Megan Chu

A12814536

Assignment 1

4/10/18

Q1.

I, Megan Chu, have read the Academic Integrity Policy, Grading Policy,

and Syllabus for BENG/BIMM/CSE 182 and agree to them.

Q2. **Program Output:**

Hello Bioinformatics

Q3. **Program Output:**

>gi|6978799|ref|NP\_036683.1| early growth response 1; nerve growth factor-induced gene [Rattus norvegicus]

508

>gi|45768856|gb|AAH67618.1| Serum/glucocorticoid regulated kinase [Danio rerio]

433

>gi|45768786|gb|AAH68134.1| Unknown (protein for MGC:95907) [Mus musculus]

423

>gi|27923854|sp|P59241|STK6\_RAT Serine/threonine kinase 6 (Aurora-A) (ratAurA)

397

>gi|45768720|gb|AAH67812.1| Cyclin L1 [Homo sapiens]

526

>gi|45768758|gb|AAH68160.1| Cdk7 protein [Mus musculus]

346

>gi|45219906|gb|AAH66834.1| Mastl protein [Mus musculus]

671

>gi|18202599|sp|Q63796|M3KC\_RAT Mitogen-activated protein kinase kinase kinase 12 (MAPK-upstream kinase) (MUK)

888

>gi|4835224|emb|CAB42902.1| protein kinase ATN1 like protein [Arabidopsis thaliana]

370

>gi|40787731|gb|AAH64804.1| SLK protein [Homo sapiens]

617

>gi|18202068|sp|O55173|PDPK\_RAT 3-phosphoinositide dependent protein kinase-1 (Protein kinase B kinase) (PkB kinase)

559

>gi|34191428|gb|AAH36504.2| C9orf96 protein [Homo sapiens]

700

>gi|29747774|gb|AAH50806.1| Gene model 711, (NCBI) [Mus musculus]

587

>gi|28856169|gb|AAH48033.1| Serine/threonine kinase 3 (STE20 homolog, yeast) [Danio rerio]

492

>gi|20071571|gb|AAH26466.1| Unknown (protein for IMAGE:4485517) [Mus musculus]

202

>gi|45709347|gb|AAH67695.1| Unknown (protein for MGC:85918) [Danio rerio]

320

Q4. **Program Output:**

Creating filteredDatabase.txt with mouse and rat sequences...

Contents of filteredDatabase:

>gi|6978799|ref|NP\_036683.1| early growth response 1; nerve growth factor-induced gene [Rattus norvegicus]

MDNYPKLEEMMLLSNGAPQFLGAAGTPEGSGGNNSSSSSSSSSGGGGGGGSNSGSSAFNP

QGEPSEQPYEHLTTESFSDIALNNEKALVETSYPSQTTRLPPITYTGRFSLEPAPNSGNT

LWPEPLFSLVSGLVSMTNPPTSSSSAPSPAASSSSSASQSPPLSCAVPSNDSSPIYSAAP

TFPTPNTDIFPEPQSQAFPGSAGTALQYPPPAYPATKGGFQVPMIPDYLFPQQQGDLSLG

TPDQKPFQGLENRTQQPSLTPLSTIKAFATQSGSQDLKALNNTYQSQLIKPSRMRKYPNR

PSKTPPHERPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQCRICMRNFSRSDHLTTHI

RTHTGEKPFACDICGRKFARSDERKRHTKIHLRQKDKKADKSVVASSAASSLSSYPSPVA

TSYPSPATTSFPSPVPTSYSSPGSSTYPSPAHSGFPSPSVATTYASVPPAFPAQVSTFQS

AGVSNSFSTSTGLSDMTATFSPRTIEIC

>gi|45768786|gb|AAH68134.1| Unknown (protein for MGC:95907) [Mus musculus]

MSTRNCQGTDSVIKHLDTIPEDKKVRVQRTQSTFDPFEKPANQVKRVHSENNACINFKSS

SAGKESPKVRRHSSPSSPTSPKFGKADSYEKLEKLGEGSYATVYKGKSKVNGKLVALKVI

RLQEEEGTPFTAIREASLLKGLKHANIVLLHDIIHTKETLTLVFEYVHTDLCQYMDKHPG

GLHPDNVKLFLFQLLRGLSYIHQRYILHRDLKPQNLLISDTGELKLADFGLARAKSVPSH

TYSNEVVTLWYRPPDVLLGSTEYSTCLDMWGVGCIFVEMIQGVAAFPGMKDIQDQLERIF

LVLGTPNEDTWPGVHSLPHFKPERFTVYSSKSLRQAWNKLSYVNHAEDLASKLLQCSPKN

RLSAQAALSHEYFSDLPPRLWELTDMSSIFTVPNVRLQPEAGESMRAFGKNNSYGKSLSN

SKH

>gi|27923854|sp|P59241|STK6\_RAT Serine/threonine kinase 6 (Aurora-A) (ratAurA)

MDRCKENCVSRPVKSTVPFGPKRVLVTEQIPSQHPGSASSGQAQRVLCPSNSQRVPPQAQ

KPVAGQKPVLKQLPAASGPRPASRLSNPQKSEQPQPAASGNNSEKEQTSIQKTEDSKKRQ

WTLEDFDIGRPLGKGKFGNVYLAREKQSKFILALKVLFKVQLEKAGVEHQLRREVEIQSH

LRHPNILRLYGYFHDATRVYLILEYAPLGTVYRELQKLSKFDEQRTATYITELANALSYC

HSKRVIHRDIKPENLLLGSNGELKIADFGWSVHAPSSRRTTLCGTLDYQPPEMIEGRMHD

EKVDLWSLGVLCYEFLVGMPPFEAHTYQETYRRISRVEFTFPDFVTEGARDLISRLLKHN

SSQRLTLAEVLEHPWIKANSSKPPTGHNSKEATSKSS

>gi|45768758|gb|AAH68160.1| Cdk7 protein [Mus musculus]

MAVDVKSRAKRYEKLDFLGEGQFATVYKARDKNTNQIVAIKKIKLGHRSEAKDGINRTAL

REIKLLQELSHPNIIGLLDAFGHKSNISLVFDFMETDLEVIIKDNSLVLTPSHIKAYMLM

TLQGLEYLHQHWILHRDLKPNNLLLDENGVLKLADFGLAKSFGSPNRAYTHQVVTRWYRA

PELLFGARMYGVGVDMWAVGCILAELLLRVPFLPGDSDLDQLTRIFETLGTPTEEQWPDM

CSLPDYVTFKSFPGVPLQHIFIAAGDDLLELIQGLFLFNPCTRTTASQALKTKYFSNRPG

PTPGCQLPRPNCPVEALKEPANPTVATKRKRAEALEQGILPKKLIF

>gi|45219906|gb|AAH66834.1| Mastl protein [Mus musculus]

SMSKPKQDYSRTPGQVLSLISSLGFFTPVGEKDQDSANMFSAPKSAAQLSRGFICPMSVD

QKEPTSYSSKLLKSCFETLSSNPEIPVKCLTSNLLQCRKRLGTSSTSSQSHTFVSSVESE

CHSNPKWERDCQSTESSGCAMSWNAVEMLYAKSTSAIKTKTELELALSPIHDSSAIPAAG

SNQVTLPRKCFREISWEARDPDNENMTIDKGQSGFCQSSQRSVNSSATSEEHLGKRNYKR

NFHLVDSSPCQEIMQSKKNCTEYEANKERQGCRANQSTGLTTEVQNLKLSGCESQQLDYA

NKENIVTYLTDRQTPEKLHIPTIAKNLMSELDEDRELSSKKDCLSSNSVCSDEDRALKTT

CVDSDSSFPGVSMMESSLEIQALEPDKSIRDYSFEEPNTEDLFVLPKCQENSLPQDDCHA

CIQDSSQVSAHPSKAPKALTSKINVVAFRSFNSHINASTNSEPSKISITSLDAMDISYDY

SGSYPMAVSPTEKGRHYTSHQTPNQVKLGTSYRTPKSVRRGAAPVDDGRILGTPDYLAPE

LLLGTAHGPAVDWWALGVCLFEFLTGIPPFNDETPQQVFQNILKRDIPWPEGEEKLSDNA

QSAMDMLLTIDDSKRAGMRELKQHPLFSEVDWENLQHQTMPFVPQPDDETDTSYFEARNN

AQHLTISGFSL

>gi|18202599|sp|Q63796|M3KC\_RAT Mitogen-activated protein kinase kinase kinase 12 (MAPK-upstream kinase) (MUK)

MACLHETRTPSPSFGGFVSTLSEASMRKLDPDTSDCTPEKDLTPTQCVLRDVVPLGGQGG

GGPSPSPGGEPPPEPFANSVLQLHEQDTGGPGGATGSPESRASRVRADEVRLQCQSGSGF

LEGLFGCLRPVWTMIGKAYSTEHKQQQEDLWEVPFEEILDLQWVGSGAQGAVFLGRFHGE

EVAVKKVRDLKETDIKHLRKLKHPNIITFKGVCTQAPCYCILMEFCAQGQLYEVLRAGRP

VTPSLLVDWSMGIAGGMNYLHLHKIIHRDLKSPNMLITYDDVVKISDFGTSKELSDKSTK

MSFAGTVAWMAPEVIRNEPVSEKVDIWSFGVVLWELLTGEIPYKDVDSSAIIWGVGSNSL

HLPVPSSCPDGFKILLRQCWNRKPRNRPSFRQILLHLDIASADVLSTPQETYFKSQAEWR

EEVKLHFEKIKSEGTCLHRLEEELVMRRREELRHALDIREHYERKLERANNLYMELNALM

LQLELKERELLRREQALERRCPGLLKSHTSRSLLHGNTMEKLIKKRNVPQKLSPHSKRPD

ILKTESLLPKLDAALSGVGLPGCPKAPPSPGRSRRGKTRHRKASAKGSCGDLPGLRAALP

PHEPGGLGSPGGLGVGPTAWDASPPALRGLHHDLLLRKMSSSSPDLLSAALGARGRGATG

GARDPGSPPPPQGDTPPSEGSAPGSTSPDSPGGAKGEPPPPVGPGEGVGLLGTGREGTTG

RGGSRAGYQHLTPAALLYRAAVTRSQKRGISSEEEEGEVDSEVELPPSQRWPQGPNMRQS

LSTFSSENPSDVEEGTASEPSPSGTPEVGSTNTDERPDERSDDMCSQGSEIPLDLPTSEV

VPERETSSLPMQHQDDQGPNPEDSDCDSTELDNSNSIDALPPPASLPP

>gi|18202068|sp|O55173|PDPK\_RAT 3-phosphoinositide dependent protein kinase-1 (Protein kinase B kinase) (PkB kinase)

MARTTSQLYDAVPIQSSVVLCSCPSPSMVRSQTEPSSSPGIPSGVSRQGSTMDGTTAEAR

PSTNPLQQHPAQLPPQPRKKRPEDFKFGKILGEGSFSTVVLARELATSREYAIKILEKRH

IIKENKVPYVTRERDVMSRLDHPFFVKLYFTFQDDEKLYFGLSYAKNGELLKYIRKIGSF

DETCTRFYTAEIVSALEYLHGKGIIHRDLKPENILLNEDMHIQITDFGTAKVLSPDSKQA

RANSFVGTAQYVSPELLTEKSACKSSDLWALGCIIYQLVAGLPPFRAGNEYLIFQKIIKL

EYDFPEKFFPKARDLVEKLLVLDATKRLGCEEMEGYGPLKAHPFFESITWENLHQQTPPK

LTAYLPAMSEDDEDCYGNYDNLLSQFGCMQVSSSSSSHSLCAVDASLPQRSGSNIEQYIH

DLDTNSFELDLQFSEDEKRLLLEKQAGGNPWHQFVENNLILKMGPVDKRKGLFARRRQLL

LTEGPHLYYVDPVNKVLKGEIPWSQELRPEAKNFKTFFVHTPNRTYYLMDPSGNAHKWCR

KIQEVWRQQYQSSPDAAVQ

>gi|29747774|gb|AAH50806.1| Gene model 711, (NCBI) [Mus musculus]

MDYYSQGTFQNIMENKRKLKAVVDTEWMHTMLSQVLDAIEYLHKLNIVHRNLKPSNIVLV

NSGYCKLQDMSSQALMTHEAKWNVRAEEDPCQKSWMAPEALKFSFSTKSDIWSLGCIILD

MATCSFLNDTEAMQLRKAIRHHPGSLKPILKTMEEKQIPGTDVYYLLLPFMLHINPSDRL

AIKDVMQVTFMSNSFKSSSVALNMQRQKVPIFITDVLLEGNMANILDVMQNFSSRPEVQL

RAINKLLTMPEDQLGLPWPTELLEEVISIIKQHGRILDILLSTCSLLLRVLGQALAKDPE

AEIPRSSLIISFLMDTLRSHPNSERLVNVVYNVLAIISSQGQISEELEEEGLFQLAQENL

EHFQEDRDICLSILSLLWSLLVDVVTVDKEPLEQLSGMVTWVLATHPEDVEIAEAGCAVL

WLLSLLGCIKESQFEQVVVLLLRSIQLCPGRVLLVNNAFRGLASLAKVSELVAFRIVVLE

EGSSGLHLIQDIYKLYKDDPEVVENLCMLLAHLTSYKEILPEMESGGIKDLVQVIRGRFT

SSLELISYADEILQVLEANAQPGLQEDQLEPPAGQEAPLQGEPLFRP

>gi|20071571|gb|AAH26466.1| Unknown (protein for IMAGE:4485517) [Mus musculus]

PTRPTRLIVSNFSQAKQKSHLVDPQILRDQSRLAPEIITATQYKKCDEFQTGILIYEMLH

LPNPFDENPELKEKEYTRTDLPRIPLRSPYSWGLQQLASCLLNPNPSERILISDAKGILQ

CLLWGPREDLFQIFTTSATLAQKNALLQNWLDIKRTLLMIKFAEKSLDREGGISLEDWLC

AQYLAFATTDSLSYIVKILQYR

Q5. **Program Output:**

Creating data.seq...

Contents of data.seq:

MDNYPKLEEMMLLSNGAPQFLGAAGTPEGSGGNNSSSSSSSSSGGGGGGGSNSGSSAFNPQGEPSEQPYEHLTTESFSDIALNNEKALVETSYPSQTTRLPPITYTGRFSLEPAPNSGNTLWPEPLFSLVSGLVSMTNPPTSSSSAPSPAASSSSSASQSPPLSCAVPSNDSSPIYSAAPTFPTPNTDIFPEPQSQAFPGSAGTALQYPPPAYPATKGGFQVPMIPDYLFPQQQGDLSLGTPDQKPFQGLENRTQQPSLTPLSTIKAFATQSGSQDLKALNNTYQSQLIKPSRMRKYPNRPSKTPPHERPYACPVESCDRRFSRSDELTRHIRIHTGQKPFQCRICMRNFSRSDHLTTHIRTHTGEKPFACDICGRKFARSDERKRHTKIHLRQKDKKADKSVVASSAASSLSSYPSPVATSYPSPATTSFPSPVPTSYSSPGSSTYPSPAHSGFPSPSVATTYASVPPAFPAQVSTFQSAGVSNSFSTSTGLSDMTATFSPRTIEIC@MTIQTETSVSAPDLTYSKTRGLVANLSAFMKQRKMGLNDFIQKLSANSYACKHPEVQSILNLTPPQDVELMNSNPSPPPSPSQQINLGPSSNPTAKPSDFDFLKVIGKGSFGKVLLARHRSDEKFYAVKVLQKKAILKKKEEKHIMSERNVLLKNVKHPFLVGLHYSFQTTDKLYFVLDYINGGELFYHLQRERCFLEPRARFYAAEIASALGYLHSLNIVYRDLKPENILLDSQGHIILTDFGLCKENIEPNGTTSTFCGTPEYLAPEVLHKQPYDRTVDWWCLGAVLYEMLYGLPPFYSRNTAEMYDNILNKPLQLKPNISNAARHLLEGLLQKDRTKRLGFTDDFTEIKNHMFFSPINWDDLNAKKLTPPFNPNVTGPNDLRHFDPEFTDEPVPNSIGCSPDSALVTSSITEATEAFLGFSYAPAMDSYL@MSTRNCQGTDSVIKHLDTIPEDKKVRVQRTQSTFDPFEKPANQVKRVHSENNACINFKSSSAGKESPKVRRHSSPSSPTSPKFGKADSYEKLEKLGEGSYATVYKGKSKVNGKLVALKVIRLQEEEGTPFTAIREASLLKGLKHANIVLLHDIIHTKETLTLVFEYVHTDLCQYMDKHPGGLHPDNVKLFLFQLLRGLSYIHQRYILHRDLKPQNLLISDTGELKLADFGLARAKSVPSHTYSNEVVTLWYRPPDVLLGSTEYSTCLDMWGVGCIFVEMIQGVAAFPGMKDIQDQLERIFLVLGTPNEDTWPGVHSLPHFKPERFTVYSSKSLRQAWNKLSYVNHAEDLASKLLQCSPKNRLSAQAALSHEYFSDLPPRLWELTDMSSIFTVPNVRLQPEAGESMRAFGKNNSYGKSLSNSKH@MDRCKENCVSRPVKSTVPFGPKRVLVTEQIPSQHPGSASSGQAQRVLCPSNSQRVPPQAQKPVAGQKPVLKQLPAASGPRPASRLSNPQKSEQPQPAASGNNSEKEQTSIQKTEDSKKRQWTLEDFDIGRPLGKGKFGNVYLAREKQSKFILALKVLFKVQLEKAGVEHQLRREVEIQSHLRHPNILRLYGYFHDATRVYLILEYAPLGTVYRELQKLSKFDEQRTATYITELANALSYCHSKRVIHRDIKPENLLLGSNGELKIADFGWSVHAPSSRRTTLCGTLDYQPPEMIEGRMHDEKVDLWSLGVLCYEFLVGMPPFEAHTYQETYRRISRVEFTFPDFVTEGARDLISRLLKHNSSQRLTLAEVLEHPWIKANSSKPPTGHNSKEATSKSS@MASGPHSTATAAAAASSAAPSAGGSSSGTTTTTTTTTGGILIGDRLYSEVSLTIDHSLIPEERLSPTPSMQDGLDLPSETDLRILGCELIQAAGILLRLPQVAMATGQVLFHRFFYSKSFVKHSFEIVAMACINLASKIEEAPRRIRDLINVFHHLRQLRGKRTPSPLILDQNYINTKNQVIKAERRVLKELGFCVHVKHPHKIIVMYLQVLECERNQTLVQTAWNYMNDSLRTNVFVRFQPETIACACIYLAARALQIPLPTRPHWFLLFGTTEEEIQEICIETLRLYTRKKPNYELLEKEVEKRKVALQEAKLKAKGLNPDGTPALSTLGGFSPASKPSSPREVKAEEKSPISINVKTVKKEPEDRQQASKSPYNGVRKDSKRSRNSRSASRSRSRTRSRSRSHTPRRHYNNRRSRSGTYSSRSRSRSRSHSESPRRHHNHGSPHLKAKHTRDDLKSSNRHGHKRKKSRSRSQSKSRDHSDAAKKHRHERGHHRDRRERSRSFERSHKSKHHGGSRSGHGRHRR@MAVDVKSRAKRYEKLDFLGEGQFATVYKARDKNTNQIVAIKKIKLGHRSEAKDGINRTALREIKLLQELSHPNIIGLLDAFGHKSNISLVFDFMETDLEVIIKDNSLVLTPSHIKAYMLMTLQGLEYLHQHWILHRDLKPNNLLLDENGVLKLADFGLAKSFGSPNRAYTHQVVTRWYRAPELLFGARMYGVGVDMWAVGCILAELLLRVPFLPGDSDLDQLTRIFETLGTPTEEQWPDMCSLPDYVTFKSFPGVPLQHIFIAAGDDLLELIQGLFLFNPCTRTTASQALKTKYFSNRPGPTPGCQLPRPNCPVEALKEPANPTVATKRKRAEALEQGILPKKLIF@SMSKPKQDYSRTPGQVLSLISSLGFFTPVGEKDQDSANMFSAPKSAAQLSRGFICPMSVDQKEPTSYSSKLLKSCFETLSSNPEIPVKCLTSNLLQCRKRLGTSSTSSQSHTFVSSVESECHSNPKWERDCQSTESSGCAMSWNAVEMLYAKSTSAIKTKTELELALSPIHDSSAIPAAGSNQVTLPRKCFREISWEARDPDNENMTIDKGQSGFCQSSQRSVNSSATSEEHLGKRNYKRNFHLVDSSPCQEIMQSKKNCTEYEANKERQGCRANQSTGLTTEVQNLKLSGCESQQLDYANKENIVTYLTDRQTPEKLHIPTIAKNLMSELDEDRELSSKKDCLSSNSVCSDEDRALKTTCVDSDSSFPGVSMMESSLEIQALEPDKSIRDYSFEEPNTEDLFVLPKCQENSLPQDDCHACIQDSSQVSAHPSKAPKALTSKINVVAFRSFNSHINASTNSEPSKISITSLDAMDISYDYSGSYPMAVSPTEKGRHYTSHQTPNQVKLGTSYRTPKSVRRGAAPVDDGRILGTPDYLAPELLLGTAHGPAVDWWALGVCLFEFLTGIPPFNDETPQQVFQNILKRDIPWPEGEEKLSDNAQSAMDMLLTIDDSKRAGMRELKQHPLFSEVDWENLQHQTMPFVPQPDDETDTSYFEARNNAQHLTISGFSL@MACLHETRTPSPSFGGFVSTLSEASMRKLDPDTSDCTPEKDLTPTQCVLRDVVPLGGQGGGGPSPSPGGEPPPEPFANSVLQLHEQDTGGPGGATGSPESRASRVRADEVRLQCQSGSGFLEGLFGCLRPVWTMIGKAYSTEHKQQQEDLWEVPFEEILDLQWVGSGAQGAVFLGRFHGEEVAVKKVRDLKETDIKHLRKLKHPNIITFKGVCTQAPCYCILMEFCAQGQLYEVLRAGRPVTPSLLVDWSMGIAGGMNYLHLHKIIHRDLKSPNMLITYDDVVKISDFGTSKELSDKSTKMSFAGTVAWMAPEVIRNEPVSEKVDIWSFGVVLWELLTGEIPYKDVDSSAIIWGVGSNSLHLPVPSSCPDGFKILLRQCWNRKPRNRPSFRQILLHLDIASADVLSTPQETYFKSQAEWREEVKLHFEKIKSEGTCLHRLEEELVMRRREELRHALDIREHYERKLERANNLYMELNALMLQLELKERELLRREQALERRCPGLLKSHTSRSLLHGNTMEKLIKKRNVPQKLSPHSKRPDILKTESLLPKLDAALSGVGLPGCPKAPPSPGRSRRGKTRHRKASAKGSCGDLPGLRAALPPHEPGGLGSPGGLGVGPTAWDASPPALRGLHHDLLLRKMSSSSPDLLSAALGARGRGATGGARDPGSPPPPQGDTPPSEGSAPGSTSPDSPGGAKGEPPPPVGPGEGVGLLGTGREGTTGRGGSRAGYQHLTPAALLYRAAVTRSQKRGISSEEEEGEVDSEVELPPSQRWPQGPNMRQSLSTFSSENPSDVEEGTASEPSPSGTPEVGSTNTDERPDERSDDMCSQGSEIPLDLPTSEVVPERETSSLPMQHQDDQGPNPEDSDCDSTELDNSNSIDALPPPASLPP@MISRMIFRNYPSHNESDDEPFHFSISRELLLDRNDVVVGEMIGEGAYSIVYKGLLRNQFPVAVKIMDPSTTSAVTKAHKKTFQKEVLLLSKMKHDNIVKFVGACIEPQLIIVTELVEGGTLQRFMHSRPGPLDLKMSLSFALDISRAMEFVHSNGIIHRDLNPRNLLVTGDLKHVKLADFGIAREETRGGMTCEAGTSKWMAPEVYSPEPLRVGEKKEYDHKADIYSFAIVLWQLVTNEEPFPDVPNSLFVPYLVSQGRRPILTKTPDVFVPIVESCWAQDPDARPEFKEISVMLTNLLRRMSSDSSIGTTLPDGEAYEGEMEESENSPLLQEHFCKVKKPKEKKKKKKLVKMRFPFFKKFKVWLYNYKP@MSFFNFRKIFKLGSEKKKKQYEHVKRDLNPEDFWEIIGELGDGAFGKVYKAQNKETSVLAAAKVIDTKSEEELEDYMVEIDILASCDHPNIVKLLDAFYYENNLWILIEFCAGGAVDAVMLELERPLTESQIQVVCKQTLDALNYLHDNKIIHRDLKAGNILFTLDGDIKLADFGVSAKNTRTIQRRDSFIGTPYWMAPEVVMCETSKDRPYDYKADVWSLGITLIEMAEIEPPHHELNPMRVLLKIAKSEPPTLAQPSRWSSNFKDFLKKCLEKNVDARWTTSQLLQHPFVTVDSNKPIRELIAEAKAEVTEEVEDGKEEDEEEETENSLPIPASKRASSDLSIASSEEDKLSQNACILESVSEKTERSNSEDKLNSKILNEKPTTDEPEKAVEDINEHITDAQLEAMTELHDRTAVIKENEREKRPKLENLPDTEDQETVDINSVSEGKENNIMITLETNIEHNLKSEEEKDQEKQQMFENKLIKSEEIKDTILQTVDLVSQETGEKEANIQAVDSEVGLTKEDTQEKLGEDDKTQKDVISNTSDVIGTCEAADVAQKVDEDSAEDTQSNDGKEVVEVGQKLINKPMVGPEAGGTKEVPIKEIVEMNEIEEKKKK@MARTTSQLYDAVPIQSSVVLCSCPSPSMVRSQTEPSSSPGIPSGVSRQGSTMDGTTAEARPSTNPLQQHPAQLPPQPRKKRPEDFKFGKILGEGSFSTVVLARELATSREYAIKILEKRHIIKENKVPYVTRERDVMSRLDHPFFVKLYFTFQDDEKLYFGLSYAKNGELLKYIRKIGSFDETCTRFYTAEIVSALEYLHGKGIIHRDLKPENILLNEDMHIQITDFGTAKVLSPDSKQARANSFVGTAQYVSPELLTEKSACKSSDLWALGCIIYQLVAGLPPFRAGNEYLIFQKIIKLEYDFPEKFFPKARDLVEKLLVLDATKRLGCEEMEGYGPLKAHPFFESITWENLHQQTPPKLTAYLPAMSEDDEDCYGNYDNLLSQFGCMQVSSSSSSHSLCAVDASLPQRSGSNIEQYIHDLDTNSFELDLQFSEDEKRLLLEKQAGGNPWHQFVENNLILKMGPVDKRKGLFARRRQLLLTEGPHLYYVDPVNKVLKGEIPWSQELRPEAKNFKTFFVHTPNRTYYLMDPSGNAHKWCRKIQEVWRQQYQSSPDAAVQ@LTHAGWGQGWTLARTRSLLIMLGPGSNRRRPTQGERGPGSPGEPMEKYQVLYQLNPGALGVNLVVEEMETKVKHVIKQVECMDDHYASQALEELMPLLKLRHAHISVYQELFITWNGEISSLYLCLVMEFNELSFQEVIEDKRKAKKIIDSEWMQNVLGQVLDALEYLHHLDIIHRNLKPSNIILISSDHCKLQDLSSNVLMTDKAKWNIRAEEDPFRKSWMAPEALNFSFSQKSDIWSLGCIILDMTSCSFMDGTEAMHLRKSLRQSPGSLKAVLKTMEEKQIPDVETFRNLLPLMLQIDPSDRITIKDVVHITFLRGSFKSSCVSLTLHRQMVPASITDMLLEGNVASILEVMQKFSGWPEVQLRAMKRLLKMPADQLGLPWPPELVEVVVTTMELHDRVLDVQLCACSLLLHLLGQALVHHPEAKAPCNQAITSTLLSALQSHPEEEPLLVMVYSLLAITTTQESESLSEELQNAGLLEHILEHLNSSLESRDVCASGLGLLWALLLDGIIVNKAPLEKVPDLISQVLATYPADGEMAEASCGVFWLLSLLGCIKEQQFEQVVALLLQSIRLCQDRALLVNNAYRGLASLVKVSELAAFKVVVQEEGGSGLSLIKETYQLHRDDPEVVENVGMLLVHLASYEEILPELVSSSMKALLQEIKERFTSSLVSDSSAFSKPGLPPGGSPQLGCTTSGGLE@MDYYSQGTFQNIMENKRKLKAVVDTEWMHTMLSQVLDAIEYLHKLNIVHRNLKPSNIVLVNSGYCKLQDMSSQALMTHEAKWNVRAEEDPCQKSWMAPEALKFSFSTKSDIWSLGCIILDMATCSFLNDTEAMQLRKAIRHHPGSLKPILKTMEEKQIPGTDVYYLLLPFMLHINPSDRLAIKDVMQVTFMSNSFKSSSVALNMQRQKVPIFITDVLLEGNMANILDVMQNFSSRPEVQLRAINKLLTMPEDQLGLPWPTELLEEVISIIKQHGRILDILLSTCSLLLRVLGQALAKDPEAEIPRSSLIISFLMDTLRSHPNSERLVNVVYNVLAIISSQGQISEELEEEGLFQLAQENLEHFQEDRDICLSILSLLWSLLVDVVTVDKEPLEQLSGMVTWVLATHPEDVEIAEAGCAVLWLLSLLGCIKESQFEQVVVLLLRSIQLCPGRVLLVNNAFRGLASLAKVSELVAFRIVVLEEGSSGLHLIQDIYKLYKDDPEVVENLCMLLAHLTSYKEILPEMESGGIKDLVQVIRGRFTSSLELISYADEILQVLEANAQPGLQEDQLEPPAGQEAPLQGEPLFRP@MEHSVPKNKLKKLSEDSLTKQPEEVFDVLEKLGEGSYGSVFKAIHKESGQVVAIKQVPVESDLQEIIKEISIMQQCDSPYVVKYYGSYFKNTDLWIVMEYCGAGSVSDIIRLRNKTLTEDEIATVLKSTLKGLEYLHFMRKIHRDIKAGNILLNTEGHAKLADFGVAGQLTDTMAKRNTVIGTPFWMAPEVIQEIGYNCVADIWSLGITSIEMAEGKPPYADIHPMRAIFMIPTNPPPTFRKPEHWSDDFTDFVKKCLVKNPEQRATATQLLQHPFIVGAKPVSILRDLITEAMDMKAKRQQEQQRELEEDDENSEEEVEVDSHTMVKSGSESAGTMRATGTMSDGAQTMIEHGSTMLESNLGTMVINSDDEEEEEDLGSMRRNPTSQQIQRPSFMDYFDKQDSNKAQEGFNHNQQDPCLISKTAFPDNWKVPQDGDFDFLKNLDFEELQMRLTALDPMMEREIEELRQRYTAKRQPILDAMDAKKRRQQNF@PTRPTRLIVSNFSQAKQKSHLVDPQILRDQSRLAPEIITATQYKKCDEFQTGILIYEMLHLPNPFDENPELKEKEYTRTDLPRIPLRSPYSWGLQQLASCLLNPNPSERILISDAKGILQCLLWGPREDLFQIFTTSATLAQKNALLQNWLDIKRTLLMIKFAEKSLDREGGISLEDWLCAQYLAFATTDSLSYIVKILQYR@MQNKENREPRVQQTPSAGVGPLRVEMNPDTHAVSGPGRVPVKSNSKVLSIDDFDIGRPLGKGKFGNVYLARERKLKVVIALKVLFKSQMVKEGVEHQLRREIEIQSHLRHPNILRFYNYFHDDTRVFLILEYAPRGEMYKELQRYGRFDDQRTATYMEEVSDALQYCHEKKVIHRDIKPENLLLGYRGELKIADFGWSVHAPSLRRRTMCGTLDYLPPEMIEGHSHDEKVDLWSIGVLCYECLVGNPPFETASHAETYKRITKVDLQFPKLVSEGARDLISKLLRHSPSMRLPLRSVMEHPWVKANSRRVLPPVCSSEPH

Creating data.in...

Contents of data.in:

6978799 0

45768856 509

45768786 943

27923854 1367

45768720 1765

45768758 2292

45219906 2639

18202599 3311

4835224 4200

40787731 4571

18202068 5189

34191428 5749

29747774 6450

28856169 7038

20071571 7531

45709347 7734

Q6. **Program Output:**

Getting contents of Q6query.txt...

Query is: MHIQITDFGTAKVLSPDS

gi # of database sequence containing query is: 18202068

Q7.

I spent around 12 hours total setting up python, learning basic

python for the assignment, and doing the problems. The actual

assignment took around 5 hours to code and package for submission.

I posted on piazza for clarification about the Q3 header, but did

not ask anyone else for help.