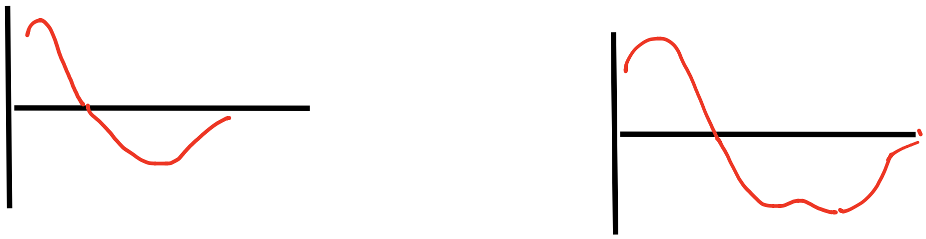
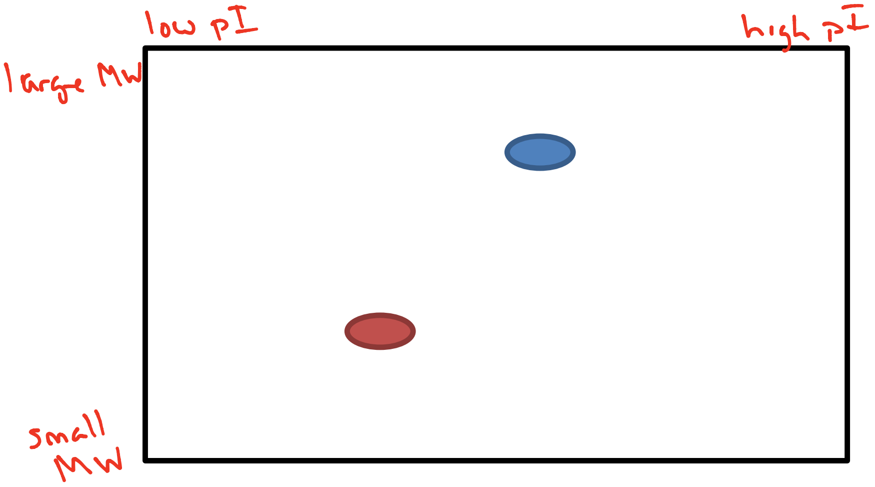
**Problem Set 4 - partial** (**Due Sept 23rd**)

1. As we discussed in class, there are several computational algorithms that allow biochemists to predict secondary structure. Let's see how good these algorithms hold up to real structural information.
   a. Access the following two pdb files: 1DNK and 7ACN.
      i. What are these proteins and what organism do they come from?
      1DNK = DNase I from Bos Taurus 7ACN = Aconitase from Sus scrofa (pig)
      ii. Exactly what position do they occupy in the genome (I showed you how to do this in class). Make sure to identify the chromosome and position.
      1DTK – chromosome 25: 2,977,213 – 2,995,476
      7ACN – chromosome 5: 4,385,951 – 4,443,595
   b. Use the Chou-Fasman prediction program on ExPASy to predict the secondary structure. Compare this prediction with the actual secondary structure observed in the pdb file. You can do this visually using Chimera or by observing the text in the pbd file – either way, map the structure vs. prediction and determine how good the prediction is. In general, there is reasonable agreement. The model gets some sections very wrong.
   c. Based on the observed secondary structure, predict what the CD spectrum of each protein might look like. Please justify your sketch. DNaseI is mostly $\beta$ sheet, so I draw it looking mostly like an all $\beta$ protein. Aconitase is about 2x more $\alpha$ helix than $\beta$ sheet, so the spectrum is drawn to reflect primarily helix.



2. Using the tools in ExPASy, determine the pI, MW, and molar absorptivity of each protein.
   1DNK pI = 5.08 MW = 29065.6 g/mol $\varepsilon_{280} = 39100\ M^{-1}cm^{-1}$
   7ACN pI = 7.20 MW = 82693.1 g/mol $\varepsilon_{280} = 80050\ M^{-1}cm^{-1}$

3. Assuming that you start with a homogenous mixture of both of these proteins (and nothing else), predict what a 2D gel electrophoresis experiment would look like. 1ACN in Blue, 7DNK in red.



4. For 1DNK, predict the MW of all peptides produced when it is treated with Cyanogen Bromide (CNBr). You are encouraged to use ExPASY to determine MW values, but you should manually determine where the peptide chain will be broken.

LKIAAFNIRTFGETKM     MW = 1840.2 g/mol<br>
SNATLASYIVRIVRRYDIVLIQEVRDSHLVAVGKLLDYLNQDDPNTYHYVVSEPLGRNSYKERYLFLFRPNKVSVLDTYQYDDGCESCGNDSFSREPAVVKFSSHSTKVKEFAIVALHSAPSDAVAEINSLYDVYLDVQQKWHLNDVM  MW = 16970 g/mol

LM  MW = 244.4

GDFNADCSYVTSSQWSSIRLRTSSTFQWLIPDSADTTATSTNCAYDRIVVAGSLLQSSVVPGSAAPFDFQAAYGLSNEM MW = 8436.2

ALAISDHYPVEVTLT      MW = 1628.8

5. Use the PeptideCutter tool in ExPASy to predict what fragments would be produced when 7ACN is digested with each of the following proteases. In this tool, make sure to select "only the following selection of enzymes and chemical" option. Also, it is quite helpful to choose the "Table of sites" display option. See attached
   a. Chymotrypsin
   b. Arg-C
   c. Thrombin – No cleavage sites

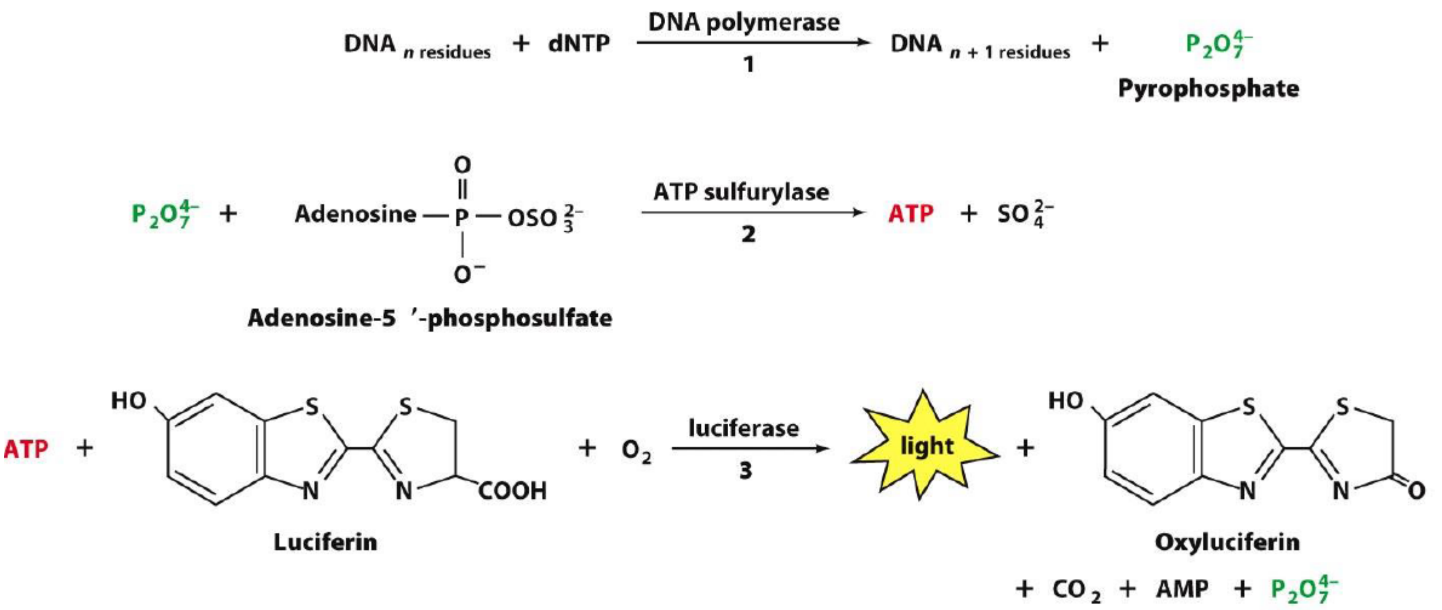
6. A tandem MS experiment results in peaks at the following m/z ratios. Determine the sequence of this peptide. GluAsn(Ile or Leu)TyrPheGlnGlyGln

| 128.2 | 185.2 | 313.3 | 460.5 | 623.7 | 736.9 | 851 | 980.1 |
|---|---|---|---|---|---|---|---|
| Gln | Gln+Gly | Gln+Gly+Gln | Q+G+Q+F | …F+Y | …Y+L or I | …L/I + N | …N+E |

A peak is also observed at m/z = 425.5. What is the source of this peak? This is the +2 peak of the 851 Da peptide. Since it carries a charge of +2, it will be observed at ½ the mass.

7. Using the attached electropherogram:
   a. Please describe how this data is generated. A PCR reaction is carried out with a small fraction of 2'3'-didexoynucleic acids. Each base on the ddNTP is modified with a chromaphore. When the modified ddNTP is incorporated into the growing DNA chain, the elongation is terminated. This terminating base is modified with the complementary base, so you know exactly what base is at a given position. These PCR fragments are separated by capillary electrophoresis and the resulting electropherogram provides single nucleotide separation so the absorbance vs. time profile can be used to determine the sequence.
   b. Determine the sequence of nucleotides 50-100 (feel free to simply highlight the sequence on the image). See attached
   c. Discuss why there are regions that are not useful and highlight those regions. This represents oligonucleotides that are too short to accurately separate by electrophoresis and/or single ddNTPs that absorb strongly.

8. What is meant by pyrosequencing? What reactions are important in this process? Pyrosequencing refers to the dependence on pyrophosphate production during the sequencing reaction. A dNTP is passed across the sequencing plate. If the complementary base is present on the next position for elongation, the coupling reaction will happen and pyrophosphate will be produced. This pyrophosphate will react with a modified AMP (sulfate on the alpha phosphate) to produce ATP. The ATP then reacts with luciferin, catalyzed by luciferase, to produce a burst of light via chemiluminescence. The slide is washed and another dNTP is added.



9. A protein is independently digested with Arg-C and Asp-N. Identify this protein.

| Asp-N | Arg-C |
|---|---|
| DSG | EIVR |
| DLT | LDLAGR |
| DIRK X | AVFPSIVGR |
| DSYVG | GYSFVTTAER |
| DLAGR | DLTDYLMKILTER |
| DETTALVC | VAPEEHPTLLTEAPLNPKANR |
| DEAGPSIVHR | KDLYANNVMSGGTTMYPGIADR |
| DIKEKLCYVAL X | HQGVMVGMGQKDSYVGDEAQSKR X |
| DNGSGLVKAGFAG | MQKEITALAPSTMKIKIIAPPER |
| DLYANNVMSGGTTMYPGIA | DEDETTALVCDNGSGLVKAGFAGDDAPR X |
| DGVTHNVPIYEGYALPHAIMRL | TTGIVLDSGDGVTHNVPIYEGYALPHAIMR |
| DFENEMATAASSSSLEKSYELP X | EKMTQIMFETFNVPAMYVAIQAVLSLYASGR |

| | |
|---|---|
| DEAQSKRGILTLKYPIEHGIITNW X | GILTLKYPIEHGIITNWDDMEKIWHHTFYNELR |
| DYLMKILTERGYSFVTTAEREIVR | CPETLFQPSFIGMESAGIHETTYNSIMKCDIDIR X |
| DAPRAVFPSIVGRPRHQGVMVGMGQK X | KYSVWIGGSILASLSTFQQMWITKQEYDEAGPSIVHR |
| DGQVITIGNERFRCPETLFQPSFIGMESAGIHETTYNSIMKC X | DIKEKLCYVALDFENEMATAASSSSLEKSYELPDGQVITIGNER X |
| DRMQKEITALAPSTMKIKIIAPPERKYSVWIGGSILASLSTFQQMWITKQEY | |
| DMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREKMTQIMFETFNVPAMYVAIQAV X | |
| LSLYASGRTTGIVL | |

Blasting this sequence tells you that the protein is **Actin**.

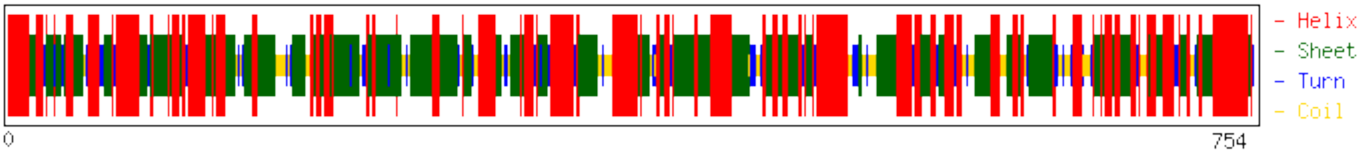
DEDETTALVCDNGSGLVKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQSKRGILTLKYPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPTLLTEAPLNPKANREKMTQIMFETFNVPAMYVAIQAV

Name of the sequence is *7ACN*.

Sequence consists of 754 amino acids.

**Target Sequence:**

```
ERAKVAMSHF EPHEYIRYDL LEKNIDIVRK RLNRPLTLSE KIVYGHLDDP ANQEIERGKT YLRLRPDRVA
MQDATAQMAM LQFISSGLPK VAVPSTIHCD HLIEAQLGGE KDLRRAKDIN QEVYNFLATA GAKYGVGFWR
PGSGIIHQII LENYAYPGVL LIGTDSHTPN GGGLGGICIG VGGADAVDVM AGIPWELKCP KVIGVKLTGS
LSGWTSPKDV ILKVAGILTV KGGTGAIVEY HGPGVDSISC TGMATICNMG AEIGATTSVF PYNHRMKKYL
SKTGRADIAN LADEFKDHLV PDPGCHYDQV IEINLSELKP HINGPFTPDL AHPVAEVGSV AEKEGWPLDI
RVGLIGSCTN SSYEDMGRSA AVAKQALAHG LKCKSQFTIT PGSEQIRATI ERDGYAQVLR DVGGIVLANA
CGPCIGQWDR KDIKKGEKNT IVTSYNRNFT GRNDANPETH AFVTSPEIVT ALAIAGTLKF NPETDFLTGK
DGKKFKLEAP DADELPRAEF DPGQDTYQHP PKDSSGQRVD VSPTSQRLQL LEPFDKWDGK DLEDLQILIK
VKGKCTTDHI SAAGPWLKFR GHLDNISNNL LIGAINIENR KANSVRNAVT QEFGPVPDTA RYYKQHGIRW
VVIGDENYGE GSSREHSALE PRHLGGRAII TKSFARIHET NLKKQGLLPL TFADPADYNK IHPVDKLTIQ
GLKDFAPGKP LKCIIKHPNG TQETILLNHT FNETQIEWFR AGSALNRMKE LQQK
```



**Secondary Structure:**

```
                   *         *         *         *         *
Query 1   ERAKVAMSHFEPHEYIRYDLLEKNIDIVRKRLNRPLTLSEKIVYGHLDDPANQEIERGKT 60
Helix 1   HHHHHHHHHHHHHH  HHHHHHHHHHHHHHH    HHHHHHH       HHHHHHHHH   60
Sheet 1      EEEE       EEEEEE    EEEEEEEEEEEEEEEEEEEEEE             E 60
Turns 1   T          T          T      T   T     T        T   T   TTT  60

                   *         *         *         *         *
Query 61  YLRLRPDRVAMQDATAQMAMLQFISSGLPKVAVPSTIHCDHLIEAQLGGEKDLRRAKDIN 120
Helix 61      HHHHHHHHHHHHHHHHHHHHHHHHHHHH      HHHHHHHHHHHHHHHHHHHHHH 120
Sheet 61  EEE     EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE            EE 120
Turns 61       TT                  T        T         T   T T          120

                   *         *         *         *         *
Query 121 QEVYNFLATAGAKYGVGFWRPGSGIIHQIILENYAYPGVLLIGTDSHTPNGGGLGGICIG 180
Helix 121 HHHHHHHHHHHH            HHHHHHHHHH    HH                     180
Sheet 121 EEEEEEEE     EEEEE      EEEEEEEEEEEEEEEEEEE          EEEEEEE 180
Turns 121 T                   T T                          T T         180

                   *         *         *         *         *
Query 181 VGGADAVDVMAGIPWELKCPKVIGVKLTGSLSGWTSPKDVILKVAGILTVKGGTGAIVEY 240
Helix 181    HHHHHHHHHHHHHHHHHHHHHHH           HHHHHHHHHHHHH     HH    240
Sheet 181 E     EEEEEEEEEEEEEEEEEEEEEEEEEEEE     EEEEEEEEEEEEEEEEEEEE  240
Turns 181                             T      TTT             T         240

                   *         *         *         *         *
Query 241 HGPGVDSISCTGMATICNMGAEIGATTSVFPYNHRMKKYLSKTGRADIANLADEFKDHLV 300
Helix 241               HHHHHHHHHHHHH        HHHHHH    HHHHHHHHHHHHHH  300
Sheet 241     EEEEEEEEEEEEEE    EEEEEEEEEEE  EEEEEEE            EEEEE  300
Turns 241  TT                 T            T       T          T   T    300

                   *         *         *         *         *
Query 301 PDPGCHYDQVIEINLSELKPHINGPFTPDLAHPVAEVGSVAEKEGWPLDIRVGLIGSCTN 360
Helix 301      HHHHHHHHHHHHHHHH       HHHHHHHHHHHHHHHHHHHHHHHHHH       360
Sheet 301     EEEEEEEEEE       EEEEEEEE                EEEEEEEEEEEEE   360
Turns 301  TT                T   T   T              TT                 360

                   *         *         *         *         *
Query 361 SSYEDMGRSAAVAKQALAHGLKCKSQFTITPGSEQIRATIERDGYAQVLRDVGGIVLANA 420
Helix 361       HHHHHHHHHHHHHHHHHHHHHH     HHHHHHHHHHHHHHHHHHHHHHHHHH  420
Sheet 361            EEEEE  EEEEEEEEEEE     EEEEE     EEEEEEEEEEEEEEEE 420
Turns 361 T      T      T         T     TT T       TT                  420

                   *         *         *         *         *
Query 421 CGPCIGQWDRKDIKKGEKNTIVTSYNRNFTGRNDANPETHAFVTSPEIVTALAIAGTLKF 480
Helix 421       HHHHHHHHHHHHHHH                HHHHHHHHHHHHHHHHHHHHHHH 480
Sheet 421 EEEEEEE            EEEEEEEEEE         EEEEEEEEEEEEEEEEEEEEE  480
Turns 421          TT   T  TT       T   TTT   T        T               480

                   *         *         *         *         *
```

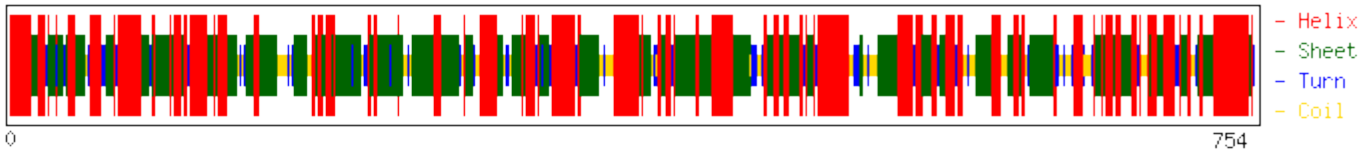
```
Query 481 NPETDFLTGKDGKKFKLEAPDADELPRAEFDPGQDTYQHPPKDSSGQRVDVSPTSQRLQL 540
Helix 481 HHHHH HHHHHHHHHHHHHHHHHHHHHHH                          HHHHH 540
Sheet 481    EEEE                          EEEE         EEEEEEEEEEEEEE 540
Turns 481  T T    TT  T                  T TT     T     T        T     540

                   *         *         *         *         *
Query 541 LEPFDKWDGKDLEDLQILIKVKGKCTTDHISAAGPWLKFRGHLDNISNNLLIGAINIENR 600
Helix 541 HHHHHHHHHHHHHHHHHHHHHHHH   HHHHHHHHHHH           HHHHHHHHHHH 600
Sheet 541 EEEE       EEEEEEEEEEEEEEEEEEE     EEE       EEEEEEEEEE      600
Turns 541         TT            T          T      T      T          TT 600

                   *         *         *         *         *
Query 601 KANSVRNAVTQEFGPVPDTARYYKQHGIRWVVIGDENYGEGSSREHSALEPRHLGGRAII 660
Helix 601 H      HHHHHHHH           HHHHHHHHH         HHHHHH       HHH 660
Sheet 601     EEEEEEEEE     EEEEEEEEEEEEEE                          EE 660
Turns 601       T   T      T      T         TT   T   T      T    T     660

                   *         *         *         *         *
Query 661 TKSFARIHETNLKKQGLLPLTFADPADYNKIHPVDKLTIQGLKDFAPGKPLKCIIKHPNG 720
Helix 661 HHHHHHHHHHHHHHHHHHHHHHHHHH   HHHHHHHHHHHHHHHHHHHHHHHHHHH     720
Sheet 661 EEEEE EEEEEEEEEEEEEEE          EEEEEEEEEE         EEEE       720
Turns 661  T       T   TT        T                      TTT         T  720

                   *         *         *
Query 721 TQETILLNHTFNETQIEWFRAGSALNRMKELQQK 754
Helix 721  HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH  754
Sheet 721 EEEEEEEEEEEEEEEEEE         EEEEE   754
Turns 721   T         T         T          T 754
```

Name of the sequence is *1DNK:A|PDBID|CHAIN|SEQUENCE*.

Sequence consists of 260 amino acids.

**Target Sequence:**

```
LKIAAFNIRT FGETKMSNAT LASYIVRIVR RYDIVLIQEV RDSHLVAVGK LLDYLNQDDP NTYHYVVSEP
LGRNSYKERY LFLFRPNKVS VLDTYQYDDG CESCGNDSFS REPAVVKFSS HSTKVKEFAI VALHSAPSDA
VAEINSLYDV YLDVQQKWHL NDVMLMGDFN ADCSYVTSSQ WSSIRLRTSS TFQWLIPDSA DTTATSTNCA
YDRIVVAGSL LQSSVVPGSA APFDFQAAYG LSNEMALAIS DHYPVEVTLT
```



**Secondary Structure:**

```
                   *         *         *         *         *
Query 1   LKIAAFNIRTFGETKMSNATLASYIVRIVRRYDIVLIQEVRDSHLVAVGKLLDYLNQDDP 60
Helix 1   HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH      60
Sheet 1      EEEEEE          EEEEEEEEEEEEEEEEEEEE       EEEEEEEE       60
Turns 1               T                            T              T  T 60

                   *         *         *         *         *
Query 61  NTYHYVVSEPLGRNSYKERYLFLFRPNKVSVLDTYQYDDGCESCGNDSFSREPAVVKFSS 120
Helix 61       HH         HHHHHHHHHHHHHHHH                  HHHHHHHHHH 120
Sheet 61   EEEEE             EEEE     EEEEEEEE                         120
Turns 61  T       T   TT   T       TT           T      TT   TT       T 120

                   *         *         *         *         *
Query 121 HSTKVKEFAIVALHSAPSDAVAEINSLYDVYLDVQQKWHLNDVMLMGDFNADCSYVTSSQ 180
Helix 121 HHHHHHHHHHHHHHH   HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH          H 180
Sheet 121      EEEEEE               EEEEEEEEEEEEEEEEEEE       EEEEEEE  180
```

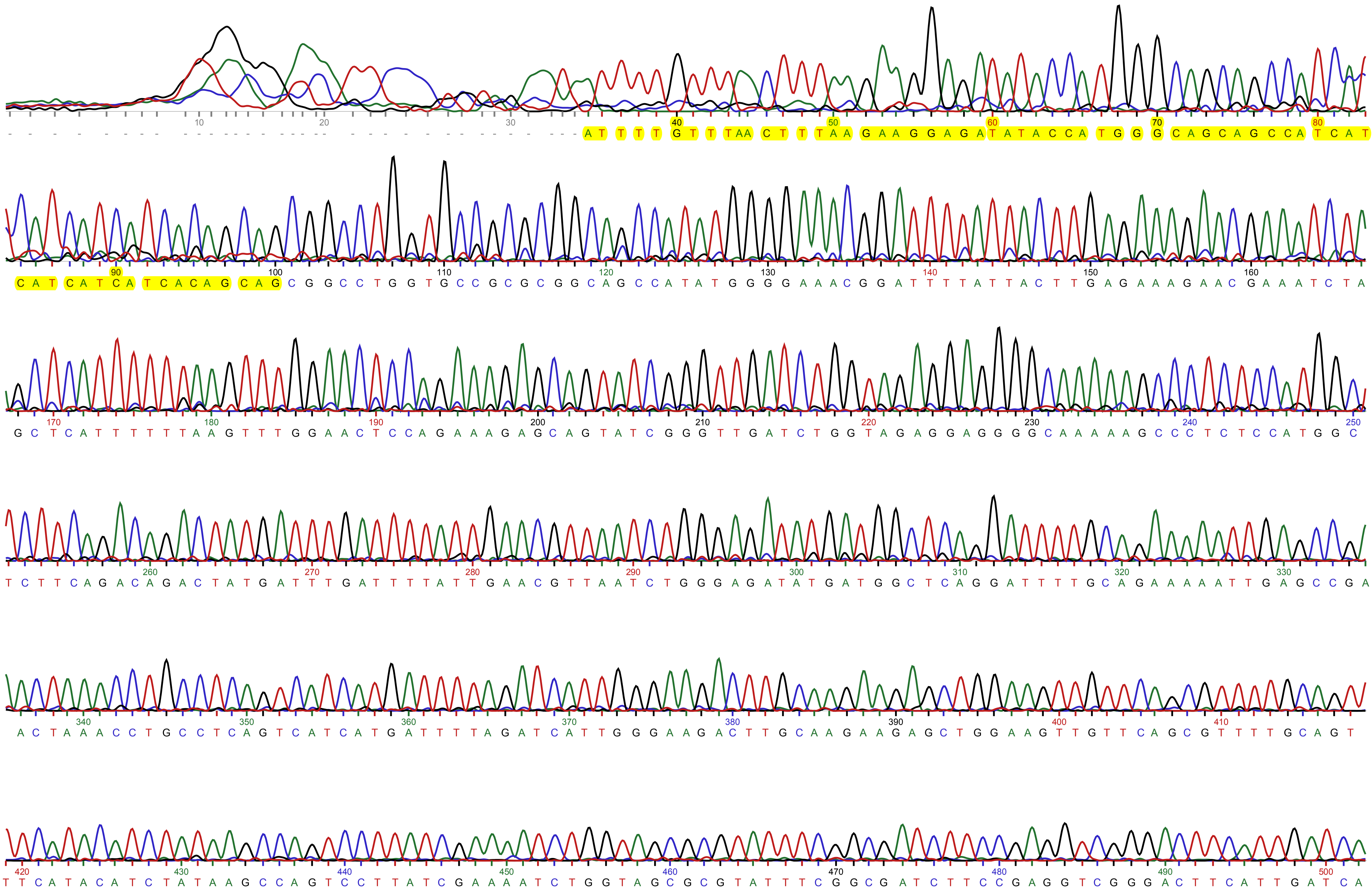
```
Turns 121                  T                 T              T       T  180

                   *         *         *         *         *
Query 181 WSSIRLRTSSTFQWLIPDSADTTATSTNCAYDRIVVAGSLLQSSVVPGSAAPFDFQAAYG 240
Helix 181 HHHHHHHHHHHHHHHHHHHHHHHH       HHHHHHHHHHHHHH    HHHHHHHHHHH 240
Sheet 181     EEEEEEEEEEE      EEEEEEEEEEEEEEEEEEEEEEEE                240
Turns 181          T       T                   T    T    T             240

                   *         *
Query 241 LSNEMALAISDHYPVEVTLT 260
Helix 241 HHHHHHHHHHH   HH     260
Sheet 241             EEEEEE   260
Turns 241   T       T          260
```

Chymotrypsin

| | |
|---|---|
| ERAKVAMSHF | 1175.371 |
| EPHEY | 673.68 |
| IRY | 450.538 |
| DLLEKNIDIVRKRLNRPLTLSEKIVY | 3139.732 |
| GHLDDPANQEIERGKTY | 1943.06 |
| LRLRPDRVAMQDATAQMAMLQF | 2563.049 |
| ISSGLPKVAVPSTIHCDHLIEAQLGGEKDLRRAKDINQEVY | 4502.125 |
| NF | 279.296 |
| LATAGAKY | 793.918 |
| GVGF | 378.428 |
| W | 204.228 |
| RPGSGIIHQIILENY | 1709.966 |
| AYPGVLLIGTDSHTPNGGGLGGICIGVGGADAVDVMAGIPW | 3879.417 |
| ELKCPKVIGVKLTGSLSGW | 2015.442 |
| TSPKDVILKVAGILTVKGGTGAIVEY | 2630.12 |
| HGPGVDSISCTGMATICNMGAEIGATTSVFPY | 3188.607 |
| NHRMKKY | 976.164 |
| LSKTGRADIANLADEF | 1720.9 |
| KDHLVPDPGCHY | 1380.541 |
| DQVIEINLSELKPHINGPF | 2163.458 |
| TPDLAHPVAEVGSVAEKEGWPLDIRVGLIGSCTNSSY | 3869.318 |
| EDMGRSAAVAKQALAHGLKCKSQF | 2546.945 |
| TITPGSEQIRATIERDGY | 2007.188 |
| AQVLRDVGGIVLANACGPCIGQW | 2340.743 |
| DRKDIKKGEKNTIVTSY | 1995.263 |
| NRNF | 549.587 |
| TGRNDANPETHAF | 1429.469 |
| VTSPEIVTALAIAGTLKF | 1831.183 |
| NPETDF | 721.722 |
| LTGKDGKKF | 993.171 |
| KLEAPDADELPRAEF | 1700.866 |
| DPGQDTY | 794.773 |
| QHPPKDSSGQRVDVSPTSQRLQLLEPF | 3047.378 |
| DKW | 447.491 |
| DGKDLEDLQILIKVKGKCTTDHISAAGPW | 3152.613 |
| LKF | 406.525 |
| RGHLDNISNNLLIGAINIENRKANSVRNAVTQEF | 3792.229 |
| GPVPDTARY | 975.069 |
| Y | 181.191 |
| KQHGIRW | 924.073 |
| VVIGDENY | 907.976 |
| GEGSSREHSALEPRHLGGRAIITKSF | 2793.095 |
| ARIHETNLKKQGLLPLTF | 2079.473 |
| ADPADY | 650.643 |
| NKIHPVDKLTIQGLKDF | 1966.311 |
| APGKPLKCIIKHPNGTQETILLNHTF | 2871.393 |
| NETQIEW | 918.958 |
| F | 165.192 |
| RAGSALNRMKELQQK | 1730.018 |

Arg-C

| | |
|---|---|
| ER | 303.318 |
| AKVAMSHFEPHEYIR | 1815.08 |
| YDLLEKNIDIVR | 1490.72 |
| KR | 302.377 |
| LNR | 401.466 |
| PLTLSEKIVYGHLDDPANQEIER | 2637.929 |
| GKTYLR | 736.869 |
| LR | 287.362 |
| PDR | 386.408 |
| VAMQDATAQMAMLQFISSGLPKVAVPSTIHCDHLIEAQLGGEKDLR | 4922.726 |
| R | 174.203 |
| AKDINQEVYNFLATAGAKYGVGFWR | 2819.172 |
| PGSGIIHQIILENYAYPGVLLIGTDSHTPNGGGLGGICIGVGGADAVDVMAGIPWELKC | 13602.736 |
| PKVIGVKLTGSLSGWTSPKDVILKVAGILTVKGGTGAIVEYHGPGVDSISCTGMATICNM | |
| GAEIGATTSVFPYNHR | |
| MKKYLSKTGR | 1211.488 |
| ADIANLADEFKDHLVPDPGCHYDQVIEINLSELKPHINGPFTPDLAHPVAEVGSVAEKE | 7255.101 |
| GWPLDIR | |
| VGLIGSCTNSSYEDMGR | 1788.965 |
| SAAVAKQALAHGLKCKSQFTITPGSEQIR | 3041.519 |
| ATIER | 588.662 |
| DGYAQVLR | 921.021 |
| DVGGIVLANACGPCIGQWDR | 2044.331 |
| KDIKKGEKNTIVTSYNR | 1994.279 |
| NFTGR | 593.64 |
| NDANPETHAFVTSPEIVTALAIAGTLKFNPETDFLTGKDGKKFKLEAPDADELPR | 5942.633 |
| AEFDPGQDTYQHPPKDSSGQR | 2360.438 |
| VDVSPTSQR | 988.065 |
| LQLLEPFDKWDGKDLEDLQILIKVKGKCTTDHISAAGPWLKFR | 4967.804 |
| GHLDNISNNLLIGAINIENR | 2190.444 |
| KANSVR | 673.77 |
| NAVTQEFGPVPDTAR | 1601.736 |
| YYKQHGIR | 1064.212 |
| WVVIGDENYGEGSSR | 1667.752 |
| EHSALEPR | 938.008 |
| HLGGR | 538.607 |
| AIITKSFAR | 1006.213 |
| IHETNLKKQGLLPLTFADPADYNKIHPVDKLTIQGLKDFAPGKPLKCIIKHPNGTQETI | 8490.815 |
| LLNHTFNETQIEWFR | |
| AGSALNR | 687.754 |
| MKELQQK | 904.092 |


10 20 30 40 50 60 70 80
A T T T T G T T TAA C T T T A A G A A G G A G A T A T A C C A T G G G C A G C A G C C A T C A T
90 100 110 120 130 140 150 160
C A T C A T C A T C A C A G C A G C G G C C T G G T G C C G C G C G G C A G C C A T A T G G G G A A A C G G A T T T T A T T A C T T G A G A A A G A A C G A A A T C T A
170 180 190 200 210 220 230 240 250
G C T C A T T T T T T A A G T T T G G A A C T C C A G A A A G A G C A G T A T C G G G T T G A T C T G G T A G A G G A G G G G C A A A A A G C C C T C T C C A T G G C
260 270 280 290 300 310 320 330
T C T T C A G A C A G A C T A T G A T T T G A T T T T A T T G A A C G T T A A T C T G G G A G A T A T G A T G G C T C A G G A T T T T G C A G A A A A A T T G A G C C G A
340 350 360 370 380 390 400 410
A C T A A A C C T G C C T C A G T C A T C A T G A T T T T A G A T C A T T G G G A A G A C T T G C A A G A A G A G C T G G A A G T T G T T C A G C G T T T T G C A G T
420 430 440 450 460 470 480 490 500
T T C A T A C A T C T A T A A G C C A G T C C T T A T C G A A A A T C T G G T A G C G C G T A T T T C G G C G A T C T T C C G A G G T C G G G A C T T C A T T G A T C A