNQQHLYNFNGNGFLTVNPSQ FLTRKISDVLGGSLTVLSVACGSLNPCNVTSSGQLFTYGSGTFG VLGHGSLESVTGLGNQFLVLILAQMSGFCTYGLGNFNNGFRL NGMLQGPTHGGKLFTWGDGDKGRLGHADSKRKLVPTCVTEL IDHDFIKVSCGWTLTVALSISGTVYTMGSSIHGQPGLMRAKDK SVNMGGFGDFYGVGALTVVVPSNGGYAGIMNGSGSMGGMA NFSSQYFQACTPGTPVLVEPLGDRLVESIACGLNLTAAICLHKE ISLNDQTACSSCKSAFGFTRRKHNCYNCGTGPFNACSSKKAVN ASLAPNKSKLSRVFMGLGNPGTGNTEFSRNVKMDNHTPRMQ MVTRRVSEDLTEKQSENEMQNLPQANRSTDGQPRWGQFGYA RGQACMFPTLSTNTNYYVSSTLHGGVSYGFNMSFSVNNTEEIE RLKAVIKNLORGCELGNEKMEECOOENORTWEVAKEEAEKS KAAKEIGKALASKLQANKEKPSNSLKTGIACNPSQVSPIFDPM LSIPYLTPITTARSQHETKQHVEKCVTKSSNRDSNIKLLVDASP AITRQLLGLVQTQDSSAEQVETFEPGVYITFTAGPCGQKTLKR NRFSRKRFSGRLAQRWWEVQGFGTLGNLFFSN

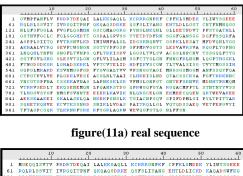
A - when a generated sequence compared with a real protein sequence of (1006) amino acids size following results has obtained :

	11162 - 722/1006 (72%)
0001	GVMPFANFL VPRDRTDEQAILALKKGAQLLKCRRRGNPKFCPFKLSMDEKYLIWYSGEEERQLR
0001	MGEQQISVTVPRDRTDEQAILALKKGAQLLKCRRRGNPKFCPFKLSMDEKYLIWYSGEEERQLR
0065	LSSVITIVRGQITPNFQKQAQSDRKEQSFSLIVANGEHTLDLLGGTCNTFNSWFKGLRAVGFGG
0065	LSSVITIVRGQITPNFQKQAQSDRKEQSFSLIVANGEHTLDLICKDKAQADSWFKGLRAVITKH
0129	LAFVGSVNHRHNGSCOMAGGRGPYNMRRRONLGLLEETPDVTPFRSLCGSPLGSTNVRCLSNGG
0129	NNIRNSVNHRSSRGAQSCINSPAGFMRRKQNLGLLEETPDVTQIRSLCGSPSTLLEERCLSNGL
0193	LGGGNDTTGSDALGPVSSYYETDYDFRNSGGFGANSGGDGFFSSQRFAASPPLSIITQPVTRSN
0193	SCSSDSFAESDALGPVSSYYETDYDFRNSDCDRSTGSELCRFSSORFAASPPLSIITOPVTRSN
0257	VLKDIMIWMFALGLIDGSKNONVTGSPKLLESATMFDVQNLYGGARMAALVTRQGEVFCWGNGN
0257	VLRDIMIWGAITGLIDGSKNONDALSPKLLESATMFDVOSISLGARMAALVTROGEVFCWGNGN
0321	SGTYYPFGSPGFPMRVGGTSLEDVAVRSVAROGFLTGPNTESGELYLWFNGNGFLTVNPSOFLT
0321	SGRLGLKVNIDIDHPRRVESLEDVAVRSVACSDNOTCAVTESGELYLWGIDGGTIEQSGSOFLT
0385	RKISDVLGGSLTVLSVACGSLNPCNVTSSGOLFTYGSGTFGVLGHGSLESVTGLGNOFLVLILA
0385	RKISDVLGGSLTVLSVACGAWHTAIVTSSGQLFTYGSGTFGVLGHGSLESVTKPKEVESLRRMK
0119	QMSGFCTYGLGNFNNGFRLNRRFYNARSCGRLFTWGDGDRGRLGHADSRRRLVFTCVTELIDHD
0449	VISVSCGPWHTAAIVETANDRKFYNAKSCGKLFTWSDSDKGRLGHADSKRKLVFTCVTELIDHD
0513	FIRVSCGWILTVALSISGTVYTMGSSIHGQFGLMRARDKSVNMGGFGDFYGVGALTVVVPSNGG
0513	FIRVSCONTLTVALSISGTVYTMGSSIHGQLGCPRARDRSVNVVLGNLTRQFVRDIASGSHHVA

Figure (10):Sequence alignment in sequence generated the 4thexperiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 72% of cases .and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 72%.

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from simulation program the figure(11a) clears the real sequences while the figure (11b) clear the generated sequences for the simulation sequences



1	ROLUQIDVIV	PRDRIDEGAL	LALKKGAULL	RURRRGHPRF	CFFRLDHDER	TLIWIDGEEL
61	RQLRLSSVIT	IVRGQITPNF	QKQAQSDRKE	QSFSLIYANG	EHTLDLICKD	KAQAD SUFKG
121	LRAVITKHHN	IRNSVNHRSS	RGAQSCINSP	AGFMRRKQNL	GLLEETPDVT	QIRSLCGSPS
181	TLLEERCLSN	GLSCSSDSFA	ESDALGPVSS	YYETDYD FRN	SDCDRSTGSE	LCRFSSORFA
241	ASPPLSIITQ	PVTRSNVLKD	INIWGAITGL	IDGSKNQNDA	LSPKLLESAT	MFDVQSISLG
301	AKHAALVTRQ	GEVFCUGNEN	SGKLGLKVNI	DIDHPKRVES	LEDVAVRSVA	CSDHQTCAVT
361	ESGELYLWGI	DGGTIEQSGS	QFLTRKISDV	LGGSLTVLSV	ACGAUHTAIV	TSSGQLFTYG
421	SGTFGVLGHG	SLESVTRPRE	VESLRRMKVI	SVSCGPUHTA	AIVETANDRK	FYNAKSCGKL
481	FTWGDGDKGR	L GHAD SKRKL	VPTCVTELID	HDFIKVSCGW	TLTVALSISG	TVYTNGSSIN
541	GQLGCPRAKD	KSVNVVLGNL	TROFVEDIAS	GSHHVAVLTS	FGNVYTWGKG	MNGQLGLGDV
601	RDRNSPVLVE	PLGDRLVESI	ACGLNLTAAI	CLHKEISLND	QTACSSCKSA	FGFTRRKHNC
661	YNCGLLFCNA	CSSKKAVNAS	LAPNKSKLSR	VCDSCFDHLW	SITEFSRNVK	HDNHTPRMQM
721	VTRRVSEDLT	EKQSENEMON	LPQANRSSDG	QPRWGQVSGP	SLFRFDKIST	SSSLNLSVSA
781	RRTSSTKIST	SSESNKILTE	EIERLKAVIK	NLOROCELON	EKMEECQQEL	DRTWEVAREE
841	AEKSKAAKEI	IKALASKLQA	NKEKPSNPLK	TGIACNPSQV	SPIFDDSMSI	PYLTPITTAR
901	SQHETKQHVE	KCVTKSSNRD	SNIKLLVDAS	PAITRTGYLQ	NETQDSSAEQ	VEQYEPGVYI
961	TFTALPCGQK	TLKRVRFSRK	RESEREAQEW	WEEKQVLVVN	KYDAEI	

figure(11b) generated sequences for the simulation program

and the computer gave these results. regarding the figure(2): .

- Ls =770
- La = Lb =1006
- SimlarityPerc =77%

As the concluded percentage of two sequence similarity is 77%.so this is an evidence that both have a similar physiochemical properties of about 77% and by this there will be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of 77%.

But when considering an amino acid H as a polar amino acid the computer gave these results:

Ls =785

La = Lb = 1006

SimlarityPerc =78

The similarity percentage when considering an amino acid H as a polar amino acid increase to be 78%.

5th experiment : When closing a sequence size N=890 in order to simulation a mitochondrial brain tissue disease protein which consist of N=890 amino acids so the following sequence had been given by the program .

PNGGGKPNSSSNSRDDGNSVFPAKASATGAGPAAAEKRLGTP PGGGGAGAKEHGNSVCFKVDGGGGGGGGGGGGGEPGFGGM DAEGPRRQYGFHANNVFNTMTVTGLFPGNMANLQNTPFEKE QERVKTAGFWIIHPYSDTPMGPNGGNGAYNSVGPFGPTGFTTF FTEQTTTGTPGFNVASDTVFLLDLIMNFRTGTVNEDSSEIILDP KVIKMNYLKGNSGGGFISSIPVDYIFLIVEKGMDSEVYKTARA LRNTHGLAQNTNNTSPGVFTNHCHGGSQTEIFHMTYDLASAV VRIFNLIGMCPLLCHWDGCLQFLVPLLQDFPPDCWVSLNEMV NGTYFLGTGFNFINGFPHMLCIGQFLTTPVSMSDLWITMLSMPI NYSTLRGTGSRSLIVFQSLDSSRRQYQEKYKQVEQYMSPGTLP ADMRQKIHDYYEHRYQGKIFDEENILNELNDPPPHVFFHGGG VIVATMPLFANADPNNPYAMLSKLRFEVFQPGDYIIREVNLPM PNGLGRNGPQGFGGGSSKEMKLTDGSYFGEICLLIMLVGLFGG SHNIYCRLYSLSVDNFNEVLEEYPMMRRAFETVAIDRLDFTVQ CNSILLQKFQKDLNTGVFNNQENEILKQSGKHDREMVQAIA

A - when a generated sequence compared with a real protein sequence of (890) amino acids size following results has obtained :

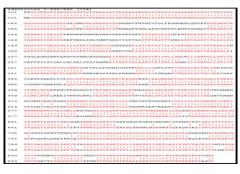


Figure (12) Sequence alignment in sequence generated the 5thexperiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 71% of cases .and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 71%.

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from simulation program the figure(13a) clears the real sequences while the figure (13b) clear the generated sequences for the simulation sequences

	10		30		50	60
1	MEGGGERPHSS				TPPGGGGAGA	KEHGNSVCFK
61		GGGGEEPAGG				
121		GFUIIHPYSD				TTTPWIIFNV
181	ASDTVFLLDL	INNFRIGIVN	EDSSEIILDP	KVIKNNYLKS	WFVVDFISSI	PVDYIFLIVE
241	KGNDSEVYKT	ARALRIVERT	KILSLLRLLR	LSRLIRVIHQ	WEEIFHMTYD	LASAVVRIFN
301	LIGHMLLLCH	WDGCLQFLVP	LLQDFPPDCW	VSLNEHVNDS	UGKQYSYALF	KANSHMLCIG
361	YGAQAPVSHS	DLUITMLSMI	VGATCYAMFV	GHATALIQSL	DSSRRQYQEK	YKQVEQYNSF
421	HKL PADMRQK	IHDYYEHRYQ	GKIFDEENIL	NELND PLREE	IVNFNCRKLV	ATMPLFANAD
481	PNFVTAMLSK	LRFEVFQPGD	YIIREGAVGK	KMYFIQHGVA	GVITKSSKEM	KLTDGSYFGE
541	ICLLTKGRRT	ASVRADTYCR	LYSLSVDNFN	EVLEEYPMMR	RAFETVAIDR	LDRIGKKNSI
601	LLQKFQKDLN	TGYFNNQENE	ILKQIVKHDR	ENVQAIAPIN	YPONTTLNST	SSTTTPTSRM
661	RTQSPPVYTA	TSLSHSNLHS	PSPSTQTPQP	SAILSPESYT	TAVCSPPVQS	PLAARTFHYA
721	SPTASQLSLM	QQQPQQQVQQ	SQPPQTQPQQ	PSPQPQTPGS	STPKNEVHKS	TQALHNTNLT
781	REVRPLSASQ	PSLPHEVSTL	ISRPHPTVGE	SLASIPQPVT	AVPGTGLQAG	GRSTVPQRVT
841	LFRQMSSGAI	PPNRGVPPAP	PPPAAALPRE	SSSVLHTDPD	AEKPRFASNL	

Figure (13a) real sequence

	10	20	30	40	50	60
1	PNGGGKPNSS	SNSRDDGNSV	FPARASATGA	GPAAAEKRLG	TPPGGGGAGA	KEHGNSVCFK
61	VDGGGGGGGGG	GGGGEEPGFG	GNDAEGPRRQ	YGFHANNVFN	THIVIGLEPG	NNANLQNTPF
121	EKEQERVKTA	GFWIIHPYSD	TPHGPNGGNG	AYNSVGPFGP	TGFTPYISNP	GTTGTPGFNV
181	ASDTVFLLDL	IMNFRTGTVN	EDSSEIILDP	KVIKMNYLKG	NSGGGFISSI	PVDYIFGGAN
241	KGND SEVYKT	ARALRNTHGL	AQNTNNTSPG	VFTNHCHGGS	QTEIFHMTYD	LASAVVRIFN
301	LIGNCPLLCH	WDGCLQFLVP	LLQDFPPDCW	VSLNEMVNGT	YFLGTGFNFI	NGFPHMLCIG
361	QFLTTPVSMS	DLWITHLSMP	INYSTLRGTG	SRSLIVFGHG	TTNTGQYQEK	YKQVEQYMSP
421	GINNYNISNL	CPGGCGLFYQ	GKIFDEENIL	NELNDPPPH	VFFHGGGVIV	ATHPLFANAD
481	PNNPYAMLSK	LRFEVVLDGL	TNSMGVNLPM	PNGLGRNGPQ	GFGGGVNTPG	LSDDGSYFGE
541	ICLLIMLVGL	FGGSHNIYCR	LYSLSVDNFN	EVLEEVPMMR	RAFETVAIDR	LDFTVQCSNF
601	PPFKFQKDLN	TOVENNOENE	ILKQSGKHDR	EMVQAIAPIN	YPONTTLNST	SSGTGGPGAL
661	QGPFLLNNGS	FGPGCFGPGS	NPYNTCDTNG	STYCSPCSYT	TAVCGPGGGL	GGSTTFNMYG
721	SPTASQLSLM	00000000000	SQPPQTQPQQ	PSPQDGASAN	STPKNEVHKS	TQALHNTNLT
781	REVRIQSASQ	PSLPHEVSTL	ISRPHPTVGE	SLASIPQPVT	AVPGTGNTGG	GRSTVPFGGR
841	GFRQMSSGAI	SRGLGVPPAP	PPPAAAMPRE	SSSVLNTDGA	TNTLFMCGFN	

Figure (13b) generated sequences for the simulation program

and the computer gave these results.

Ls =714

La = Lb = 890

SimlarityPerc =80%

As the concluded percentage of two sequence similarity is 80%.so this is an evidence that both have a similar physiochemical properties of about 80% and by this there will be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of 80%.

But when considering an amino acid H as a polar amino acid the computer gave these results:

Ls =726

La = Lb = 890

SimlarityPerc =82%

The similarity percentage when considering an amino acid H as a polar amino acid increase to be 82%.

 6^{th} experiment :When closing a sequence size N=774in order to simulation a mitochondrial brain tissue disease protein which consist of 774 amino acids so the following sequence had been given by the program .

```
TCGGORPAAGASEGATPGLELGPPVAPPPATAASGGLGSLFPE
PKRRHLGTLLQPTVNKFSLRVFGSHKAVEIEQERVKSAGAWII
HPYSDFRFYWDLCNGFGVLMCTYLGPNTNGPIFTGHCHFGGV
GYPNVNGTTLGPFLLNFRTGIVVEEGAEILLVTMTIRTRYLRT
WFLVDLISSTGTLTIANIIGFEPRLDAEVYKTAPDPTPTTPGNGF
SHFLGLRLSRLIRYIHQWEEFMGTLPANLSAVVRIFNLIGMML
LLCHWDGCLQFLVPMLQDFPPDCWVSINHMVNHSWGQGYSH
ALFKAMSHMLMTRGRLDGQHGNGTHWLTMLSMIVGATCYA
MFIGHATALIQSLDSSRRQYQEKYKQVEQYMSFHKLPADTRQ
RIHEYYEHRYQGKMFDEESILGGAHMPLREEIINFTCRGLVAH
MPLFAHADPSFVTAVLTKLRFEVFOPGDLVVREGSVGRNGFD
PQHGLLSVLLSNTNDTRLTDGSYFGEICLLTRGRRTASVRADT
YCLGFLLSVDHFNAVGEEFPMMRRAFETVAMDMHAVNFGQQ
RGGQRKRSEPSPGSSGGIMESLPPNNIFTLANGVRGRAPSTGA
QLSGKPVLWRGGVHAPLQAAAVTSNVAIALTHQRGPLPLSPD
SPATLLARSAWRSAGSPASPLVPVRAGPWASTSGLPAPPARTL
HASLSRAGRSQVSLLGPPPMMVGPYPGPRGRPLSASQPSLPQR
ATGDGSPGRKGSGSERLTPPGLLAKPPRHAQPPRPPVCFITAAT
OTOLSANM
```

A - when a generated sequence compared with a real protein sequence of (774) amino acids size following results has obtained :

Iden	tities = 601/774 (78%)
001	TCGGQRPAAGASEGATPGLELGPPVAPPPATAASGGLGSLFPEPKRRHLGTLLQPTVNKFSLRV
001	MEAEORPAAGASEGATPGLEAVPPVAPPPATAASGPIPKSGPEPKRRHLGTLLOPTVNKFSLRV
065	FGSHKAVEIEOERVKSAGAWIIHPYSDFRFYWDLCNGFGVLMCTYLGPNTNGPIFTGHCHFGGV
065	FGSHKAVEIEOERVKSAGAWIIHPYSDFRFYWDLIMLLLMVGNLIVLPVGITFFKEENSPPWIV
129	GYPNVNGTTLGPFLLNFRTGIVVEEGAEILLVTMTIRTBYLRTWFLVDLISSTGTLTIANIIGF
	1 1111111111111111111111111111111111111
129	FNVLSDTFFLLDLVLNFRTGIVVEEGAEILLAPRAIRTRYLRTWFLVDLISSIPVDYIFLVVEL
129	PRVESDIFFEEDEVENTRIGIVVEEGRETEERPRATRIKTERIWFEVDEISSIFVDIFFEVVEE
193	EPRLDAEVYKTAPDPTPTTPGNGF5HFLGLRLSRLIRYIHOWEEFMGTLPANLSAVVRIFNLIG
193	
193	
193	EPRLDAEVYKTARALRIVRFTKILSLLRLLRLSRLIRYIHQWEEIFHMTYDLASAVVRIFNLIG
257	MMLLLCHWDGCLQFLVPMLQDFPPDCWVSINHMVNHSWGQGYSHALFKAMSHMLMTRGRLDGQH
257	MMLLLCHWDGCLOFLVPMLODFPPDCWVSINHMVNHSWGROYSHALFKAMSHMLCIGYGOOAPV
321	GNGTHWLTMLSMIVGATCYAMFIGHATALIQSLDSSRRQYQEKYKQVEQYMSFHKLPADTRQRI
321	GMPDVWLTMLSMIVGATCYAMFIGHATALIQSLDSSRRQYQEKYKQVEQYMSFHKLPADTRQRI
385	HEYYEHRYQGKMFDEESILGGAHMPLREEIINFTCRGLVAHMPLFAHADPSFVTAVLTKLRFEV
385	HEYYEHRYQGKMFDEESILG ELSE PLREEIINFTCRGLVAHMPLFAHADPSFVTAVLTKLRFEV
449	FOFGDLVVREGSVGRNGFDFOHGLLSVLLSNTNDTRLTDGSYFGEICLLTRGRRTASVRADTYC
449	FOFGDLVVREGSVGRKMYFIOHGLLSVLARGARDTRLTDGSYFGEICLLTRGRRTASVRADTYC
513	LGFLLSVDHFNAVGEEFFMMRRAFETVAMDMHAVNFGQQRGGQRKRSEFSFGSSGGIMESLFFN
513	RLYSLSVDHFNAVLEEFPMMRRAFETVAMDRLLRIGKKNSILORKRSEPSPGSSGGIMEOHLVO

Figure (14): Sequence alignment in sequence generated the 6thexperiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 78% of cases .and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 78%.

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from simulation program the figure(15a) clears the real sequences while the figure (15b) clear the figure(13a) generated sequences for the simulation program and the computer gave these results.

	10	20	30	40	50	60
1	MEAEQRPAAG	ASEGATPGLE	AVPPVAPPPA	TAASGPIPKS	GPEPKRRHLG	TLLQPTVNKF
61	SLRVFGSHKA	VEIEQERVKS	AGAWIIHPYS	DFRFYWDLIM	LLLMVGNLIV	LPVGITFFKE
121	ENSPPWIVFN	VLSDTFFLLD	LVLNFRTGIV	VEEGAEILLA	PRAIRTRYLR	TWFLVDLISS
181	IPVDYIFLVV	ELEPRLDAEV	YKTARALRIV	RFTKILSLLR	LLRLSRLIRY	IHQUEEIFHM
241	TYDLASAVVR	IFNLIGHMLL	LCHWDGCLQF	LVPMLQDFPP	DCWVSINHMV	NHSWGRQYSH
301	ALF <mark>KAMSH</mark> ML	CIGYGQQAPV	GMPDVWLTML	SMIVGATCYA	MFIGHATALI	QSLDSSRRQY
361	QEKYKQVEQV	MSFHKLPADT	RORIHEYYEH	RYQGKMFDEE	SILGELSEPL	REEIINFTCR
421	GLVAHMPLFA	HAD PSFVTAV	LTKLRFEVFQ	PGDLVVREGS	VGRKMYFIQH	GLLSVLARGA
481	RDTRLTDGSY	FGEICLLTRG	RRTASVRADT	YCRLYSLSVD	HFNAVLEEFP	MMRRAFETVA
541	MDRLLRIGKK	NSILQRKRSE	PSPGSSGGIM	EQHLVQHDRD	MARGVRGRAP	STGAQLSGKP
601	VLW <mark>e</mark> plv <mark>h</mark> ap	LQAAAVTSNV	AIALTHQRGP	LPLSPDSPAT	LLARSAURSA	GSPASPLVPV
661	RAGPWASTSR	LPAPPARTLH	ASLSRAGRSQ	VSLLGPPPGG	GGRRLGPRGR	PLSASQPSLP
721	QRATGDGSPG	RKGSGSERLP	PSGLLAKPPR	TAQPPRPPVP	EPATPRGLQL	SANM

Figure (15a) real sequences

	10	20	30	40	50	60
	ليتتبينين	Jummed	human	human	Jummed	human
1	TCGGQRPAAG	ASEGATPGLE	LGPPVAPPPA	TAASGGLGSL	FPEPKRRHLG	TLLQPTVNKF
61	SLRVFGSHKA	VEIEQERVKS	AGAWIIHPYS	DFRFYUDLCN	GFGVLHCTYL	GPNTNGPIFT
121	GHCHFGGVGY	PNVNGTTLGP	FLLPPGQMLG	FGLRVPGNPV	TMTIRTRYLR	TWFLVDLISS
181	TGTLTIANII	GFE PRLDAEV	YKTAPD PTPT	TPGNGFSHFL	GLRLSRLIRY	IHQUEE FMGT
241	LPANLSAVVR	IFNLIGMMLL	LCHWDGCLQF	LVPMLQDFPP	DCWVSINHMV	NHSVGQGYSH
301	ALFKAMSHML	MTRGRLDGQH	GNGTHULTML	SMIVGATCYA	MFIGHATALI	QSLDSSRRQY
361	QEKYKQVEQY	MSFHKLPADT	RORIHEYYEH	RYQGKMFDEE	SILGGAHMPL	REEIINFTCR
421	GLVA <mark>H</mark> MPLFA	HAD PSFVTAV	LTKLRFEVFQ	PGDLVVREGS	VGRNG FD PQH	GLLSVLLSNT
481	NDTRLTDGSY	FGEICLLTRG	RRTASVRADT	YCLGFLLSVD	HFNAVGEEFP	MMRRAFETVA
541	HD MHAVN FGQ	QRGGQRKRSE	PSPGSSGGIM	ESLPPNNIFT	LANGVRGRAP	STGAQLSGKP
601	VLURGGVHAP	LQAAAVTSNV	AIALTHORGP	LPLSPDSPAT	LLMPNNSLSA	GSPASPLVPV
661	RAGPWASTSG	LPAPPARTLH	ASLSRAGRSQ	VSLLGPPPMM	VGPYPGPRGR	PLSASQPSLP
721	QRATEDESPE	RKGSGSERLT	PPGLLAKPPR	HAQPPRPPVC	FITAATQTQL	SANH

Figure (15b) generated sequences for the simulation program

Ls = 633

La = Lb = 774

SimlarityPerc = 82%

As the concluded percentage of two sequence similarity is 82%.so this is an evidence that both have a similar

physiochemical properties of about 82% and by this there will be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of 82%.

But when considering an amino acid H as a polar amino acid the computer gave these results:

Ls = 656

La = Lb = 774

SimlarityPerc =% 85

The similarity percentage when considering an amino acid H as a polar amino acid increase to be 85%.

7th experiment : When closing a sequence size N=888 in order to simulation a mitochondrial brain tissue disease protein which consist of 888 amino acids so the following sequence had been given by the program .

NNFLPPHVAPQNTFLDTIIRKFEGQSRKFIIANARVMGGDGNPG FLNGSGNGGYSRAEVMQRPCTCDFLHGPRTQRRAAAQIAQAL LGAEERKVEIAFYRKDGSCFLCLVDVGLHNGILYGQLNFFNN GHFGLGPTPYGNNDLIFNTQPTLFGNRPMGLYNAPGGKLPALL ALTARESSVRSGGAGGAGAGAPGAVGVNVDLTPAAPSSESLALD EVTAMDNHVAGLGPAEERRAFLNFANSFGVVNNIQYTYSAHS LNPDASGSSCSLIALTSPVTGMLRRSGGNACMGNTNGNYRTG PTIPHASTGAMHPLRSLTGTGMALTQGAGPTPMFPYTGTMTN FVDLKGDPFLASPTSDRELMAPKIKERTHNVTEKVGHNCLYH NLSPTTNSGNQHNISGLHGGFGSPFKAMGSGNFGVTVIYTAVF TPYSAAFLLKETEEGPPATECGYACQIFGNHFVFFGNMFIVDIL INFRTTYVNIPTVNVSHPGRIAVHYFKGWFLIDMVAAIPFDLLI FGSGSEELIGLLKTARLLRLVPQARKLDRYSEYGAAVLFLLMC TFALIAHWLACIWYAIGNMEOPHMDSRIGWLHNRNGYOGKP **YNSSGLGGPFLOFPALLNLYFTFSSLTSVGFGNVSPNTNVTTTP** GGFNNLIGSLMYASIFGNVSANMNGPYSGTARYHTQMLRNTG NLGFHQIPNPLRQRLEEYFGAFGPNSNGIDMNAVLKGGRSYN VCQFGLHLNRSLLQHCKPFRGATKYMLRALAMKFKTTHAPP GDTLVHAGDLLTALYFISRGSIEILRGDVVVAILGMGWGAGTG LEMPSAAFRGASLLNMQSLGLWTWDCLQGHWAPLIHLNSGP PSGAMERNHTWGEAAELWGSHNOGVGNGRHKQTLFASLK

A - when a generated sequence compared with a real protein sequence of (888) amino acids size following results has obtained :



Figure (16) Sequence alignment in sequence generated the 7thexperiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 70% of cases and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 70%.

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from

simulation program the figure(17a) clears the real sequences while the figure (17b) clear the generated sequences for the simulation program

	10	2.0	30	40	50	60
1	MPVRRGHVAP	QNTFLDTIIR	KFEGQSRKFI	IANARVENCA	VIYCNDGFCE	LCGYSRAEVH
61	QRPCTCDFLH	GPRTORRAAA	QIAQALLGAE	ERKVEIAFYR	KDGSCFLCLV	DVVPVKNEDG
121	AVINFILNEE	VVHEKDHVGS	PAHDTNHRGP	PTSULAPGRA	KTFRLKLPAL	LALTARESSV
181	REGGAGGAGA	PGAVVVDVDL	TPAAPSSESL	ALDEVTAMDN	HVAGLGPAEE	RRALVGPGSP
241	PRSAPGOLPS	PRAHSLNPDA	SGSSCSLART	RSRESCASVR	RASSADDIEA	MRAGVLPPPP
301	RHASTGANHP	LRSGLLNSTS	DSDLVRYRTI	SKIPQITLNF	VDLKGDPFLA	SPTSDREIIA
361	PKIKERTHNV	TERVIQUESE	GADVLPEYKL	QAPRIHRWTI	LHYSPFKAVU	DULILLUIY
421	TAVFTPYSAA	FLLKETEEGP	PATECGYACQ	PLAVVDLIVD	IMFIVDILIN	FRTTYVNANE
481	EVVSHPGRIA	VHYFKGUFLI	DMVAAIPFDL	LIFGSGSEEL	IGLLKTARLL	RLVRVARKLD
541	RYSEYGAAVL	FLLHCTFALI	AHULACIUYA	I GNME Q PHMD	SRIGWLHNLG	DOIGKPYNSS
601	GLGGPSIKDK	YVTALYFTFS	SLTSVGFGNV	SPNTNSEKIF	SICVMLIGSL	MYASIFGNVS
661	AIIORLYSGT	ARYHTOMLRV	REFIRFHOIP	NPLRORLEEY	FOHAWSYTNG	IDHNAVLKGF
721	PECLQADICL	HLNRSLLQHC	KP FRGATKGC	LRALAMKFKT	THAPPGDTLV	HAGDLLTALY
781	FISRGSIEIL	RGDVVVAILG	MGUGAGTGLE	MPSAASRGAS	LLNNQSLGLW	TWDCLQGHWA
841	PLIHLNSGPP	SGARERSPTU	GEAAELUGSH	ILLPFRIRHK	OTLFASLK	
			(1 =			

figure (17a) real sequences

	10	20	30	40	50	60
						TTTTTTTTTT
1	NNFLPPHVAP	QNTFLDTIIR	KFEGQSRKFI	IANARVHGGD	GNPGFLNGSG	NGGYSRAEVM
61	QRPCTCDFLH	GPRTORRAAA	QIAQALLGAE	ERKVEIAFYR	KDGSCFLCLV	DVGLHNGILY
121	GQLNFFNNGH	FGLGPTPYGN	NDLIFNTQPT	LFGNRPHGLY	NAPGGKLPAL	LALTARESSV
181	REGGAGGAGA	PGAVGVNVDL	TPAAPSSESL	ALDEVTANDN	HVAGLGPAEE	RRAFLNFANS
241	FGVVNNIQYT	YSAHSLNPDA	SGSSCSLIAL	TSPVTGHLRR	SCGNACMONT	NGNYRTGPTI
301	PHASTGAMMP	LRSLTGTGHA	LTQGAGPTPM	FPYTGTHTNF	VD LKGD PFLA	SPTSDRELMA
361	PKIKERTHNV	TEKVGHNCLY	HNLSPTTNSG	NQHNISGLHG	GFGSPFKAHG	SGNFGVTVIY
421	TAVFTPYSAA	FLLKETEEGP	PATECGYACO	IFGNHFVFFG	NHFIVDILIN	FRTTYVNIPT
481	VNVSHPGRIA	VHYFKGWFLI	DNVAAIPFDL	LIFGSGSEEL	IGLLKTARLL	RLVPQARKLD
541	RYSEYGAAVL	FLLMCTFALI	AHWLACIWYA	I GNME Q PHMD	SRIGULHNRN	GYQGKPYNSS
601	GLGGPFLQFP	ALLNLYFTFS	SLTSVGFGNV	SPNTNVTTTP	GGFNNLIGSL	HYASIFGHVS
661	ANMNGPYSGT	ARYHTOMLRN	TGNLGFHQIP	NPLRORLEEY	FGAFGPNSNG	IDMNAVLKGG
721	RSYNVCQFGL	HLNRSLLQHC	KPFRGATKYM	LRALAMKFKT	THAPPGDTLV	HAGDLLTALY
781	FISRGSIEIL	RGDVVVAILG	MGWGAGTGLE	MP SAAFRGAS	LLNMQSLGLW	TWDCLQGHWA
841	PLIHLNSGPP	SGAMERNHTU	GEAAELWGSH	NQGVGNGRHK	QTLFASLK	
_						

figure (17b) generated sequences for the simulation program

and the computer gave these results.

Ls =682

La = Lb = 888

SimlarityPerc = 77%

As the concluded percentage of two sequence similarity is 77%.so this is an evidence that both have a similar physiochemical properties of about77% and by this there will be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of 77%.

But when considering an amino acid H as a polar amino acid the computer gave these results:

Ls =707

La = Lb = 888

SimlarityPerc = 80%

The similarity percentage when considering an amino acid H as a polar amino acid increase to be 80%.

8thexperiment :When closing a sequence size N=337 in order to simulation aprion protein which consist of 337 amino acids so the following sequence had been given by the program . TFSTASPAADGGRGPWEGGLVSWPPAPPLTLPWTWMGPSWG QHPGHWGFPALTDPSASPAASLGIFEVRYVLDASGCSMIFGPG GGAARFSSYLLSRARKVNGGLPLSPCGVPELCSISTSRATTGY GMGNMAAMVPFPPQRYHYFLVLDFEATCDKPQIHPQEIJEFPIL KLNLNPMEIESTFHMYVQPVVHPQLGTFCTELTGIIQAMTDGQ PSLQQVLERVDWMAKEGGLLDPNVKSIFVTCGDWDLKVMLP GQCHYLGLPADYFKQWINNLKKAYSFAMGWPKNGGCGDMN KGLSLQHIGRPHSGIDDCKNANIMGLGLNLYGVGQYQTSKPF

A - when a generated sequence compared with a real protein sequence of (337) amino acids size following results has obtained :

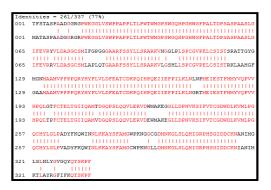


figure (18): Sequence alignment in sequence generated the 8thexperiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 77% of cases .and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 77%..

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from simulation program the figure(19a) clears the real sequences while the figure (19b) clear the generated sequences for the simulation program

	10	20	30	40	50	60
	human	hunni		hunner	human	hunning
1	MATASPAADG	GRGRPWEGGL	VSWPPAPPLT	LPWTWMGPSW	GQHPGHWGFP	ALTDPSASPA
61	ASLGIFEVRR	VLDASGCSML	APLQTGAAR F	SSYLLSRARK	VLGSHLLSPC	GVPELCSIST
121	RKLAAHGFGA	AMAAMVPFPP	QRYHYFLVLD	FEATCDKPQI	HPQEIIEFPI	LKLNGRTHE I
181	ESTFHMYVQP	WWHPQLTPFC	TELTGIIQAM	VDGQPSLQQV	LERVDEUMAK	EGLLDPNVKS
241	IFVTCGDWDL	KVMLPGQCHY	LGLPVADYFK	QUINLKKAYS	FANGCWPKNG	LLDMNKGLSL
301	QHIGRPHSGI	DDCKNIANIM	KTLAYR GFIF	KQTSKPF		



	10	20	30		50	60
1	TFSTASPAAD	GGRGPHEGGL		LPHTHMGPSH		ALTDPSASPA
	ASLGIFEVRY					
121	SRATTGYGNG	NMAAMVP FPP	QRYHYFLVLD	FEATCDKPQI	HPQEIIE FPI	LKLNLNPME I
181	ESTFHMYVQP	WWHPQLGTFC	TELTGIIQAM	TDGQPSLQQV	LERVDUMAKE	GGLLDPNVKS
241	IFVTCGDWDL	KVMLPGQCHY	LGLPA <mark>DYFKQ</mark>	WINNL <mark>KK</mark> AYS	FAMGWPKNGG	CGDMNKGLSL
301	QHIGRPHSGI	DDCKNANING	LGLNLYGVGQ	YQTSKPF		

Figure (19b) generated sequences for the simulation program

and the computer gave these results.

Ls =277

La = Lb = 337

SimlarityPerc =82%

As the concluded percentage of two sequence similarity is 82%.so this is an evidence that both have a similar physiochemical properties of about82% and by this there will be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of 82%.

But when considering an amino acid H as a polar amino acid the computer gave these results:

Ls =286

La = Lb = 337

SimlarityPerc =85%

The similarity percentage when considering an amino acid H as a polar amino acid increase to be 85%.

 9^{th} experiment :When closing a sequence size N=1196 in order to simulation a prion protein which consist of 1196 amino acids so the following sequence had been given by the program.

```
SVPKKFKLMNSFLDDOPKDPNLVASPFGGYFKNPAADAGSNN
ASKKSSYQQQRNWKQGGNYQQGGYQSYNSNYNNYNNYNN
YNNYNNYNNYNKYNGQGYQKSTYKQSAVTPNQSGTPTPSAS
TTSLTSLNEFNSAGSTSCISQFLSKITFGFGFTDCKNQIKAFFPG
NGGGGNSTGEKIEEWKIVFNGIFLFKPKNPSLVRESAMLIISNIA
QFFSGKPPQEASPLGGNNTFIIFNVMGTHTVKRAAQHAIDSLL
NCFPMEALTCFVLPTILDYLSSGAKWQAKMAALSVVDRIRED
SANDLLELTFKDAVPVLTDVATDFKPELAKQGYKTLLDYVSIL
DNLDLSPRYKLGGGNLQDPSKVPESVKSLSYTIGTNPLTEPSLS
LLVPILNRSLNLSSSSQEQLRQTGGVVENLTRLVNNRNEIESFIP
LLLPGIQKVVDTATRPEVRELANMNLNVLKEDDEADKENKFS
GRLTLEEGRDFLLDHLKDIKADDSCFVKPYMNLETVIKYMSKI
LTVDSNVNDWKRLEDFLTPGHLHRGTTFNGQPNGPQHNLRAL
FYQEKERADEDEGIEIVNTDANGLNRRVNGLNKTNLRLLKGH
RYGLCGRNGAGKSTLMRAIANGQLDGFPDKDTLRTCFVEHKL
QGEEGDLDLVSFIALDEESQSTSREEIAAALTLTIANIIGFFFSVG
SLSGGWKMKLELARAMLQKADILLLDEPGNHLDVSNVKWLE
GTLPAHTDITSLIVSHDSGFLDTVCTDIIHYENKKLAYYNPNLA
AFVEQKPEAKSYYTLTDSNAQMRPPPGILLTGVKSNTRAVAK
MTDVTFSYFLQGNHDDGGAMTNTTGPHGALNLGPNGAGKGL
PVKLLTGELVPNEGKVEKHPNLRTLLLQNVIPQHVNEHKEKT
ANQYLQWRYQFGDDREVLLKEMMTSDHMVRGMMTKEIDID
DGRGKRAIEAIVGRQKLKKSFPNGFFLCCALSNTNPNGTQMT
NLNLFGPAFLQKFDDHEFLYLGLGYRELLSGTYTKHFEDVGV
TSFATQTALPGMMHAGQLVKVVIAGAMWNNPHLLVLDEPTN
YLDRDSLGALAVAIRDWSGGVVMISHNNEFRGALCPEOWIVE
NGKMVQKGSAQVDQSHVGTDVNMARVLLLMPNNSLPSVDD
DTGPANIKVKNFTGPDTRNEKKLMAERRRLRYIEWLSSPKGTP
KPVDTGPYPV
```

A - when a generated sequence compared with a real protein sequence of (1196) amino acids size following results has obtained :

Ident 0001	ities = 942/1196 (79%) SVPKKFKLMNSFLDDQPKDPNLVASPFGGYFKNPAADAGSNNASKKSSYQQQRNWKQGGNYQQG
0001	MPPKKFKDLNSFLDDQPKDPNLVASPFGGYFKNPAADAGSNNASKKSSYQQQRNWKQGGNYQQG
0065	GYQSYNSNYNNYNNYNNYNNYNNYNKYNGQGYQKSTYKQSAVTPNQSGTPTPSASTTSLTS
0065	GYQSYNSNYNNYNNYNNYNNYNNYNNYNNYNKYNGOGYQKSTYKQSAVTPNOSGTPTPSASTTSLTS
0129	LNEFNSAGSTSCISQFLSKITFGFGFTDCKNQIKAFFPGNGGGGNSTGEKIEEWKIVFNGIFLF
0129	LNEKLSNLELTPISOFLSKIPECOSITDCKNOIKLIIEEFGKEGNSTGEKIEEWKIVDVLSKFI
0193	KPKNPSLVRESAMLIISNIAQFFSGKPPQEASPLGGNNTFIIFNVMGTHTVKRAAQHAIDSLLN
0193	KPKNPSLVRESAMLIISNIAQFFSGKPPQEAYLLPFFNVALDCISDKENTVKRAAQHAIDSLLN
0257	CFPMEALTCFVLPTILDYLSSGAKWQAKMAALSVVDRIREDSANDLLELTFKDAVPVLTDVATD
0257	CFPMEALTCFVLFTILDYLSSGAKWQAKMAALSVVDRIREDSANDLLELTFRDAVPVLTDVATD
0321	FKPELAKQGYKTLLDYVSILDNLDLSPRYKLGGGNLQDPSKVPESVKSLSYTIGTNPLTEPSLS
0321	FKPELAKQGYKTLLDYVSILDNLDLSPRYKLIVDTLQDPSKVPESVKSLSSVTFVAEVTEPSLS
0385	LLVPILNRSLNLSSSSQEQLRQTGGVVENLTRLVNNRNEIESFIPLLLPGIQKVVDTATRPEVR
0385	LLVPILNRSLNLSSSSQEQLRQTVIVVENLTRLVNNRNEIESFIPLLLPGIQKVVDTASLPEVR
0449	ELANMNLNVLKEDDEADKENKFSGRLTLEEGRDFLLDHLKDIKADDSCFVKPYMNLETVIKYMS
0449	ELAEKALNVLKEDDEADKENKFSGRLTLEEGRDFLLDHLKDIKADDSCFVKPYMNDETVIKYMS
0513	KILTVDSNVNDWKRLEDFLTPGHLHRGTTFNGOPNGPOHNLRALFYOEKERADEDEGIEIVNTD
0513	KILTVDSNVNDWKRLEDFLTAVFGGSDSQREFVKQDFIHNLRALFYQEKERADEDEGIEIVNTD

Figure (20): Sequence alignment in sequence generated the9th experiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 79% of cases .and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 79%.

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from simulation program the figure(21a) clears the real sequences

while the figure (21b) clear the generated sequences for the simulation program

	10	20	30	40	50	60
1	MPPKKFKDLN	SFLDDQPKDP	NLVASPFGGY	FKNPAADAGS	NNASKKSSYQ	QQRNWKQGGN
61	YQQGGYQSYN	SNYNNYNNYN	NYNNYNNYNN	YNKYNGQGYQ	KSTYKQSAVT	PNQSGTPTPS
121	ASTTSLTSLN	EKLSNLELTP	ISQFLSKIPE	CQSITDCKNQ	IKLIIEEFGK	EGNSTGERIE
181	EWKIVDVLSK	FIKPKNPSLV	RESAMLIISN	IAQFFSGKPP	QEAYLLPFFN	VALDCISDKE
241	NTVKRAAQHA	IDSLLNCFPM	EALTCFVLPT	ILDYLSSGAK	WQAKMAALSV	VDRIRED SAN
301	DLLELTFKDA	VPVLTDVATD	FKPELAKQGY	KTLLDYVSIL	DNLDLSPRYK	LIVDTLQDPS
361	KVPESVKSLS	SVTFVAEVTE	PSLSLLVPIL	NRSLNLSSSS	QEQLRQTVIV	VENLTRLVNN
421	RNEIESFIPL	LLPGIQKVVD	TASLPEVREL	AEKALNVLKE	DDEADKENKF	SGRLTLEEGR
481	DFLLDHLKDI	KADDSCFVKP	YMNDETVIKY	MSKILTVDSN	VND WKRLED F	LTAVFGGSDS
541	QREFVKQDFI	HNLRALFYQE	KERADEDEGI	EIVNTDFSLA	YGSRMLLNKT	NLRLLKGHRY
601	GLCGRNGAGK	STLMRAIANG	QLDGFPDKDT	LRTCFVEHKL	QGEEGDLDLV	SFIALDEELQ
661	STSREEIAAA	LESVGFDEER	RAQTVGSLSG	GWKMKLELAR	AMLQKADILL	LDEPTNHLDV
721	SNVKWLEEYL	LEHTDITSLI	VSHDSGFLDT	VCTDIIHYEN	KKLAYYKGNL	AAFVEQKPEA
781	KSYYTLTDSN	AQMRFPPFGI	LTGVKSNTRA	VAKMTDVTFS	YPGAQKPSLS.	HVSCSLSLSS
841	RVACLGPNGA	GKSTLIKLLT	GELVPNEGKV	EKHPNLRIGY	IAQHALQHVN	EHKEKTANQY
901	LQWRYQFGDD	REVLLKESRK	ISEDEKEMMT	KEIDIDGDRG	KRAIEAIVGR	QKLKKSFQYE
961	VKWKYWKPKY	NSUVPKDVLV	EHGFEKLVQK	FDDHEASREG	LGYRELIPSV	ITKHFEDVGL
1021	DSEIANHTPL	GSLSGGQLVK	VVIAGANUNN	PHLLVLDEPT	NYLDRDSLGA	LAVAIRDUSG
1081	GVVMISHNNE	FVGALCPEQU	IVENGKMVQK	GSAQVDQSKF	EDGGNADAVG	LKASNLAKPS
1141	VDDDDSPANI	KVKQRKKRLT	RNEKKLQAER	RRLRYIEWLS	SPKGTPKPVD	TDDEED

figure (21a) real sequences

	10	20	30	40	50	60
						OORNWKOGGN
1	SVP <mark>KKFKLMN</mark>	SFLDDQPKDP	NLVASPFGGY	FENPAADAGS	NNASKKSSYQ	
61	YQQGGYQSYN	SNYNNYNNYN	NYNNYNNYNN	ANKANCÓCAŐ	KSTYSLMGLT	PNQSGTPTPS
121	ASTTSLTSLN	EFNSAGSTSC	PYFGHQCTTF	GFGFSTTFPG	GGAFFPGNGG	GGNSTGEKIE
181	EWKIVFNGIF	LFFPKNPSLV	RESAMLIISN	IAQFFSGKPP	QEASPLGGNN	TFIIFNVMGT
241	HPGGFSLMLF	TDSLLNCFPM	EALTCFVLPT	ILDYLSSGAK	UQAKMAALSV	VDRIRED SAN
301	DLLELTFKDA	VPVLTDVATD	FKPELAKQGY	KTLLDYVSIL	DNLDLSPRYK	LGGGNLQDPS
361	KVPESVKSLS	YTIGTNPLTE	PSLSLLVPIL	NRSLNLSSSS	QEQLEQTEGV	VENLTRLVNN
421	RNEIESFIPL	LLPGIQKVVD	TATRLTNPPT	FNMNLNVLKE	DDEADKENKF	SGRLTLEEGR
481	DFLLDHLKDI	KADDSCFVKP	YMNLETVIKY	MSKILTVDSN	VNDWKRLEDF	LTPGHLHRGT
541	TFNGQPNGPQ	HNLRALFYQE	KERADEDEGI	EIVNTDANGL	NRRVNGLNKT	NLRLLKGHRY
601	GLCGRNGAGK	STLMRAIANG	QLDGFPDKDT	LRTC FVEHKL	QGEEGDLDLV	SFIALDEESQ
661	STSREEIAAA	LTLTIANIIG	FFFSMGLGSG	GUKHKLELAR	AMLQKADILL	LDEPGNHLDV
721	SNVKWLEGTL	PAHTDITSLI	VSHDSGFLDT	VCTDIIHYEN	KKLAYYNPNL	AAFVEQKPEA
781	KSYYTLTDSN	AQMRPPPGIL	LTGVKSNTRA	VAKHTDVTFS	YFLQGNHDDG	GAMTNTTGPH
841	GALNLGPNGA	GKGLPVKLLT	GELVPNEGKV	EKHPNLRTLL	LQNVIPQHVN	EHKEKTANQY
901	LQWRYQFGDD	REVILKEMMT	SDHMVRGMMT	KEIDIDDGRG	KRAIEAIVGR	QKLKKSFPNG
961	FFLCCALSNT	NPNGTQMTNL	NLFGPAFLFG	GGLSGFLYLR	GALGFLLSGT	YMNNFGLGGV
1021	TSFATQTALP	GMMHAGQLVK	VVIAGANUNN	PHLLVLDEPT	NYLDRDSLGA	LAVAIRDWSG
1081	GVVMISHNNE	FRGALCPEQU	IVENGKNVQK	GSAQVDQSHV	GTD VNMARVL	LLMPNNSLPS
1141	VDDDTGPCDG	MGRNFTGPDT	RNEKKLMAER	RRLRYIEWLS	SPKGTPKPVD	TGPYPV

figure (21b) generated sequences for the simulation program

and the computer gave these results .

Ls = 1006

La = Lb = 1196

SimlarityPerc = 84%

As the concluded percentage of two sequence similarity is 84%.so this is an evidence that both have a similar physiochemical properties of about84% and by this there will be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of84%.

But when considering an amino acid H as a polar amino acid the computer gave these results:

Ls = 1025

La = Lb = 1196

SimlarityPerc = 86%

The similarity percentage when considering an amino acid H as a polar amino acid increase to be 86%.

10ndexperiment :When closing a sequence size N=405 in order to simulation a prion protein which consist of 405 amino acids so the following sequence had been given by the program.

MDTDKLISEAESHFSQGNHAEAVAKLTSAAQSNPNDEQMSTIE SLIQKIAGYVMDNRSGGSDASQDRAAGGGSSFMNTLMADSK GSSQTQLGKLALLATVMTHSSNKGSSNRGFDVGTVMSMLSGS GGGSQSMGASGLAALASQFFKSGNNSQGQGQGQGQGQGQG QGQGQGSFTALASLASSFMNSNNNNQQGQNQSSGGSSFGALA SMASSFMHSNNNQNSNNSQQGYNQSYQNGNQNSQGYNNQQ YQGGNGGYQQQQGQSGAFSSLASMAQSYLGGGQTQSNQQ QYNQQGQNNQQQYQQGQNYQHQQQGQQQQGHSSSFSAL ASMASSYLGNNSNSNSSYGGQQQANEYGRPQQNGQQQSNEY GRPQYGGNQNSNGQHESFNFSGNFSQQNNNGNQNRY

A - when a generated sequence compared with a real protein sequence of (405) amino acids size following results has obtained :

Iden	tities = 312/405 (77%)
001	NLTGNGGSNCESHFSQGNHANLSPKLTSAAQSNPNDEQMSTIESLIQKIAGYVMDNRSGGSDAS
	1 I IIIIIIII IIIIIIIIIIIIIIIIIIIIIIIII
001	MDTDKLISEAESHFSQGNHAEAVAKLTSAAQSNPNDEQMSTIESLIQKIAGYVMDNRSGGSDAS
001	ADIDKETSERESHESCONNRERVARITSRACSNEHEECHSTTESTICKTROTVADARSOSSDAS
065	QDRAAGGGSSFMNTLMAGRFTFIFLQLGKLALLATNTSHSSNKGSSNRGFDVGTVMSMLSGSGG
065	QDRAAGGGSSFMNTLMADSKGSSQTQLGKLALLATVMTHSSNKGSSNRGFDVGTVMSMLSGSGG
129	GSOSMGPNTLOCHFLOFFKSGNNSOGOGOGOGOGOGOGOGOGOGSFTALASFFDPSFMNSNNNN
129	GSOSMGASGLAALASOFFKSGNNSOGOGOGOGOGOGOGOGOGOGOGOGSFTALASLASSFMNSNNNNO
193	QQGQN55GG55FGALASMA55FMH55NNNQNLMPTLRNGNQYQNGNQN5QGYNNQQYQGGNGGY
	1 IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII
193	QGQNQSSGGSSFGALASMASSFMHSNNNONSNNSQQGYNQSYQNGNQNSQGYNNQQYQGGNGGY
257	0000GG05GGAF55LA5M05YLGGG0T05N000YYN000NN000Y000G0NY0H000G00000G
257	OOOOGOSGGAFSSLASMAOSYLGGGOTOSNOOOYNOOGONNOOOYOOOGONYOHOOOGOOOOOG
201	0000003034133143140311333010380000110003011000001000000000
321	HSSSFSALASMASSYLGNNSNSNSSYGGQQQANMVGRPQQNGQQQSOFGLLPTNFLQATGNNLV
321	HSSSFSALASMASSYLGNNSNSNSSYGGOOOANEYGRPOONGOOOSNEYGRPOYGGNONSNGOH
385	TSFNFSGNFSQQNNPGNQNRY
385	ESFNFSGNFSOONNNGNONRY

figure (22) Sequence alignment in sequence generated the 10th experiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 77% of cases .and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 77%.

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from simulation program the figure (23a) clears the real sequences while the figure (23b) clear the generated sequences for the simulation program.

	10	20	30	40	50	60
	ليتتبيني	human	human	human	hunner	hunning
1	MDTDKLISEA	ESHFSQGNHA	EAVAKLTSAA	QSNPNDEQMS	TIESLIQKIA	GYVHDNRSGG
61	SDASQDRAAG	GGSSFHNTLM	ADSKGSSQTQ	LGKLALLATV	MTHSSNKGSS	NRGFDVGTVM
121	SMLSGSGGGS	QSMGASGLAA	LASQFFKSGN	NSQGQGQGQG	0000000000	QGSFTALASL
181	ASSFHNSNNN	NQQGQNQSSG	GSSFGALASM	ASSFMHSNNN	QNSNNSQQGY	NQSYQNGNQN
241	SQGYNNQQYQ	GGNGGYQQQQ	GQSGGAFSSL	ASMAQSYLGG	GQTQSNQQQY	NQQGQNNQQQ
301	YQQQGQNYQH	0000000000	HSSSFSALAS	MASSYLGNNS	NSNSSYGGQQ	QANEYGRPQQ
361	NGQQQSNEYG	RPQYGGNQNS	NGQHESFNFS	GNFSQQNNNG	NQNRY	

figure (23a) real sequences

	10	20	30	40	50	60
1	NLTGNGGSNC	ESHFSQGNHA	NLSPKLTSAA	QSNPNDEQNS	TIESLIQKIA	GYVMDNRSGG
61	SDASQDRAAG	GGSSFMNTLM	AGRFTFIFLQ	LGKLALLATN	TSHSSNKGSS	NRGFDVGTVM
121	SMLSGSGGGS	QSMGPNTLQC	HFLQFFKSGN	NSQGQGQGQG	0000000000	QGSFTALASF
181	FDPSFMNSNN	NNQQGQNSSG	GSSFGALASM	ASSEMHSSNN	NQNLMPTLRN	GNQYQNGNQN
241	SQGYNNQQYQ	GGNGGYQQQQ	GGQSGGAFSS	LASMQSYLGG	GQTQSNQQQY	YNQOQNNQQQ
301	YQQQGQNYQH	0006000006	HSSSFSALAS	MASSYLGNNS	NSNSSYGGQQ	QANMVGRPQQ
361	NGQQQSQFGL	LPTNFLQATG	NNLVTSFNFS	GNFSQQNNPG	NQNRY	

figure (23b) generated sequences for the simulation program

and the computer gave these results

Ls=350

La = Lb = 405

SimlarityPerc=86%

As the concluded percentage of two sequence similarity is 86%.so this is an evidence that both have a similar physiochemical properties of about86% and by this there will be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of86%.

But when considering an amino acid H as a polar amino acid the computer gave these results:

International Journal of Computer Applications (0975 – 8887) Volume 141 – No.13, May 2016

Ls=356

La = Lb = 405

SimlarityPerc=88%

The similarity percentage when considering an amino acid H as a polar amino acid increase to be 88%.

11ndexperiment :When closing a sequence size N=254 in order to simulation a prion protein which consist of 254 amino acids so the following sequence had been given by the program.

QNNLGYWLLALFVTTCTDVGLCKKRPKPGGWNTGGSRYP GQGSPGGNRYPPQSGGTWGQPHGGGWGQPHGGGWGQP HGGGWGQPHGGGWSQGGGTHNQWNKPSKPKTNLKHVA GAAAAGAVVGGGMFVNAGGQPLAVFFDGGPFWEDRYYR ENMYRYPNQFNNNPVDQYSNQNNFVHDCVNITIKQHTVTT TTKGENFTETGGPGAFGGMPFNQVTQYQKESQAYYDGRR SSAVLFSSPFLNTNVSVFFVGLGL

A - when a generated sequence compared with a real protein sequence of (254) amino acids size following results has obtained :

Ident	titles = 203/254 (80%)
001	QNNLGYWLLALFVTTCTDVGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQSGGTWGQPHGGG
001	MANLGYWLLALFVTTCTDVGLCKKRPKPGGWNTGGSRYPGQGSPGGNRYPPQSGGTWGQPHGGG
065	WGQPHGGGWGQPHGGGWGQPHGGGWSQGGGTHNQWNKPSKPKTNLKHVAGAAAAGAVVGGGMFV
065	WGQPHGGGWGQPHGGGWGQPHGGGWSQGGGTHNQWNKPSKPKTNLKHVAGAAAAGAVVGGLGGY
129	NAGGOPLAVFFDGGPFWEDRYYRENMYRYPNOFNNNPVDQYSNONNFVHDCVNITIKOHTVTTT
129	MLGSAMSRPMLHFGNDWEDRYYRENMYRYPNQVYYRPVDQYSNQNNFVHDCVNITIKQHTVTTT
193	TKGENFTETGGPGAFGGMPFNQVTQYQKESQAYYDGRRSSAVLFSSPFLNTNVSVFFVGLGL
193	TKGENFTETDVKMMERVVEQMCVTQYQKESQAYYDGRRSSAVLFSSPPVILLISFLIFLIVG

Figure (24): Sequence alignment in sequence generated the 11th experiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 80% of cases .and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 80%.

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from simulation program the figure(25a) clears the real sequences while the figure (25b) clear the generated sequences for the simulation program

	10	20	30	40	50	60
1	MANLGYULLA	LFVTTCTDVG		WNTGGSRYPG	QGSPGGNRYP	PQSGGTWGQP
61	HGGGWGQPHG	GGWGQPHGGG	NGQP <mark>H</mark> GGGWS	QGGGTHNQUN	KPSKPKTNLK	HVAGAAAAGA
121	VVGGLGGYML	GSAMSRPML <mark>H</mark>	FGNDWEDRYY	RENMYRYPNQ	VYYRPVDQYS	NQNNFVHDCV
181	NITIKQHTVT	TTTKGENFTE	TDVKMMERVV	EQMEVTQYOK	ESQAYYDGRR	SSAVLFSSPP
241	VILLISFLIF	LIVG				
	figure(25a) real sequences					

	10	20	30	40	50	60
	أسسسا			l		لسسس
1	QNNLGYWLLA	LFVTTCTDVG	LCKKRPKPGG	WNTGGSRYPG	QGSPGGNRYP	PQSGGTWGQP
61	HGGGWGQPHG	GGWGQP <mark>H</mark> GGG	WGQP <mark>H</mark> GGGWS	QGGGTHNQWN	KPSKPKTNLK	HVAGAAAAGA
121	VVGGGMFVNA	GGQPLAVFFD	GGP FW <mark>EDR</mark> YY	RENMYRYPNQ	FNNNPVDQYS	NQNNFVHDCV
181	NITIKQHTVT	TTTKGENFTE	TGGPGAFGGM	PFNQVTQYQK	ESQAYYDGRR	SSAVLFSSPF
241	LNTNVSVFFV	GLGL				
6.	(251	>	4 1	e		1 4*

figure (25b) generated sequences for the simulation program

and the computer gave these results

Ls = 215

La = Lb = 254

SimlarityPerc = 85%

As the concluded percentage of two sequence similarity is 85%.so this is an evidence that both have a similar physiochemical properties of about85% and by this there will be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of 85%.

But when considering an amino acid H as a polar amino acid the computer gave these results:

Ls = 223

La = Lb = 254

SimlarityPerc = 88%

The similarity percentage when considering an amino acid H as a polar amino acid increase to be 88%.

 12^{nd} experiment :When closing a sequence size N=151 in order to simulation a prion protein which consist of 151 amino acids so the following sequence had been given by the program.

GNWAPHSNWALLLAAAFLCDSGAAKGGRGGARGSARGG VRGGARGASRVRVRPAQRYGAPGSSLRVAAAGAAAGAAA AGAAAGGLPSGWRRAAGPGERLGLEDEEGVPGGNGTGPG IYSGRAWTPQFTPTRGPRLCLVLGGAFFTVGLLRP

A - when a generated sequence compared with a real protein sequence of (151) amino acids size following results has obtained :

Iden	tities = 125/151 (83%)
001	GNWAPHSNWALLLAAAFLCDSGAAKGGRGGARGSARGGVRGGARGASRVRVRPAORYGAPGSSL
001	MNWAPATCWALLLAAAFLCDSGAAKGGRGGARGSARGGVRGGARGASRVRVRPAQRYGAPGSSL
065	RVAAAGAAAGAAAGAAAGGLPSGWRRAAGPGERLGLEDEEGVPGGNGTGPGIYSGRAWTPOFT
003	
065	RVAAAGAAAGAAAGAAAGLAAGSGWRRAAGPGERGLEDEEDGVPGGNGTGPGIYSYRAWTSGAG
129	PTRGPRLCLVLGGAFFTVGLLRP
129	PTRGPRLCLVLGGALGALGLLRP

Figure (26):Sequence alignment in sequence generated the 12th experiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 83% of cases .and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 83%..

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from simulation program the figure(27a) clears the real sequences while the figure (27b) clear the generated sequences for the simulation program



Figure (27b) generated sequences for the simulation program

and the computer gave these results.

Ls = 137

La = Lb = 151

SimlarityPerc = 91%

As the concluded percentage of two sequence similarity is 91%.so this is an evidence that both have a similar physiochemical properties of about91% and by this there will be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of 91%.

But when considering an amino acid H as a polar amino acid the computer gave these results:

Ls = 137

La = Lb = 151

SimlarityPerc = 91%

The similarity percentage when considering an amino acid H as a polar amino acid increase to be 91%.

13 the xperiment : When closing a sequence size N=89 in order to simulation a prion protein which consist of 89 amino acids so the following sequence had been given by the program .

PQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHG GGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGW GQPHGGGWGQPHGGGWGQ

A - when a generated sequence compared with a real protein sequence of (89) amino acids size following results has obtained :

	ntities = 74/89 (83%)
01	FHGGGGWGQPHGGGWGQPHGGGWGGFHGGGWGQPHGGGWGQGGGWGQPHGGTPNQPHGGGQ
01	PQGGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWGQPHGGGWG
65	OPHGGGGGOPH PHGWGOPHGGGWGF
65	QPHGGGWGQPHGGGWGQPHGGGWGQ

Figure (28): Sequence alignment in sequence generated the 13th experiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 83% of cases and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 83%.

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from simulation program the figure(29a) clears the real sequences while the figure (29b) clear the generated sequences for the simulation sequences

	10	20	30	40 l	50 l	60
1						HGGGWGQPHG
61	GGWGQP <mark>H</mark> GGG	WGQP <mark>H</mark> GGGWG	QPHGGGWGQ			

Figure (29a) real sequences

	10	20	30	40	50	60
	ليتتبين		munul	hunnel	human	muuul
1	FHGGGGWGQP	HGGGWGQPHG	GGWGGFHGGG	WGQP <mark>H</mark> GGGWG	WGQGGGWGQP	HGGTPNQPHG
61	GGQGQPHGGG	GGQPHPHGWG	QPHGGGWGF			

Figure (29b) generated sequences for the simulation program

and the computer gave these results.

La = Lb = 89

SimlarityPerc = 81%

As the concluded percentage of two sequence similarity is 81%.so this is an evidence that both have a similar physiochemical properties of about81% and by this there will be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of 81%.

But when considering an amino acid H as a polar amino acid the computer gave these results:

$$Ls = 83$$

La = Lb = 89

SimlarityPerc = 93%

The similarity percentage when considering an amino acid H as a polar amino acid increase to be 93%.

 $14^{th}experiment$: When closing a sequence size N=2202 in order to prion protein which consist of 2202amino acids so the following sequence had been given by the program.

PHRCNFONSLNGTGNPNFOPVOTNAGGFGHOMAOTGAAAAA AATGQYYNPMLQQQQYLNFGFGMNYNNLFDFQAQQQQQQ YLMQQQQQQQLHHQQQQQHQNIAAQPQAQHHQMNMFTQ HQMLQLMQQQQQQQQQVQQIQRQQPIAFFPLSTGPTHPGRT GTPTGTVFNSHAIFMTCVISSHQKLYEEQCRQIEKERKEQEERK RKQELEEQRKRNETPGGGINITTTYTNGPGNDVNVHMNRERL AEIKRLEEATAFGIPMILTDQFGMGPQAMLQKMQGEQNKHIA EVERORSELHGTFMFLMILGLLVGTHFGFNFLDMIPFPYESMV DSTLPOVFDMERDSALNASGLPAGGRPLPGGONHLNQLFRGP PHNGVLKALLKLQTIKYFTYLFSSSLTRMDKNRHDDKETNDL FLDKLPPIIQAVVNYNTSALDVDSHNDNSQFIGNFVMMTEDIT RTTATMTSSSSYNNHHQNSIVMMTSSSVSMSERNPMNLVTMN HHDVDEEFPAPISIEKRRQMNVFLAPVKAGGGGGQNQRKKRD **MVENLYDNLTDNFVPTDTGRRGRRGRGSDDDEDELLQVNV** TQIEEMEKGVKLPASGVTGFTTTEDVQHFFGSQGFGRRKEDR RKDRSPTPEDVIESRDAEWOERLRLKMEREPSRKADEESONA WSPQALADNETFTRFCQTVDGVLEQGDSLDTELKMPKNKKR RSGGDHHHKKGTDFMESDEEEEMDEIDPDLRIELYYILEELRR GSARLRENALQAVGADKLLKLVAMLLDRNIRTISADNORLLV PCDDDVDVGDVLEKEICEERVKRAGFAAVVALNIMSSHRMH KQVIIEDVIDRCVGLTRLLLIHLIYPASDSIYKSVNSKKKDRAM PTGRRRKKAGVCTRDKFSEYIYERITEAIGLLAVLVKSESMLTT GGTNAGSVALTPFFVANVGSLQIVYMYSASNIFSRAEDSLRFS MITDLLSSLHRAPQFGCRNIGLQAHGPDGSWISTTTALFIQLVQ STIKIPKFKGHADEDELAKRSKKEEAMVREGFLKASKVTNAFL NGFLAKCSOKGNKMDGEEDYTFTPGFLNOELLSAHRSPEWPA AEMILTALGSLLVKNFRSKSSDMTIRQASLDYLGNITAKLRKD QKEAIAGERRLDAVVKKSFLLLSDKGVEDYESVDISNLKQND **KLKVLETSLIDYLVITNSSDIIVYACNFYVGFPMYEVAEDLESA** RSKLKOTVDTNESEKDVKKAERKYEKIQYRGAEMKVFLSKIL DKPFLKRRLEKSNKVKMLDSDAFWAVKFLAQSREFTQSFDTY LKHIVFGAGSETIVALRSKALKCLSSTAVFDSSVLILEDVQQAV HTRMVDSHAQVRESAVELIGRFVLYDMVLGGGYSQIAERILD TGVAVDLPVIRIMREICEKFPTFEMIPDMLARMTPRVTDEEGV KKLVFETFTLLWFQPVDTRIYTNAVGTFTITMCSVAQHCIKDA MSDYLEQLHLLLNPGFTFGSGMSVANRQIIDSLVDHILNLEQH

KSSGMFLEVELMRRKEQEEKYMAYLSTLAVFSKGARLLLTSH VEVLLPYLTFSGAKTNAENQVTKCMIGMLERVIPLVPFGDFYV LDSIDENLCKVIHSLDMALVVNPVSCVASIYKKFKRGATKTID VFSTYLKHLEVFKRNFDSNPRYDLQYGFFPILSRSIFTLGVLSR YGQFEEFVKEDPTEEKVEASPNFALLHGGHNSRYHKGGLRQK ALTAMGHFCAQHSTYLTKRQLTNTYFTPGNAANSPQQQQQR LLVLQNLEMFLQCEEQKLAASHDKWDENKEAQNLKEMELSG SGLGSSVIQYWKAVLESYYVDADIQLRRAAQVVWLTLNQGL GFMPGASIPTLIAMTTDPVDVIRNRIDILKEIIDSKYSGMVQSKA MQGVRLSYKLHLKLMLTQQEKFVRGFRCDDFHLNTLPNALPE THDGMAVTLSGLYQSLRTNRQQRRFFLQSMVKLFSEFFSHDK PQLMEYIFIADNLAMFPYQMDENGNLGNMQIDQNIAQTGQSL LVQYKLQLRMQESEDEDIVFLDENMMSRLSQLGQIETFYLFLD SQVPSSLLLYVRTFMHNLYGFNETKVAEYQPSEAAKVYEKAV TNTOIHMFKPITALEALNFPFEWGSFOHTGTMFGGGPOGRKM LLSLDQVEEVEVSNTTAANDDYDEEEDGGEDQNNGFMGPGM HH

A - when a generated sequence compared with a real protein sequence of (2202) amino acids size following results has obtained :



figure(30):Sequence alignment in sequence generated the 14th experiment

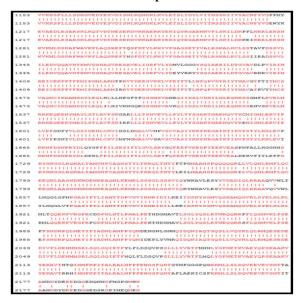


figure (30)-follow:Sequence alignment in sequence generated the 14th experiment

And us it is cleared from the previous figure the matching percentage between a real sequences cases and the generated sequences cases from simulation program reach to 80% of cases .and through the presence of such a match between the primary protein composition so it is possible to predict the position and the crop of the 2ndary protein composition which generational from the simulation program with success percentage of 80%.

B /regarding the similarity percentage between the real sequence cases and the generated sequence cases from simulation program the figure(33a) clears the real sequences while the figure (33b) clear the generated sequences for the simulation program

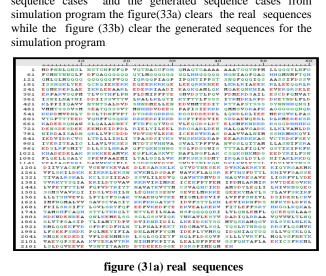


figure (31a) real sequences

	10		30		50	60
1	PHRCNFONSL	NGTGNPNFQP	VOTNAGGFGH	OMAQTGAAAA	AAATGQYYNP	MLQQQQYLNF
61	GFGMNYNNLF	DFQAQQQQQ	GATW000000	00THH00000	HONIAAOPOA	QHHQMNMFTQ
151	HOMLOLMOOO	000000PV00	IORGOPIAFF	PLSTGPTHPG	RTGTPTGTVF	NSHAIFMTCA
181	GNAHOKLYEE	QCRQILGLAP	NLGTFFPVME	LEEQRKRNET	PEGGINITTT	YTNGPGNDVN
241	VHMNNQSPPA	QTGQFNATAF	GIPMILTDQF	GHCPQAMLQK	MQGEQNEHIA	EVERQRSELH
301	GTFMFLMILG	LNNNLAHGFN	FLDMIPFPYE	SMVDSTLPQV	FDMERDSALN	ASGLPAGGRP
361	LPGGQNHLNQ	LFRGPPHNGI	HKALLKLQTI	RYFTYLFSSS	LTRMDENRHD	DEETNDLFLD
421	KLPPIIGAVV	NYNTSALDVD	SHNDNSOFIG	NFVMMTEDIT	RTTATHTSSS	SYNNHHONSI
401	VMMTSSSVSM	SERNPHNLVT	MNIGHDVDEEF	PAPISIEKRR	OMNVFLAPVK	ACCCCCCNOR
541	KKRDNVENLY	DNLTDNFVPT	DIGRRGRRRG	RGSDDDEDEL	LOVNVTQIEE	MERGVELPAS
601	GVTGFTTTED	VONFFGSOGF	GRRKEDRRKD	RSPTPEDVIE	SRDAEWGERL	RLEMEREPSR
661	KADEESQNAW	SPOALADNET	FTRFCQTVDG	VLEQGDSLDT	ELKMPENEER	REGEDHHHKK
721	GTDFMESDEE	EEMDEIDPDL	RIELYVILEE	LRRGSARLRE	NALQAVGADE	LLKLVAMLLD
781	RNIRTISADN	ORLLVPCDDD	VDVGDVLEKE	ICEERVERAG	FAAVVALNIM	SSHRMHKQVI
841	IEDVIDRCVG	LTRLLLIHLI	YPASDSIYKS.	VNSREKDRAM	PTORRRKAG	VCTRDKFSEY
901	IVERITEAIG	LLAVLVKSES.	HLTTGGTNAG	SVALTPFFVA	NVGSLOIVYM	YSASNIFSRA
961	EDSLRFSMIT	DLLSSLHRAP	QFGCRNIGLQ	AHGPDGSWIS	TTTALFIQLV	QSTIKIPKFK
1021	GHADEDELAK	RSEKEEANVR	EGFLEASEVT	NAFLNGFLAK	CSOKGNEMDG	EEDYTFTPGF
1081	LNOELLSAHR	SPEUPAAEMI	LTALGSLLVK	NFRSKSSDMT	IROASLDYLG	NITAKLERDO
1141	REAIAGERRL	DAVVERSFLL	LSDKGVEDYE	SVDISNLKQN	DELEVIETSL	IDYLVITNSS
1201	DIIVYACNEY	VOFPMYEVAE	DLESARSKLK	OTVDTNESEK	DVKKAERKYE	KIQYRGAEM
1261	VFLSKILDKP	FLERRLEKSN	KYKHLDSDAF	WAVEFLAGSR	EFTOSFDTYL	KHIVFGAGSI
1321	TIVALESKAL	KCLSSTAVFD	SSVLILEDVO	OAVHTRHVDS	HAGVRESAVE	LIGRFYLYDI
1301	VLOGGYSOIA	ERILDTGVAV	DLPVIRIMRE	ICERFPTFEM	IPPHLAPHTP	RVTDEEGVEI
1441	LVFETFTLLM	FOPVDTRIYT	NAVOTETITH	CSVAOHCIKD	ANSDYLEGLH	LLLNPOFTF
1501	SGMSVANROI	IDSLVDHILN.	LEONKSSONF	LEVELMERKE	OFFEVMATLS	TLAVESKGAL
1561	LLLTSHVEVL	LPYLTFSGAK	TNAENOVTKC	MIGHLERVIP	LVPFGDFYVL	DSIDENLCK
1621	INSLDMALVV	NPVSCVASIY	KEFERGATET	IDVFSTYLKH	LEVFERNFDS	NPRYPLOYGI
1681	FPILSRSIFT	LOVLSRYCOF	EEFVKEDPTE	EKVEASPNFA	LLHGGHNSRY	HEGGLEOKA
1741	TANGHECAOH	STYLTEROLT	NTYFTPGNAA	NSPOODORL	LVLONLEMFL	OCCEPORLAA
1801	HDEWDENKEA	ONLKEMELSO	SCLOSSVICY	WKAVLESYYV	DADIOLPPAA	OVVULTINO
1861	LGENPGASIP	TLIAMTTEPY	DVIRNRIDIL	REIIDSKYSG	NVOSKANOGY	RLSYKLHLK
1921	MLTOOFKEVE	OFRCDDFHLN	TLPNALPETH	DOMAYTLEGL	YOSLETNEOO	PRFFLOSMV
1981	LFSEFFSHDK	POLNEYIFIA	DNLAMFPYON	DENGNLONMO	IDONIAOTGO	SLLVOYELOI
2041	RMOESEDEDT	VELDENMMSR	LSOLNYGTYC	VLFLDS0VPS	SLLLYVRTEN	
2101	VAEYOPSEAA	EVYFEAVENT	OTHMFKPTTA	LEALNEPPEM	GSFONTGTMF	GCCPOCREM
2161				ONNOFMOROM		and a derenant

figure (31b) generated sequences for the simulation program

and the computer gave these results.

Ls =1853

La = Lb = 2202

SimlarityPerc = 84%

As the concluded percentage of two sequence similarity is 84%.so this is an evidence that both have a similar physiochemical properties of about84% and by this there will be a shared evolutional relationship between them which leads to liability of the generated sequence to be replaced by the real one success percentage of 84% .

But when considering an amino acid H as a polar amino acid the computer gave these results:

Ls = 1900

La = Lb = 2202

SimlarityPerc = 86%

The similarity percentage when considering an amino acid Has a polar amino acid increase to be 86%.

7. CONCLUSION

The next schedule summarizes the previous experiments results the second column of the schedule clears names of protein that have been simulated by different sized amino acids N (amino acid).and the 4th and 5th column gives the matching /similarity percentage between the factions and generated sequences and the last column gives the similarity percentage when the amino acids H considered as one of the polar amino acids.

Table 2: Summary of amino acids simulation experiments.

simil arity perc enta ge of pola r ami no acid s (%)	simi larit y perc enta ge (%)	matching percentag e(%)	Protein size(N)	the name of the simulation Protein	experime nt number
84	81	77	1075	brain tissue disease (1) protein	1
81	78	71	818	brain tissue disease (2)protein	2
81	79	73	421	brain tissue disease (3)protein	3
78	77	72	1006	brain tissue disease (4)protein	4
82	80	71	890	sudden spasmodic (1) epilepsy	5
85	82	78	774	sudden spasmodic)2(epilepsy	6
80	77	70	888	sudden spasmodic (3) epilepsy	7
85	82	77	337	(1) Prion protein	8
86	84	79	1196	(2) Prion protein	9
88	86	77	405	(3) Prion protein	10
88	85	80	254	(4) Prion protein	11
91	91	83	151	(5) Prion protein	12
93	81	83	89	(6) Prion protein	13
86	84	80	2202	(7) Prion protein	14

From the previous simulation programs a lot of conclusion obtained which are:

- 1. The simulation program which built for this study is of good efficacy .because it can generates an amino acids sequences simulates to an acceptable degree to the real (factious) sequence to those amino acids .
- 2. The matching percentage between a factious sequences cases and generated sequences cases from simulation program became between 71% and 83% of cases .
- 3. The similarity percentage between a factious sequences cases and generated sequences cases from simulation program became between 78% and 93%.
- 4. The matching and simulation good sharing between the factious and generated sequences excludes the random matching and this indicates that the sequences descended from one developmental origin.
- 5. The detection of the location and corps of secondary

protein composition is possible if the proteins 2ndory composition that match's its initial one is known which means that match's its initial one is known . which means that this proteins matching in initial composition leads to make this matching in 2ndory composition and function but if the proteins are matched in composition and function is not necessary leads to matching in initial composition and through the presence of such matching in between the initial compositions so it is possible to detect the 2ndory protein composition that matched with in the initial one is known

6. When amino acids H considered from the polar amino acids so the percentage between the two sequences increase . and this leads to similarity increasing between the physiochemical characters of both sequences which may leads to increase the shared developmental relationship between the sequences and by this the two sequences became more liable to exchange between each other without a great influences

8. **REFERENCES**

- [1] ALkhafaji .Z. and ALShaykhli ,A.H. 2012 . Bioinformatics University of Bagdad ,(in Arabic).
- [2] Mathews ,c.k., vanHolde ,k.e and hern,k. g 2000.Biochemistry Third Editition ,An Imprint of Addison Wesley long man ,inc
- [3] Zamzwm ,F. M. 2016.Modeling and Simulation Biological Sequences with the application, Unpublished Ph.D. thesis, , Faculty of Computer Sciences & Mathematics , University of Mosul.,2016. (in Arabic).
- [4] AL-Khayat, B.Y. probability and Random Variables and their Applications using MATLAB. (under preparation),(in Arabic).
- [5] Chou, P. Y. and G.D. Fasman. 1978. Empirical Predictions of Protein Conformation., *Annual Reviews of Biochemistry* 47:251-276. 1978
- [6] Deustscher, M.P. 1990. Guide to Protein Purification, Method in Enzymology, 182:751-776.
- [7] Delamarche, C. 2000 .Color and graphic display (CGD): Programs for multiple sequence alignment analysis in spreadsheet software., BioTechniques 29:100-107,
- [8] Lesk, A.M. 2004. Introduction to Protein Science Architecture: Function and Genomics, Oxford University Press, Univer. of Cambridge,
- [9] Lee, C., Grasso ,C. and Sharlow , M. 2002. "Multiple sequence alignment using partial order graphs", Bioinformatics . 18:452-464,
- [10] Ussery ,D., Jensen , M., Poulsen, T. and Hallin, P. 2004.: Genome update: alignment of bacterial chromosomes, Microbiology 150:2491-2,
- [11] Whitford, D. , John Wiley & Sons Ltd 2005. Proteins Structure and Function., England,
- [12] http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/inform ation3.html.