ANIMPLEMENTATIONFORPERFORMINGACOMPUTERBASEDMUTATIONANALYSIS

BrwaABUBAKER[1],HalgurdMOHAMMED[2],RıdvanSARAÇOĞLU[1,*]

[1]YüzüncüYılUniversity,EngineeringandArchitectureFaculty,Electric-ElectronicsEngineeringDepartment,VanTURKEY

[2]YüzüncüYılUniversity,FacultyofMedicine,MedicalBiologyDepartment,Van TURKEY

brwa.pshdary@gmail.com,mhalgurd@ymail.com,*ridvansaracoglu@yyu.edu.tr

Abstract

ThehistoryofMutationAnalysiscanbesketchedbackfrom1971byRichardLipton.Itisvitaltoidentifythevariationsoccurredin DNAduetomutation.TheaimofthisworkistodevelopanewsoftwarethathelpstopredictthemutatedsequencepositionfoundbetweentheanytwosequenceswhetheritmaybeDNAorProteinoritmaybeboth.Moreoverthisapproachismosteffectiveandaccuratetoanalyzesequences.Thesoftwareisdevelopedthathelpstoprovidenecessaryinputandgetdesiredoutput.Theoutputfilewillshowthepositionwerethemutationoccurforprotein1mutationoccurin1forKandWandforCmutationoccurin40position.Thus,thesystemrunstoprogressqualityoftestingandprovideadvanceefficiencybymeansofvariousmutationoperators.Computerizedmutationanalysisisperformedwithoutmanualintervention.

Keywords: MutationAnalysis,Computerizedmutationanalysis,DNAorProtein

BİLGİSAYARTABANLIMUTASYONANALİZİİÇİNBİRUYGULAMA

Özet

Mutasyonanalizitarihi1971yılındaRichardLiptontarafındanyapılançalışmalaradayanmaktadır.MutasyonnedeniyleDNAiçerisindekioluşanvaryasyonlarınbelirlenmesikritikönemtaşımaktadır.Buçalışmanınözü;DNA,Proteinveyaherikisideolabilenherhangiikisıraarasındabulu

nanmutasyongeçirmişpozisyonlarıntahminedilmesineyardımcıolacakyenibiryazılımgeliştir mektir.Üsteliksıraanaliziiçinçokverimlivedoğrusonuçüretenbiryaklaşımdır.Buyazılım,gerekl igirişlerinkolaycasağlamasıvearzuedilençıkışlarınalınmasınayardımcıolacakşekildegeliştiril miştir.Çıktıdosyası,40pozisyoniçindeki1pozisyondakioluşanK,WveCmutasyonununyerinig österecektir.Böylecebusistemle,kalitelibirtestsürecigerçekleştirilmekteveçeşitlimutasyonope ratörlerivasıtasıylaverimlilikteilerlemesağlanmaktadır.Bilgisayartabanlımutasyonanalizi,ma nüelmüdahaleolmaksızıngerçekleştirilmişolmaktadır.

AnahtarKelimeler: MutasyonAnalizi,Bilgisayarlımutasyonanalizi,DNAveyaProtein

1. Introduction

ThehistoryofMutationTestingcanbesketchedbackfrom1971byRichardLipton[1].Thebir thofthefieldcanalsobeidentifiedinotherpaperspublishedinthelate1970sbyDeMilloetal.[2]and Hamlet[3].ItisvitaltoidentifythevariationsoccurredinDNAduetomutation.Forthatgeneticcod ewhichisusedplaysacrucialrole.DNAisamajorcontrollerofON/OFFmechanismofgenes.Som epartsofDNAarenothavinganyfunctionalpropertiesandsomehavethepropertiesoftranslationt oprotein.Whenthereisanerrorlikeabasedeletedoraddedorawrongbaseincorporatedintheseque nceofDNA,itiscalledamutation.

Existingnucleicacidmoleculesinalivingorganismactsasageneticmplatetotransferthege neticinfofromonegenerationtothenext.Nucleicacidmoleculesareorganizedasgeneswhichcod eforaparticularphenotypeviaspecificproteinsandthegeneexpressionisregulatedbybothextern alandinternalfactorswhichaidthedevelopmentalprocessofanorganism.Thisrelationbetweeng enesandproteinsformsthe“centraldogmaoflife”.

Theproteinishavingcompletesetofaminoacidsandeveryproteinhasuniqueaminoacidarr angedinaspecificsequence.Theinformationtosynthesizeproteinswithuniqueaminoacidsequen ceisprovidedbythenucleicacidpresentwithinthenucleus.Inapre-setsequence,DNApresentinth enucleusgiverisetothespecificRNAsequenceandthatinturnguidethecellularmachinerytosynth

esizeprotein.

Thegeneticcodeisconventionalinformationthattranslatestheinformationencodedingenetic materialintoproteinsinlivingcells.TheDNAcodeswithfourlettersA,T,G,andC.Theseproteincodi ngDNAaresaidtobeCodons.Thesecodonsareagroupofthreeadjacentnucleotidesspecifythesignalst oprotein.Thestopcodonimpliesthecompletionoftheafreshfabricatedprotein.

ManyComputationalprogramdesignlanguagesasawhiteboxunittestmethod.Forexample,FORTRANprograms[4-6],Adaprograms[7],[8],Cprograms[9-11],Javaprograms[12-14],C#programs[15-19],SQLcode[20,21]andAspectprograms[22,23].C#isamodest,object-orientedprogramminglanguageestablishedbyMicrosoftandpermittedbyEuropeanComputerManufacturersAssociationandInternationalStandardsOrganization.ItisbasedonCandC++programminglanguage[16].

ItwasdevelopedbyAndersHejlsbergandhisteamusing.NetFramework.C#isintendedforCommonLanguageInfrastructure(CLI),consistsoftheexecutablecodeandruntimesituationthatpermitsvarioushigh-levellanguagesondifferentcomputerplatformsandarchitectures.

ThereasonsbehindC#awidelyusedprofessionallanguageismodernwithwell-structuredlanguage,objectaswellascomponentoriented,produceefficientprograms,andcompilevarietyofplatforms.

The.Netframeworkapplicationsaremulti-platformapplications.Thesehasbeenapplicableforc#,C++,VisualBasic,Jscript,COBOL,etc.,foraccesstheframeworkaswellasconversewitheachother[18].The.Netframeworkcontainsenormouslibrarycodesusedbytheclientlanguagessuchasc#.Somecomponentsof.NetframeworkareCommonLanguageRuntime,ASP.NetandASP.NetAJAX,etc.

C#sourcecodefilescanbemadeusingabasictexteditor,likeNotepad,andcompilethecodeintoassembliesusingthecommand-linecompiler,whichisagainapartofthe.NETFramework.Monoisanopen-sourceversionofthe.NETFrameworkwhichincludesaC#compilerandrunsonsevaloperatingsystems,includingvariousflavorsofLinuxandMacOS.

Thepurposeofthisworkistodevelopanewsoftwarethathelpstopredictthemutatedsequencepositionfoundbetweentheanytwosequencesof DNAandthosesequenceswillprocessedfortranslationtoProteinsequences.Itispossibletotrackmutationinproteinsequencesaswell.Moreoveritmosteffectiveandaccuratetoanalysessequences.Thesoftwareisdevelopedbased on C#Programlanguagethathelpstoprovidenecessaryinputandgetdesiredoutput.

2. MaterialsandMethods

2.1.DNAMatching

DNAsequenceisfabricatedwithfourbases(A,C,T,andG),anwell-organizedfixed-lengthencodingsystem[24]canbeused.Inmolecularbiology,DNAsequencescarryvitalinformationforeachspeciesandacomparisonbetweenDNAsequencesisaninterestingandmorecomplicated.Therearenumerouscomparisontoolstoprovideapproximatematching.OurDNAmatchingalgorithmarefastmatchingalgorithmtomatchlengthysequencesinfastestapproach.

2.2.ImplementationofMutationAnalysisProgram

FASTAformat:AsequencebookinaFASTAformatincluding(firstline)asingle-linedescription(sequencename),followedbyline(s)or(secondline)ofsequencedata.Thefirstcharacteroftheedenotelineisagreater-than(">")symbol.likethat

>HSBGPGHumangeneforboneglaprotein(BGP)

GGCAGATTCCCCCTAGACCCGCCCGCACCATGGTCAGGCATGCCCCTCCTCATCGCTGGGCACAGCCCAGAGGGT

FASTAcanbeutilizedtodeducefunctionalandevolutionarylinkagesamidstsequencesalsohelpidentifymembersofgenefamilies[25].

“Protein”

- ✓ ProteintoproteinFASTA.
- ✓ ProteintoproteinSmith–Waterman(ssearch).
- ✓ Globalproteintoprotein(Needleman–Wunsch)(ggsearch)

- ✓ Global/localproteintoprotein(glsearch)
- ✓ Proteintoproteinwithunorderedpeptides(fasts)
- ✓ Proteintoproteinwithmixedpeptidesequences(fastf)

"Nucleotide"

- ✓ Nucleotidetonucleotide(DNA/RNAfasta)
- ✓ Orderednucleotidesvsnucleotide(fastm)
- ✓ Unorderednucleotidesvsnucleotide(fasts)

InFASTAalgorithmNucleotideorproteinsequenceistakenasinput.

Thehurryandsensitivityiscontrolledbytheparametercalledktup,whichspecifiesthegaugeoftheword.Thisprogramusesthewordhitstoidentifypotentialmatchesbetweenthequerysequenceanddatabasesequence (Fig. 2.1).Initiallyitreviewforsegment'scontainingseveralthereabouthits.

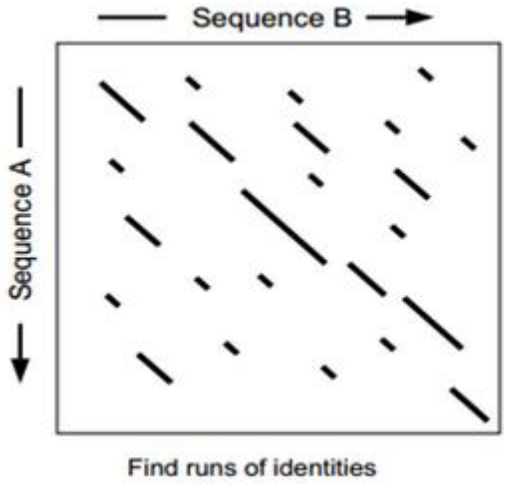
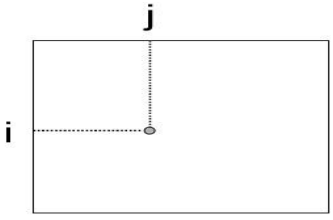


Fig. 2.1.FASTAalgorithm(FASTAAlignments)

FASTAalgorithmhasDotmatrixcomparisonsWordsmatchesin2sequencesI&Jcanberepresentedasadotmatrix(as shown Fig.2.2),thus



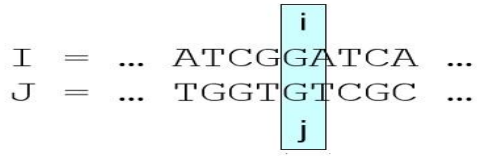


Fig. 2.2Dotmatrixcomparisons

Theflowchartofprogram'salgorithmisshowninFigure2.3inthattheinputersequencesofDNAareintheformofFASTAformat.OncetheDNAisinFASTAformatthenthecomparisonbetweenthetwosequenceshastobedonebasedoncolordifferences.Followedbytranscriptionandtranslationto RNAandProtein.Thencomparisonbetweenthesetwomutatedproteinsequenceshastobeanalysed.Theresulthastobeshownindatagridview.

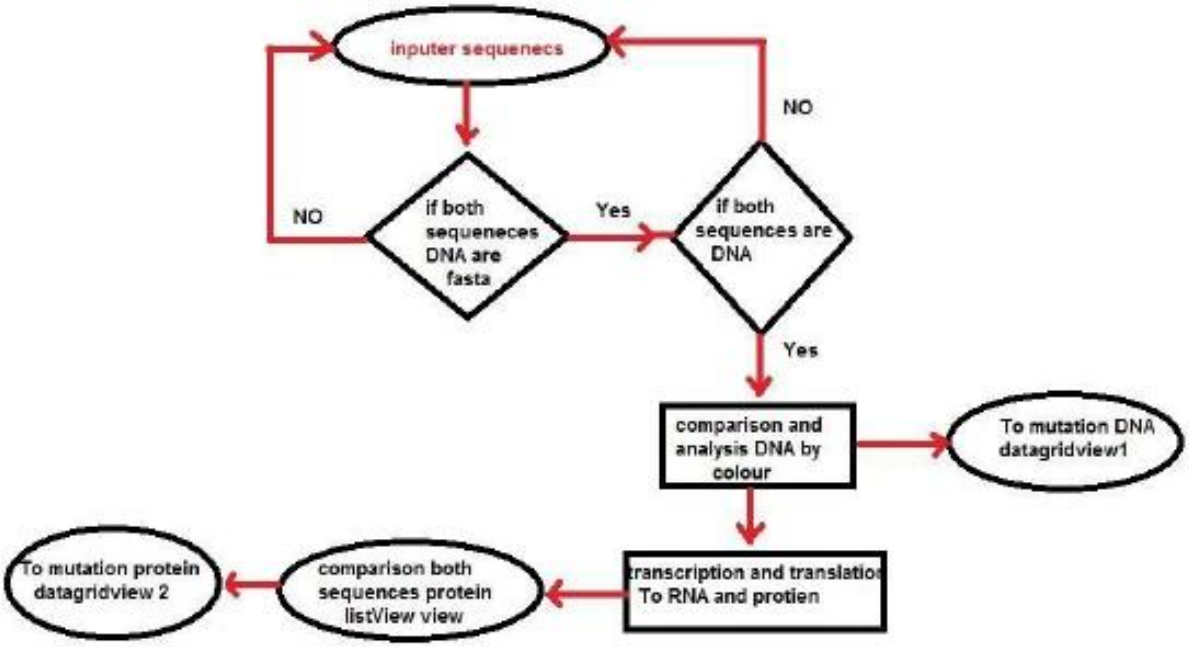


Fig.2.3.Overviewofprogram

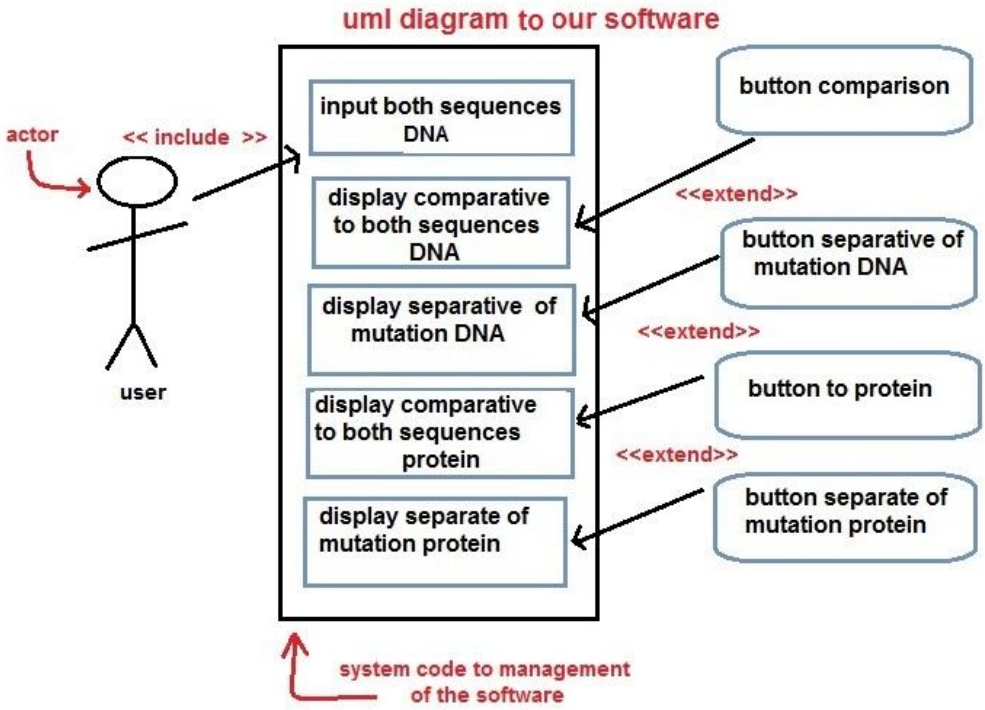


Fig.2.4.UMLdaigramofoursoftware

UML daigram of our softwareis shown in Fig. 2.4.

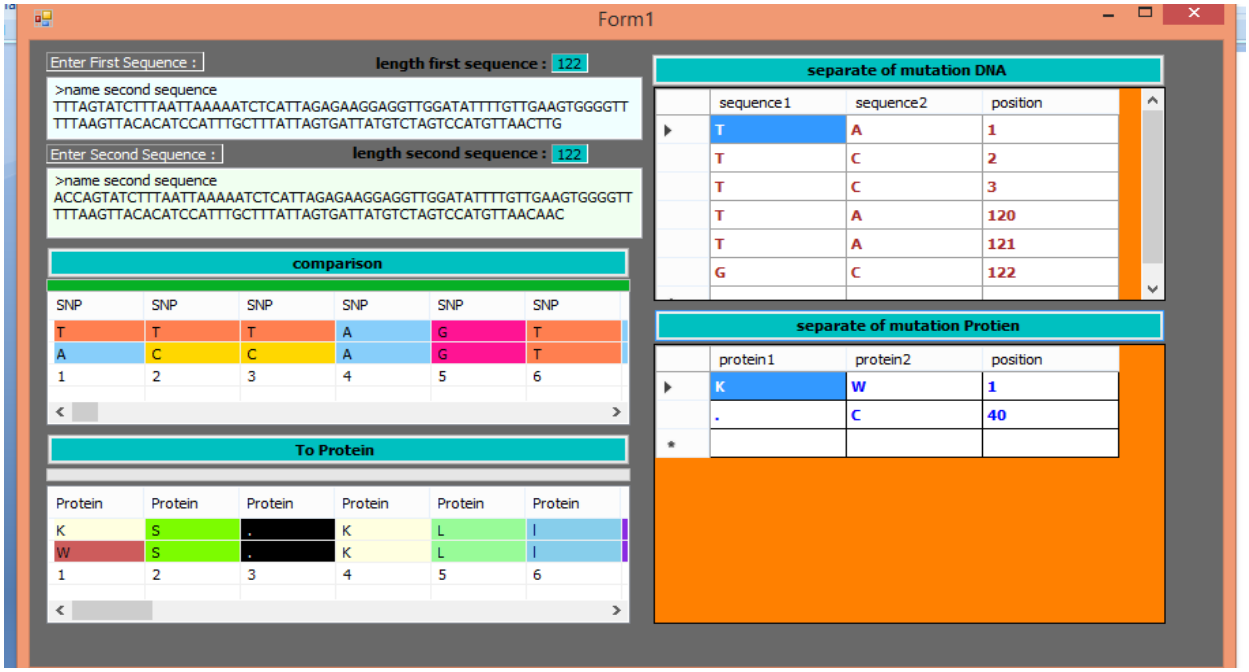
2.3.Retrievesequencesfromdatabase

Thesequencewhichisgoingtobeanalyzedhastoberetrievedfromthespecificproteinsdata baseforanalysis.Importantpointissequencesaremustbeintheformof FASTAformat.ThoseFAS TAsequencesareimportedtooursoftwarebyusingasuitablecods.

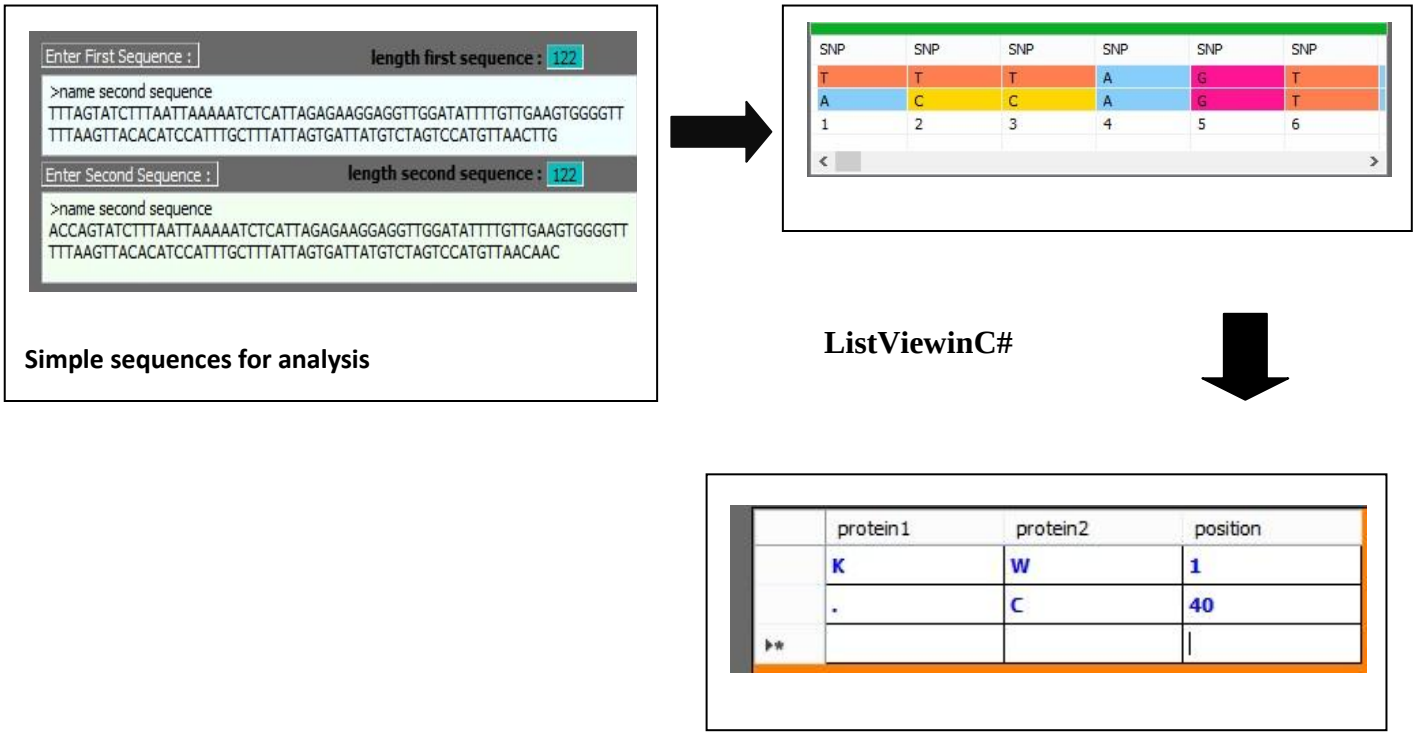
3. ExperimentalResults

Thecompleteviewofoursoftwareinthatthesequenceswhicharegoingtobesequencedareretrieve dandpastetothefollowingboxandselectRUN.Thencomparisonwillstartprocessingoncetheproc essiscompletetheresultwillshowinrightsideofthedialoguebox(asshowninFig.3.1).

Fig. 3.1.RepresenttheWholeSoftware



outputfile



DataGridViewOutputfile

Fig. 3.2.Outputfileshowsseparatemutationofprotein.

Weselecttwosequenceswhicharegoingforanalysisisretrievedasa FASTA Formatandthes equencehastobeundergoneformutationanalysis.Beforethatnucleotidesequencesvariationdone bymeansoflistviewcommand.Thethymineresiduesareinorangecolor,adenineresiduesareinblu ecolor,guanineisinRoseandcytosineisinyellow.Theoutputfileprovidethepositionwerethemuta tionoccurforprotein1mutationoccurin1forKandWandforCmutationoccurin40position(Fig.3.2).

Comparebetweenoursoftwarewithanothertool(bynameTranscriptionandTranslationToo l)isshowninTable3.1.

BlastandFastaaretwoalgorithmsthesareutilizedtocomparesequencesofaminoacids,DN A,proteinsandnucleotidesofdiversespeciesandlookforthesimilarities.thosegeneticalgorithms werewrittenkeepingspeedinmindinordertoasthedatabankofthesequencesswelledonceDNAwa

sisolatedinthelabbythescientistsin1980sthereincreasedaneedtocompareandfindcorrespondinggenesformoreresearchathighspeed.

Table3.1ComparisonofSoftware

Ourtool	**TranscriptionandTranslationTool**
Withoutinternetiswork	Itisneedinternettowork
ItisutilizeFASTAformat	ItisutilizePlainsequenceformat
ItcouldusecolortoDNAsequences	ItcouldnotusecolortoDNAsequences
IthasaccountlengthofsequencesDNA&protein	Ithasn'taccountlengthofsequences
Itcanloadingtwosequences	Itcanloadingonlyonesequence
ItcanseparatemutationDNAsequences	ItcannotseparatemutationDNAsequences
ItcannotdisplayRNA,immediatelyDNAtoprotein	ItwillshowRNAbeforeprotein
Itcouldusecolortoproteinsequences	Itcouldnotusecolortoproteinsequences
ItwillshowpositiontosequencesDNA&protein	Itcannot

FASTAwasthemostvastlyutilizedproteinandDNAsequencedatabasesearchprogramnextthecomingofBLAST.ItisidenticalwithBLASTinmanyroutes,andisstillrepeatedlyutilized.SuchasBLAST,itisaheuristicforapproximatingtheSmith-Watermanalgorithm,bututilizesdiverseheuristicmethodstoraisespeed.BLASTandFASTAaswellutilizeslightlydifferentmethodstocalculatestatisticalsignificance.OursoftwarehasutilizedFASTAthereforeallsoftwareonFASTAformatcouldnotseparatepartofmutationforsegmentofDNAandsegmentofprotein,onthatoursoftwarewasadditionalpartsofmutationforproteinsandnucleotidesbybestqualitycolour.

4. Conclusion

Thepurposeoftheworkistoperformamutationanalysisofeach DNAsequencsfollowedbycomparisontotrackthepositionaswell,thestructureofthesequancesofDNAis4typesofbasesthatsymbolizebyfourletterA,C,GandT.thissoftwarecoluredallthebasesofDNAsequencesbydifferentcoloureachcolourindicatestospecialnucleotideasdeeppinkcolourtoG,goldtoC,lightskyblueto

AandthecoraltoTthatpropertyofthissoftwaregivetheuserdetailsaboutthecontainofeachtypeofnucleotideafterthattranslatetheDNAtoproteinandcomparethemalsobymeansofthissoftware.

Thiswillbemoreaccurate,alsosequenceofproteinissymbolizebyfourletterA,C,GandUandeachthreesymbolizestooneaminoaciddependontheaminoacidcoden.alsointhisbioinformaticstoolgiveeachsymbolspecialcolourtoindicatethatfourdifferentcharacterslesstime,easytopredictthoseregionswhicharemutated.Thus,thesystemrunstoprogressqualityoftestingandprovideadvanceefficiencybymeansofvariousmutationoperators.Computerizedmutationtestingisperformedwithoutmanualinterventin.

Inthebiologicalscienceanychangeinthestructureany DNAsequenceallowtochangeinproteinsequenceandthatmaybeappearabonormaltyinhumanbodythatcalledmutation.

Inthisworkreslutofthissoftware,itissimpletounderstandingfromtheuser.ifcomparethissoftwarefromspeedandefficiencysides,ithashighefficiencyandmuchspeed. And on the otherhandthissoftwareisworkofflineandeasytodownloadonthewindowssystem.

Acknowledgments

Theauthorsaregratefulforthesupportprovided byYüzüncüYılUniversity.

References

[1]MathurP.“MutationTesting”,inEncyclopediaofSoftwareEngineering,J.J.Marciniak,Ed.,1994,pp.707–713.

[2]DeMilloRA,LiptonRJ,SaywardFG.“HintsonTestDataSelection:HelpforthePracticingProgrammer,”Computer,vol.11,no.4,pp.34–41,April1978.

[3]HamletRG,“TestingProgramswiththeAidofaCompiler,”IEEETransactionsonSoftwareEngineering,July1977,3(4):279–290,

[4].Acree,A.T.,Budd,T.A.,DeMillo,R.A.,Lipton,R.J.,andSayward,F.G.,“MutationAnalysis,” GeorgiaInstituteofTechnology,Atlanta,Georgia,TechniqueReportGIT-ICS-79/08,1979.

[5].BuddTA,DeMilloRA,LiptonRJ,SaywardFG.“TheDesignofaPrototypeMutationSystemforProgramTesting,”inProceedingsoftheAFIPSNationalComputerConference,vol.74.Anaheim,NewJersey:ACM,5-8June1978,pp.623–627.

[6]BuddTA,SaywardFG.“UsersGuidetothePilotMutationSystem,”YaleUniversity,NewHaven,Connecticut,TechniqueReport114,1977.

[7].BowserJH.“ReferenceManualforAdaMutantOperators,”GeorgiaInstituteofTechnology,Atlanta,Georgia,TechniqueReportGITSERC-88/02,1988.

[8].Offutt,A.J.,Voas,J.,andPayn,J.,“MutationOperatorsforAda,”GeorgeMasonUniversity,Fairfax,Virginia,TechniqueReportISSE-TR-96-09,1996.

[9].AgrawalH,DeMilloRA,HathawayB,HsuW,HsuW,KrauserEW,MartinRJ,MathurAP,SpaffordE.“DesignofMutantOperatorsfortheCProgrammingLanguage,”PurdueUniversity,WestLafayette,Indiana,TechniqueReportSERC-TR-41-P,March1989.

[10]DelamaroME,MaldonadoJC,MathurAP.“InterfaceMutation:AnApproachforIntegration Testing,”IEEETransactionsonSoftwareEngineering,May2001,27(3):228–247.

[11]VilelaP,MachadoM,WongWE.“TestingforSecurityVulnerabilitiesinSoftware,”inSoftwareEngineeringandApplications,2002.

[12]ChevalleyP.“ApplyingMutationAnalysisforObject-orientedProgramsUsingaReflective Approach,”inProceedingsofthe8thAsia-PacificSoftwareEngineeringConference(APSEC01), Macau,China,4-7December2001,p.267.

[13]ChevalleyP,Th´evenod-FosseP.“AMutationAnalysisToolforJavaPrograms,”InternationalJournalonSoftwareToolsforTechnologyTransfer,November2002,5(1):90–103.

[14]Ma,Y.S.,Offutt,A.J.andKwon,Y.R.,“MuJava:AnAutomatedClassMutationSystem,”SoftwareTesting,Verification&Reliability,vol.15,no.2,pp.97–133,June2005.

[15]Derezi´nskaA.“Object-orientedMutationtoAssesstheQualityofTests,”inProceedingsofthe29thEuromicroConference,Belek,Turkey,1-6September2003,pp.417–420.

[16]Derezi´nskaA.“AdvancedMutationOperatorsApplicableinC#Programs,”WarsawUniver

sityofTechnology,Warszawa,Poland,TechniqueReport,2005.

[17]Derezi´nskaA.“QualityAssessmentofMutationOperatorsDedicatedforC#Programs,”inProceedingsofthe6thInternationalConferenceonQualitySoftware(QSIC’06),Beijing,China,27-28October2006.

[18]Derezi´nskaA,SzustekA.“CREAM-ASystemforObject-OrientedMutationofC#Programs,”WarsawUniversityofTechnology,Warszawa,Poland,TechniqueReport,2007.

[19]Derezi´nskaA,SzustekA.“Tool-SupportedAdvancedMutationApproachforVerificationofC#Programs,”inProceedingsofthe3thInternationalConferenceonDependabilityofComputerSystems(DepCoS-RELCOMEX’08),SzklarskaPorˆeba,Poland,26-28June2008,pp.261–268.

[20]ShahriarH,ZulkernineM.“MUSIC:Mutation-basedSQLInjectionVulnerabilityChecking,”inProceedingsofthe8thInternationalConferenceonQualitySoftware(QSIC’08),Oxford,UK,12-13August2008,pp.77–86.

[21]TuyaJ,CabalMJS,delaRivaC.“SQLMutation:AToooltoGenerateMutantsofSQLDatabaseQueries,”inProceedingsofthe2ndWorkshoponMutationAnalysis(MUTATION’06).Raleigh,NorthCarolina:IEEEComputerSociety,November2006,p.1.

[22]AnbalaganP,XieT.“AutomatedGenerationofPointcutMutantsforTestingPointcutsinAspectJPrograms,”inProceedingsofthe19thInternationalSymposiumonSoftwareReliabilityEngineering(ISSRE’08).Redmond,Washingto:IEEEComputerSociety,11-14November2008,pp.239–248.

[23]FerrariFC,MaldonadoJC,RashidA.“MutationTestingforAspect-OrientedPrograms,”inProceedingsofthe1stInternationalConferenceonSoftwareTesting,Verification,andValidation(ICST’08).Lillehammer,Norway:IEEEComputerSociety,9-11April2008,pp.52–61.

[24]KimJW,KimE,ParkK.FastmatchingmethodforDNAsequences.InCombinatorics,Algorithms,ProbabilisticandExperimentalMethodologies,volume4614ofLNCS,pages271–281,2007.

[25]

Setubal&Meidanis.IntroductiontoComputationalMolecularBiology,PWSPublishingCompany,1997.Chapter3.