February 6, 2012

Ran the following code after creating my usual PyMOL session with ezbeta\_api.py:

import csv

from sundries import one\_letter

stored.one\_letter = one\_letter

output = list()

for name, group in groupdict.items():

stored.sequence = ''

cmd.iterate('{}.molecule & n. ca & polymer'.format(name),

'stored.sequence += stored.one\_letter[resn]')

output.append((name, stored.sequence))

with open(r'C:\cygwin\home\alex\temp\sequences.csv', 'wb') as f:

fwriter = csv.writer(f)

fwriter.writerows(output)

from this I received the output

2POR,EVKLSGDARMGVMYNGDDWNFSSRSRVLFTMSGTTDSGLEFGASFKAHESVGAETGEDGTVFLSGAFGKIEMGDALGASEALFGDLYEVGYTDLDDRGGNDIPYLTGDERLTAEDNPVLLYTYSAGAFSVAASMSDGKVGETSEDDAQEMAVAAAYTFGNYTVGLGYEKIDSPDTALMADMEQLELAAIAKFGATNVKAYYADGELDRDFARAVFDLTPVAAAATAVDHKAYGLSVDSTFGATTVGGYVQVLDIDTIDDVTYYGLGASYDLGGGASIVGGIADNDLPNSDMVADLGVKFKF

1THQ,TTFRENIAQTWQQPEHYDLYIPAITWHARFAERPWGGGFGLSRWDEKGNWHGLYAMAFKDSWNKWEPIAGYGWESTWRPLADENFHLGLGFTAGVTARDNWNYIPLPVLLPLASVGYGPVTFQMTYIPGTYNNGNVYFAWMRFQFLE

1QFG,ESAWGPAATIAARQSATGTKTDTPIQKVPQSISVVTAEEMALHQPKSVKEALSYTPGVSVGTRGASNTYDHLIIRGFAAEGQSQNNYLNGLKLQGNFYNDAVIDPYMLERAEIMRGPVSVLYGKSSPGGLLNMVSKRPTTEPLKEVQFKAGTDSLFQTGFDFSDSLDDDGVYSYRLTGLARSANAQQKGSEEQRYAIAPAFTWRPDDKTNFTFLSYFQNEPETGYYGWLPKEGTVEPLPNGKRLPTDFNEGAKNNTYSRNEKMVGYSFDHEFNDTFTVRQNLRFAENKTSQNSVYGYGVCSDPANAYSKQCAALAPADKGHYLARKYVVDDEKLQNFSVDTQLQSKFATGDIDHTLLTGVDFMRMRNDINAWFGYDDSVPLLNLYNPSSHHHHHHGSSVNTDFDFNAKDPANSGPYRILNKQKQTGVYVQDQAQWDKVLVTLGGRYDWADQESLNRVAGTTDKRDDKQFTWRGGVNYLFDNGVTPYFSYSESFEPSSQVGKDGNIFAPSKGKQYEVGVKYVPEDRPIVVTGAVYNLTKTNNLMADPEGSFFSVEGGEIRARGVEIEAKAALSASVNVVGSYTYTDAEYTTDTTYKGNTPAQVPKHMASLWADYTFFDGPLSGLTLGTGGRYTGSSYGDPANSFKVGSYTVVDALVRYDLARVGMAGSNVALHVNNLFDREYVASCFNTYGCFWGAERQVVATATFRF

1KMO,ALTVVGDWLGDARENDVFEHAGARDVIRREDFAKTGATTMREVLNRIPGVSAPENNGTGSHDLAMNFGIRGLNPRLASRSTVLMDGIPVPFAPYGQPQLSLAPVSLGNMDAIDVVRGGGAVRYGPQSVGGVVNFVTRAIPQDFGIEAGVEGQLSPTSSQNNPKETHNLMVGGTADNGFGTALLYSGTRGSDWREHSATRIDDLMLKSKYAPDEVHTFNSLLQYYDGEADMPGGLSRADYDADRWQSTRPYDRFWGRRKLASLGYQFQPDSQHKFNIQGFYTQTLRSGYLEQGKRITLSPRNYWVRGIEPRYSQIFMIGPSAHEVGVGYRYLNESTHEMRYYTATSSGQLPSGSSPYDRDTRSGTEAHAWYLDDKIDIGNWTITPGMRFEHIESYQNNAITGTHEEVSYNAPLPALNVLYHLTDSWNLYANTEGSFGTVQYSQIGKAVQSGNVEPEKARTWELGTRYDDGALTAEMGLFLINFNNQYDSNQTNDTVTARGKTRHTGLETQARYDLGTLTPTLDNVSIYASYAYVNAEIREKGDTYGNLVPFSPKHKGTLGVDYKPGNWTFNLNSDFQSSQFADNANTVKESADGSTGRIPGFMLWGARVAYDFGPQMADLNLAFGVKNIFDQDYFIRSYDDNNKGIYAGQPRTLYMQGSLKF

1A0S,SGFEFHGYARSGVIMNDSGASTKSGAYITPAGETGGAIGRLGNQADTYVEMNLEHKQTLDNGATTRFKVMVADGQTSYNDWTASTSDLNVRQAFVELGNLPTFAGPFKGSTLWAGKRFDRDNFDIHWIDSDVVFLAGTGGGIYDVKWNDGLRSNFSLYGRNFGDIDDSSNSVQNYILTMNHFAGPLQMMVSGLRAKDNDERKDSNGNLAKGDAANTGVHALLGLHNDSFYGLRDGSSKTALLYGHGLGAEVKGIGSDGALRPGADTWRIASYGTTPLSENWSVAPAMLAQRSKDRYADGDSYQWATFNLRLIQAINQNFALAYEGSYQYMDLKPEGYNDRQAVNGSFYKLTFAPTFKVGSIGDFFSRPEIRFYTSWMDWSKKLNNYASDDALGSDGFNSGGEWSFGVQMETWF

3EMN,MAVPPTYADLGKSARDVFTKGYGFGLIKLDLKTKSENGLEFTSSGSANTETTKVNGSLETKYRWTEYGLTFTEKWNTDNTLGTEITVEDQLARGLKLTFDSSFSPNTGKKNAKIKTGYKREHINLGCDVDFDIAGPSIRGALVLGYEGWLAGYQMNFETSKSRVTQSNFAVGYKTDEFQLHTNVNDGTEFGGSIYQKVNKKLETAVNLAWTAGNSNTRFGIAAKYQVDPDACFSAKVNNSSLIGLGYTQTLKPGIKLTLSALLDGKNVNAGGHKLGLGLEFQA

3DWO,AGFMVPTTNTAGWGRAMAGGSLFPNDPSAAFNNPAAMAFIDKRIAQLTVNYADIDIKYNGDAYDYQGNPMTGGYQDGPGTPELGTNDGGQAGFGAWLPTGFLVVPINDRFAFGLSQVVPMGMRSTWDPNWKGRDFAVDTKIETIGLTGSLSFKVNDNFSLGAGVIIQRTSGFVSQNLDLYASAANSPGMGGIPFPASNSSALMRVKVDNTSPGFFAGAVWKPTDRDTLGFAYHAKIRNKLKGHYNLYDHDGGLTEGAIEGGTPGLAYPGLDLRMGASASARLDIPAYASLDWVHQFNDRLSLGASATWTEWSSFQDLTLKSHGNTIVSIPYTYRNTWTLAVGGDYKVTDQWTMRAGVAYDQTPTHNATRDPRIPDGDRYFASLGAGYRFQSMPELSIDAAYSRQFVKEVPLKTVNQDRLGGGRLDGRATSKGQVFSLSATYDFH

2WJR,ALDVRGGYRSGSHAYETRLKVSEGWQNGWWASMESNTWNTINDVQVEVNYAIKLDDQWTVRPGMLTHFSSNGTRYGPYVKLSWDATKDLNFGIRYRYDWKAYRQQDLSGDMSRDNVHRWDGYVTYHINSDFTFAWQTTLYSKQNDYRYANHKKWATENAFVLQYHMTPDITPYIEYDYLDRQGVYNGRDNLSENSYRIGVSFKL

2F1V,EAGEFFMRAGSATVRPTEGGFSVTNNTQLGLTFTYMATDNIGVELLAATPFRHKIGTRATGDIATVHHLPPTLMAQWYFGDASSKFRPYVGAGINYTTFFDNGFNDHGKEAGLSDLSLKDSWGAAGQVGVDYLINRDWLVNMSVWYMDIDTTANYKLGGAQQHDSVRLDPWVFMFSAGYRFH

2ERV,ADVSAAVGATGQSGMTYRLGLSWDWDKSWWQTSTGRLTGYWDAGYTYWEGGDEGAGKHSLSFAPVFVYEFAGDSIKPFIEAGIGVAAFSGTRVGDQNLGSSLNFEDRIGAGLKFANGQSVGVRAIHYSNAGLKQPNDGIESYSLFYKIPI

1E54,SSVTLFGIVDTNVAYVNKDAAGDSRYGLGTSGASTSRLGLRGTEDLGGGLKAGFWLEGEIFGDDGNASGFNFKRRSTVSLSGNFGEVRLGRDLVPTSQKLTSYDLFSATGIGPFMGFRNWAAGQGADDNGIRANNLISYYTPNFGGFNAGFGYAFDEKQTIGTADSVGRYIGGYVAYDNGPLSASLGLAQQKTAVGGLATDRDEITLGASYNFGVAKLSGLLQQTKFKRDIGGDIKTNSYMLGASAPVGGVGEVKLQYALYDQKAIDSKAHQITLGYVHNLSKRTALYGNLAFLKNKDASTLGLQAKGVYAGGVQAGESQTGVQVGIRHAF

1TLY,LSDWWHQSVNVVGSYHTRFGPQIRNDTYLEYEAFAKKDWFDFYGYADAPVPLFMEIEPRFSIDKLTNTDLSFGPFKEWYFANNYIYDMGRNKDGRQSTWYMGLGTDIDTGLPMSLSMNVYAKYQWQNYGAANENEWDGYRFKIKYFVPITDLWGGQLSYIGFTNFDWGSDLGDDSGNAINGIKTRTNNSIASSHILALNYDHWHYSVVARYWHDGGQWNDDAELNFGNGNFNVRSTGWGGYLVVGYNFHHH

1I78,STETLSFTPDNINADISLGTLSGKTKERVYLAEEGGRKVSQLDWKFNNAAIIKGAINWDLMPQISIGAAGWTTLGSRGGNMVDQDWMDSSNPGTWTDEARHPDTQLNYANEFDLNIKGWLLNEPNYRLGLMAGYQESRYSFTARGGSYIYSSEEGFRDDIGSFPNGERAIGYKQRFKMPYIGLTGSYRYEDFELGGTFKYSGWVESSDNDEHYDPKGRITYRSKVKDQNYYSVAVNAGYYVTPNAKVYVEGAWNRVTNKKGNTSLYDHNNNTSDYSKNGAGIENYNFITTAGLKYTF

1QJ8,ATSTVTGGYAQSDAQGQMNKMGGFNLKYRYEEDNSPLGVIGSFTYTEKSRTASSGDYNKNQYYGITAGPAYRINDWASIYGVVGVGYGKFQTTEYPTYKNDTSDYGFSYGAGLQFNPMENVALDFSYEQSRIRSVDVGTWIAGVGYRF

2F1C,NDWHFNIGAMYEIENVGLAEPSVYFNAANGPWRIALAYYQEGPVDYSKRGTWFDRPELEVHYQFLENDDFSFGLTGGFRNYGYHYVDEPGKDTANMQRWKIAPDWDVKLTDDLRFNGWLSMYKFANDLNTTGYADTRVETETGLQYTFNETVALRVNYYLERGFNMDDSRNNGEFSTQEIRAYLPLTLGNHSVTPYTRIGLDRWSNDFNRVGLFYGYDFQNGLSVSLEYAFEWQDDKFHYAGVGVNYSFHHH

3JTY,GFLEDAQANLTLRNFYFNRNFTNPTKAQGKAEEWTQSFILDAKSGFTQGTVGFGMDVLGLYSLKLDGGKGTGGTQLLPLDHDGRPADNFGRLGVAFKARLSQTEVKVGEWMPVLPILRSDDGRSLPQTFRGGQITSKEIAGLTLYGGQFRANSPRDDSSMSDMSMFGKAAFTSDRFNFQGAEYAFNDKRTQIALWNAQLKDIYSQQFINLIHSQPLGDWTLGANLGFFYGKEDGSARAGDMENRTWSGLFSAKYGGNTFYVGLQKLTGDSAWMRVNGTSGGTLANDSYNASYDNAKEKSWQVRHDYNFAALGVPGLTLMNRYISGSNVHTATVSDGKEWGRESEVAYTVQSGTLKNLNLKWRNSTMRRDFSNNEFDENRLIISYPLSL

1T16,AGFQLNEFSSSGLGRAYSGEGAIADDAGNVSRNPALITMFDRPTFSAGAVYIDPDVNISGTSPSGRSLKADNIAPTAWVPNMHFVAPINDQFGWGASITSNYGLATEFNDTYAGGSVGGTTDLETMNLNLSGAYRLNNAWSFGLGFNAVYARAKIERFAGDLGQLVAGQIMQSPAGQTQQGQALAATANGIDSNTKIAHLNGNQWGFGWNAGILYELDKNNRYALTYRSEVKIDFKGNYSSDLNRAFNNYGLPIPTATGGATQSGYLTLNLPEMWEVSGYNRVDPQWAIHYSLAYTSWSQFQQLKATSTSGDTLFQKHEGFKDAYRIALGTTYYYDDNWTFRTGIAFDDSPVPAQNRSISIPDQDRFWLSAGTTYAFNKDASVDVGVSYMHGQSVKINEGPYQFESEGKAWLFGTNFNYAFHHHHHH

3CSL,NDWVYDEPRSVSVISREQMDNRPARHAADILEQTTGAYSSVSQQDPALSVNIRGIQDYGRVNMNIDGMRQNFQKSGHGQRNGTMYIDSELLSGVTIDKGTTGGMGSAGTLGGIATFNTVSASDFLAPGKELGGKLHASTGDNGTHFIGSGILALGNETGDILLAASERHLGDYWPGNKGDIGNIRINNDTGNYDRYAESIKNNKIPDTHYRMHSRLAKVGWNLPANQRLQLSYLQTQTASPIAGTLTNLGTRPPYELGWKRTGYTDVMARNAAFDYSLAPEDVDWLDFQAKLYYVDTQDDSDTYSTSSLLDNGYATRTRLRTYGAQAQNTSRFSLAPGHDFRANYGLEFYYDKATSDSSRQGMEGVTPAGNRSVASLFANLTYDYDGWLTLEGGLRYDRYRLRGQTGLSYPDLAKDGQRYTIDNPCKALRLTGCSTTTREDWDVDRDQGKLSPTLAVAVRPGVEWLELYTTYGKSWRPPAITETLTNGSAHSSSTQYPNPFLQPERSRAWEVGFNVQQPDLWFEGDRLVAKVAYFDTKVDNYINLAIDRNKPGLVQPSIGNAAYVNNLSKTRFRGLEYQLNYDAGVFYADLTYTHMIGKNEFCSNKAWLGGRLRYGDGSRRGNFYVEPDAASNDFVTCDGGTQFGSAAYLPGDRGSVTLGGRAFDRKLDAGVTVRFAPGYQDSSVPSNYPYLADWPKYTLFDLYASYKLTDSLTLRGSVENLTNRAYVVSYGETLANTLGRGRTVQGGVEYRF

1QJP,APKDNTWYTGAKLGWSQHENKLGAGAFGGYQVNPYVGFEMGYDWLGRMPYAYKAQGVQLTAKLGYPITDDLDIYTRLGGMVWRADTYSNVYGKNHDTGVSPVFAGGVEYAITPEIATRLEYQWTNGMLSLGVSYRFG

1AF6,VDFHGYARSGIGWTGSGGEQQCFQTTGAQSKYRLGNECETYAELKLGQEVWKEGDKSFYFDTNVAYSVAQQNDWEATDPAFREANVQGKNLIEWLPGSTIWAGKRFYQRHDVHMIDFYYWDISGPGAGLENIDVGFGKLSLAATRSSEAGGSSSFASNNIYDYTNETANDVFDVRLAQMEINPGGTLELGVDYGRANLRDNYRLVDGASKDGWLFTAEHTQSVLKGFNKFVVQYATDSMTSQGKGLSQGSGVAFDNEKFAYNINNNGHMLRILDHGAISMGDNWDMMYVGMYQDINWDNDNGTKWWTVGIRPMYKWTPIMSTVMEIGYDNVESQRTGDKNNQYKITLAQQWQAGDSIWSRPAIRVFATYAKWDEKWGYDYTGNADNNANFGKAVPADFNGGSFGRGDSDEWTFGAQMEIWW

2J1N,AEVYNKDGNKLDLYGKVDGLHYFSDNKDVDGDQTYMRLGFKGETQVTDQLTGYGQWEYQIQGNSAENENNSWTRVAFAGLKFQDVGSFDYGRNYGVVYDVTSWTDVLPEFGGDTYGSDNFMQQRGNGFATYRNTDFFGLVDGLNFAVQYQGKNGNPSGEGFTSGVTNNGRDALRQNGDGVGGSITYDYEGFGIGGAISSSKRTDAQNTAAYIGNGDRAETYTGGLKYDANNIYLAAQYTQTYNATRVGSLGWANKAQNFEAVAQYQFDFGLRPSLAYLQSKGKNLGRGYDDEDILKYVDVGATYYFNKNMSTYVDYKINLLDDNQFTRDAGINTDNIVALGLVYQF

1P4T,EGASGFYVQADAAHAKASSSLGSAKGFSPRISAGYRINDLRFAVDYTRYKNYKAPSTDFKLYSIGASAIYDFDTQSPVKPYLGARLSLNRASVDLGGSDSFSQTSIGLGVLTGVSYAVTPNVDLDAGYRYNYIGKVNTVKNVRSGELSAGVRVKF

3PRN,MISLNGYGRFGLQYVEDRGVGLEDTIISSRLRINIVGTTETDQGVTFGAKLRMQWDDGDAFAGTAGNAAQFWTSYNGVTVSVGNVDTAFDSVALTYDSEMGYEWSSFGDAQSSFFAYNSKYDASGALDNYNGIAVTYSISGVNLYLSYVDPDQTVDSSLVTEEFGIAADWSNDMISLAAAYTTDAGGIVDNDIAFVGAAYKFNDAGTVGLNWYDNGLSTAGDQVTLYGNYAFGATTVRAYVSDIDRAGADTAYGIGADYQFAEGVKVSGSVQSGFANETVADVGVRFDF

1UYN,AATVYADSTAAHADMQGRRLKAVSDGLDHNGTGLRVIAQTQQDGGTWEQGGVEGKMRGSTQTVGIAAKTGENTTAAATLGMGRSTWSENSANAKTDSISLFAGIRHDAGDIGYLKGLFSYGRYKNSISRSTGADEHAEGSVNGTLMQLGALGGVNVPTGDLTVEGGLRYDLLKQDAFAEKGSALGWSGNSLTEGTLVGLAGLKLSQPLSDKAVLFATAGVERDLNGRDYTVTGPHTRLVAGLGADVEFGNGWNGLARYSYAGSKQYGNHSGRVGVGYRF

2QDZ,QLLPGARDLNRIDDRQRKEQLQRDIERAATSGHTVTVHAVDLDFGVEGRLFDPAPLVQDYLNRPLDNEQLFLLVKALSAALYDRGYATSIVTFVPPGVVDGVLKLKVEWGRIKGWLIDGKPLEGTRDRMMVFSAMPGWQDKVLNVFDIDQAIYNINNGGKTGNITIVPADEYGYSYLDLQLQRRALPRVSLGMDNSGPGTPENGRYKYNASVTANDLLGLNDTLGLYIGNRYYRDAGHDAERNYDLMYSVPLGRTRLDLQTGYSTYRNLLKTRYGQYQSAGNSRSFGLKATRLLYRDTRSQFSVYGGLKLRQNKNYLAGTRLDVSSKHYSDVTVGMQYSTQRGANAYFGDLSFTRGVGVGNVSRFNGSLAWTRYMALAGQPIQWASQLGFQYSRQQLLNSYQITVGDEYTVRGYNLRTSQSGDSGVYLSNTLTVPVQFSLLGKQASVAPFVGADVGALKSNHPDARTIRMAGLAAGVRFDLPYARMSFTYSKPVGAQPGGAPRAPVWLYINAGLSF

3DZM,HAAKFSVEAGAGFYGGFGGQLAVVAEDLAPGLPLGVRLGVGFATSDALDDGYDLGGGTTWGDVKEAGKFSEWGQNVTLSLDVLYKPSGLGLPVEVAPYFGVRYNFFSGGYTDPEDNLTIKAQTISSNQLGLGLGVRAAYPLMPNLSLVGDLGVDYYFQACFTRVEEDDSGNKSQSSVCPGDSGYEDVNKFVTQPEWVLKLRLGAAYRF

2O4V,GTVTTDGADIVIKTKGGLEVATTDKEFSFKLGGRLQADYGRFDGYYTNNGNTADAAYFRRAYLEFGGTAYRDWKYQINYDLSRNVGNDSAGYFDEASVTYTGFNPVNLKFGRFYTDFGLEKATSSKWVTALERNLTYDIADWVNDNVGTGIQASSVVGGMAFLSGSVFSENNNDTDGDSVKRYNLRGVFAPLHEPGNVVHLGLQYAYRDLEDSAVDTRIRPRMGMRGVSTNGGNDAGSNGNRGLFGGSSAVEGLWKDDSVWGLEGAWALGAFSAQAEYLRRTVKAERDREDLKASGYYAQLAYTLTGEPRLYKLDGAKFDTIKPENKEIGAWELFYRYDSIKVEDDNIVVDSATREVGDAKGKTHTLGVNWYANEAVKVSANYVKAKTDKISNANGDDSGDGLVMRLQYVF

2GUF,QDTSPDTLVVTANRFEQPRSTVLAPTTVVTRQDIDRWQSTSVNDVLRRLPGVDITQLSSIFIRGTNASHVLVLIDGVRLNLAGVSGSADLSQFPIALVQRVEYIRGPRSAVYGSDAIGGVVNIITTRDEPGTEISAGWGSNSYQNYDVSTQQQLGDKTRVTLLGDYAHTHDGFLSKTLYGALEHNFTDAWSGFVRGYGYDNRTNYDTRKLYSQSWDAGLRYNGELIKSQLITSYSHSKDYNYDPHYGRYDSSATLDEMKQYTVQWANNVIVGHGSIGAGVDWQKQTTTPGTGYVEDGYDQRNTGIYLTGLQQVGDFTFEGAARSDDNSQFGRHGTWQTSAGWEFIEGYRFIASYGTSYKAPNLGQLYGFYGNPNLDPEKSKQWEGAFEGLTAGVNWRISGYRNDVSDLIDYDDHTLKYYNEGKARIKGVEATANFDTGPLTHTVSYDYVDARNAITDTPLLRRAKQQVKYQLDWQLYDFDWGITYQYLGTRYDKDYSSYPYQTVKMGGVSLWDLAVAYPVTSHLTVRGKIANLFDKDYAGREYTLSGSYTF

3BS0,DLEGYGAISRAMGGTSSSYYTGNAALISNPATLSFAPDGNQFELGLDVVTTDIKVHDSHGAEAKSSTGPYVGPQLSYVAQLDDWRFGAGLFVSSGLGTEYGSKSFLSQTENGIQTSFDNSSRLIVLRAPIGFSYQATSKLTFGASVDLVWTSLNLELLLPSSQVGALTAQGNLSGGLVPSLAGFVGTGGAAHFSLSRNSTAGGAVDAVGWGGRLGLTYKLTDNTVLGAMYNFKTSVGDLEGKATLSAISGDGAVLPLDGDIRVKNFEMPASLTLGLAHQFNERWVVAADIKRAYWGDVMDMNVAFISQLGGIDVALPHRYQDITVASIGTAYKYNNDLTLRAGYSYAQLILPVIPAYLKRHVTFGGEYDFDKDSRINLAISFGLRERVQTTEMLRQSHSQINAVVSYSKNFHHH

3FHH,TETMTVTATGNARSSFEAPMMVSVIDTSAPENQTATSATDLLRHVPGITLDGTGRTNGQDINMRGYDHRGVLVLVDGIRQGTDTGHLNGTFLDPALIKRVEIVRGPSALLYGSGALGGVISYDTVDAKDLLQEGQSSGFRVFGTGGTGDHSLGLGASAFGRTENLDGIVAWSSRDRGDLRQSNGETAPNDESINNMLAKGTWQIDSAQSLSGLVRYYNNDAREPKNPQTVEASESSNPMVDRSTIQRDAQLSYKLAPQGNDWLNADAKIYWSEVRINAQEYREQITKGARLENRSTLFADSFASHLLTYGGEYYRQEQGSHTTGFPQAKIDFSSGWLQDEITLRDLPITLLGGTRYDSYRGSSDGYKDVDADKWSSRAGMTINPTNWLMLFGSYAQAFRAPTMGEMYNDSKHFSIGRFYTNYWVPNPNLRPETNETQEYGFGLRFDDLMLSNDALEFKASYFDTKAKDYISTTVDFAAATTMSYNVPNAKIWGWDVMTKYTTDLFSLDVAYNRTRGKDTDTGEYISSINPDTVTSTLNIPIAHSGFSVGWVGTFADRSTHISSKQPGYGVNDFYVSYQGQQALKGMTTTLVLGNAFDKEYWSPQGIPQDGRNGKIFVSYQW

1QD6,FTLYPYDTNYLIYTQTSDLNKEAIASYDWAENARKDEVKFQLSLAFPLWRGILGPNSVLGASYTQKSWWQLSNSEESSPFRETNYEPQLFLGFATDYRFAGWTLRDVEMGYNHDSNGRSDPTSRSWNRLYTRLMAENGNWLVEVKPWYVVGNTDDNPDITKYMGYYQLKIGYHLGDAVLSAKGQYNWNTGYGGAELGLSYPITKHVRLYTQVYSGYGESLIDYNFNQTRVGVGVMLNDLF

1K24,QTANEFTVHTDLSSISSTRAFLKEKHKAAKHIGVRADIPFDANQGIRLEAGFGRSKKNIINLETDENKLGKTKNVKLPTGVPENRIDLYTGYTYTQTLSDSLNFRVGAGLGFESSKDSIKTTKHTLHSSRQSWLAKVHADLLSQLGNGWYINPWSEVKFDLNSRYKLNTGVTNLKKDINQKTNGWGFGLGANIGKKLGESASIEAGPFYKQRTYKESGEFSVTTKSGDVSLTIPKTSIREYGLRVGIKF

1FEP,DDTIVVTAAEQNLQAPGVSTITADEIRKNPVARDVSKIIRTPGVNLTGNSTSGQRGNNRQIDIRGGPENTLILIDGKPVSSRNSVRQGWRGERDTRGDTSWVPPEIERIEVLRGPAAARYGNGAAGGVVNIITKKGSGEWHGSWDAYFNAPEHKEEGATKRTNFSLTGPLGDEFSFRLYGNLDKTQADAWDINQGHQSARAGTYATTLPAGREGVINKDINGVVRWDFAPLQSLELEAGYSRQGNLYAGDTQNTNSDSYTRSKYGDETNRLYRQNYALTWNGGWDNGVTTSNWVQYEHTRNSRIPEGLAGQDFVDIDLDDVLHSEVNLPIDFLVNQTLTLGTEWNQQRKDLSSNTQADRSPYSKAEIFSLFAENNELTDSTIVTPGLRFDHHSIVGNNWSPALNISQGLGDDFTLKGIARAYKAPSLYQTNPNYILYSKGQGCYLQGNDDLKAETSINKEIGLEFKRDGWLAGVTWFRNDYRNKIEAGYVAVGQNAVGTDLYQWDNVPKAVVEGLEGSLNVPVSETVWTNNITYLKSENKTTGDRLSIIPEYTLNSTLSWQAREDLSQTTFTWYGKQQPKKYNYKGQPAVGPETKEISPYSIVGLSATWDVTKNVSLTGGVDNLFDKRLWRAGNAQTTGDLAGANYIAGAGAYTYNEPGRTWYSVNTHF

2VQI,LLDYNLNGTVSRNYQGGDSHQFSYNGTVGGNLPWRLRADYQGSQEQSRYNGEKTTNRNFTWSRFYLFRAIPRWRANLTLGENNINSDIFRSWSYTGASLESLPPRLRGYAPQITGIAETNARVVVSQQGRVLYDSVPAGPFSIQDLDSSVRGRLDVEVIEQNGRKKTFQVDTASVPYLTRPGQVRYKLVSGRSRGYEGPVFATGEASWGLSNQWSLYGGAVLAGDYNALAAGAGWDLGVPGTLSADITQSVARIEGERTFQGKSWRLSYSKRFDDITFAGYRFSERNYTEQYLNARYRNDYSSREKEYTVTLNKNVADWNTSFNLQYSRQTYWDIRKTDYYTVSVNRYFNVFLQGVAVGLSASRSKYLGRDNDSAYLRISVPLGTGTASYSGSSNDRYVNAGYTDTFNDGLDSYSLNAGLNSGLTSQRQINAYYSHRSPLANLSANIASLQKGYTSFGVSASGGATITG

3EFM,EISDLTEGTNAYTTEAMSTATGLTLSPRETPQSVSVVTRQQIEDQGLTDTGAILATAPGISVTRSDSNRYSFSARGFTIDNFQFDGLVSPILSQWNYGSTDMDAAIYDHVEIVRGATGLMTGSGNPSAAVNFVRKRPLREFAATFNASVGSWDYVRGDADISVPITEDGRIRSRLVAAYSQGDLDTRRRTFYGVVSADLTPDTVLTTSVEYQHNHSNGPWARQDTEATTYFVDLTHRFTNDWKLRAAYSHTDGRYLMKHVFSNYDGNLDRDDIHFSLSAPFEAFGLRHEVALGWMSIDNHSDIQRYAMTLSPADDVRTKQTGAYLVGRFALAEPLHLIVGDRWSDWKTKQMYFGSRREYRIKNQFTPYAGLTYDINDTYTAYASYTEIFQPQNARDTSGGILPPIKSKSYELGLKAAYLEGRLNTSAALFQTRQDNLAQVIPGSSIPGFPNMQASRAASGAKVEGIDLEASGQILPDWNIGASYTHFTTKDASGNPINTNHPRSLFKLYTTYRLPGALHRLTVGGGVDWQSRMQDSYALVSLMARFDFNKKLSATLNVNNLRNVMLNLRAQY

I entered each sequence as an HHOMP search using the default settings, which are

Database: HHompDB\_1.0

Max. PSI-BLAST iterations: 8

E-value threshold for PSI-BLAST: 1E-3

Min. coverage of PSI-BLAST hits: 20

Min. sequence identity of PSI-BLAST hits with query: 0

Score beta-barrel structure: yes

Alignment mode: local

I clicked the "Save" button on each HHOMP output page, and saved it in the folder "hhomp search results" with the pdbid in the filename.

There were only 34 files, but I had 35 sequences.

I used the python function os.walk to generate a list of filenames in that folder.

I ran the following code from within my ezbeta\_api.py session to check which one was missing:

missing = []

for pdbid in groupdict.keys():

number\_in = 0

for filename in x:

if pdbid in filename:

number\_in += 1

if number\_in > 1:

raise Exception('found too many of ' + pdbid)

if number\_in == 0:

missing.append(pdbid)

I received the following input:

PyMOL>print missing

['2F1V', '1AF6']

Turns out I had mislabeled 2F1V as 1F1V, and hadn't entered 1AF6's sequence as an HHOMP search. I renamed the mislabeled file, and ran a search for 1AF6.

Found 2POR was located to many subgroups of 16.2, and 1E54 as well.