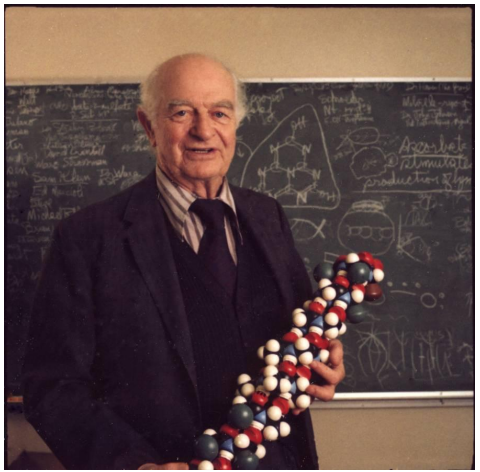
université PARIS DIDEROT PARIS 7

Programming Project : Secondary Structure Assignment

by Hélène Kabbech, bioinformatics master student

1951

α-helices and pleated β-sheets predicted by L. Pauling and R. Corey

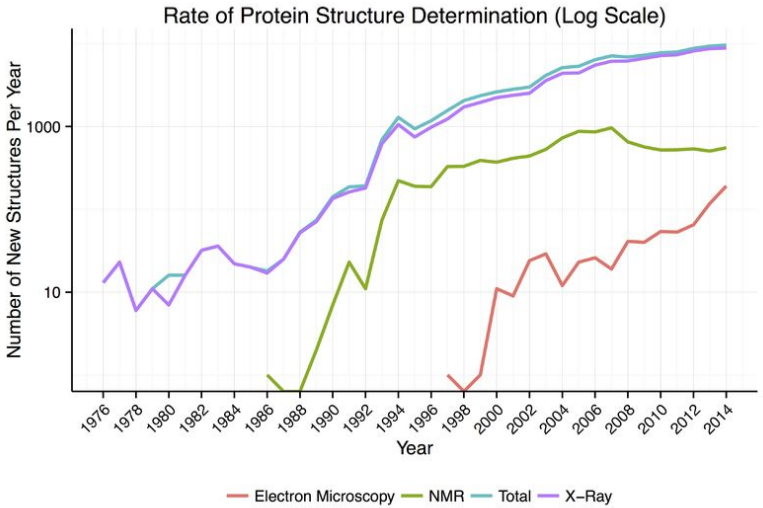


1971

Protein Data Bank (PDB)

1983

DSSP program by W. Kabsch and C. Sander



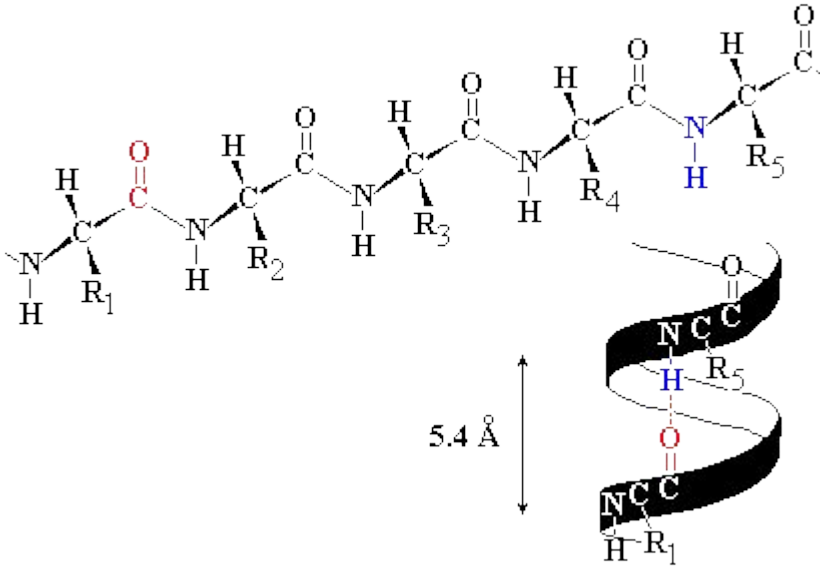
Context

université PARIS DIDEROT PARIS 7

Secondary structure recognition algorithm mainly on H-bonding patterns

H-bond between the CO of residue i and NH of residue j

Cutoff : E [Hbond(res i, res j)] < 0.5 kcal/mol



W. Kabsch and C. Sander *Biopolymers* 22.12 (1983): 2577-2637.

www.chembio.uoguelph.ca

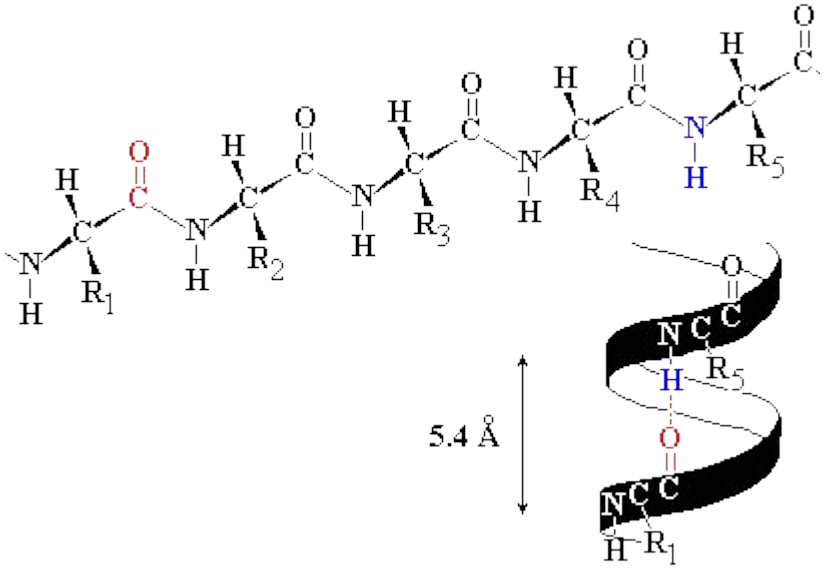
How Secondary Structures are Assigned ?

Secondary structure recognition algorithm mainly on H-bonding patterns

H-bond between the CO of residue i and NH of residue j

Cutoff : E [Hbond(res i, res j)] < 0.5 kcal/mol

n-turn : Residue i and i+n (with n = [3,4,5])
bridge : H-bonds between residues not near each other in sequence



W. Kabsch and C. Sander *Biopolymers* 22.12 (1983): 2577-2637.

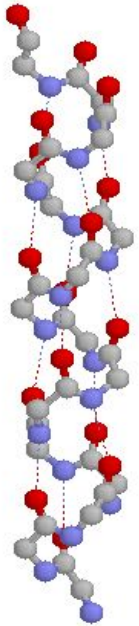
www.chembio.uoguelph.ca

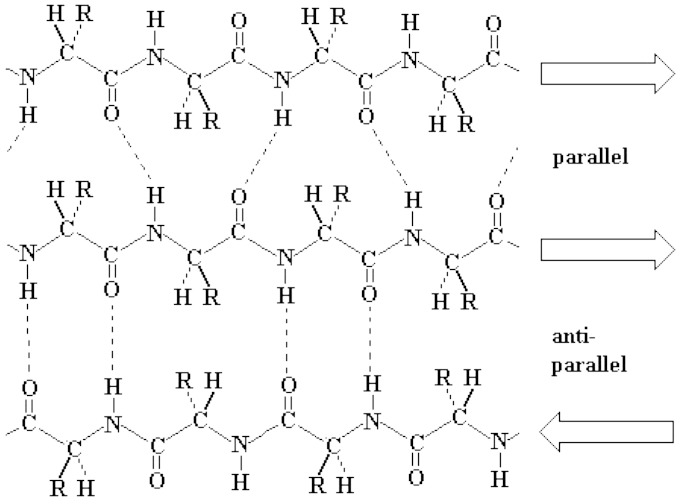
How Secondary Structures are Assigned ?

Repetitive n-turns : n-helix (**G**, **H, I**)
Single n-turn (**T**)

Repetitive β-bridges : β-structures (**E**)
Isolated β-bridge (**B**)

Curvature of at least 70° : Bend (**S**)





W. Kabsch and C. Sander *Biopolymers* 22.12 (1983): 2577-2637.

www.chembio.uoguelph.ca

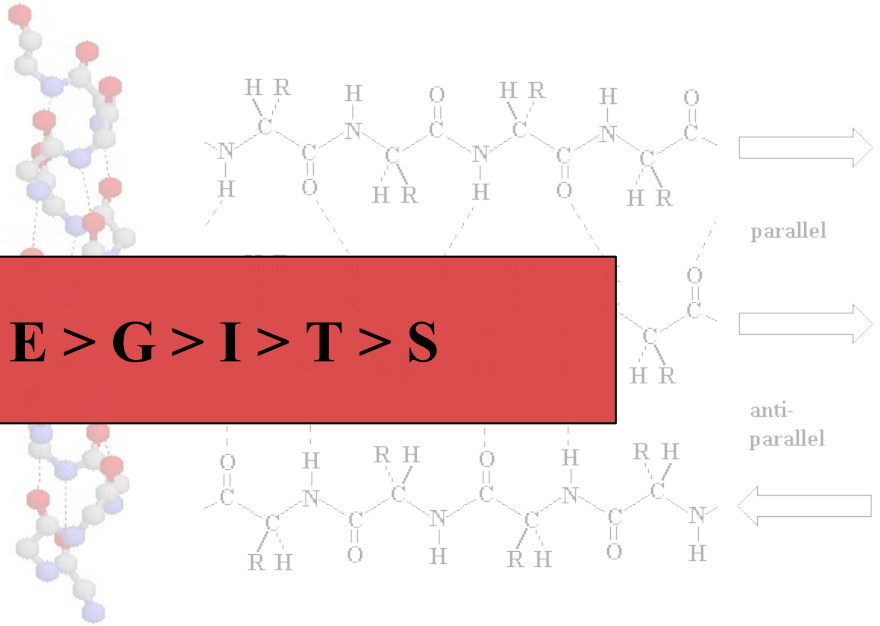
How Secondary Structures are Assigned ?

Repetitive n-turns : n-helix (**G**, **H**, **I**)
Single n-turn (**T**)

Repetitive β-b
Isolated β-Bri

Curvature of a t least 70° : Bend (**S**)

Priority order : H > B > E > G > I > T > S



W. Kabsch and C. Sander *Biopolymers* 22.12 (1983): 2577-2637.

www.chembio.uoguelph.ca

How Secondary Structures are Assigned ?

université PARIS DIDEROT PARIS 7

Implementation the Secondary Structure Recognition Algorithm described by Kabsch and Sander using the Python programming language

Python 3.6 and Biopython library 1.72

Version control Git/GitHub
https://github.com/kabhel/DSSP_implt

Reduce 3.23 software in order to add H atoms into PDB file

Programming Tools used

université PARIS DIDEROT PARIS 7

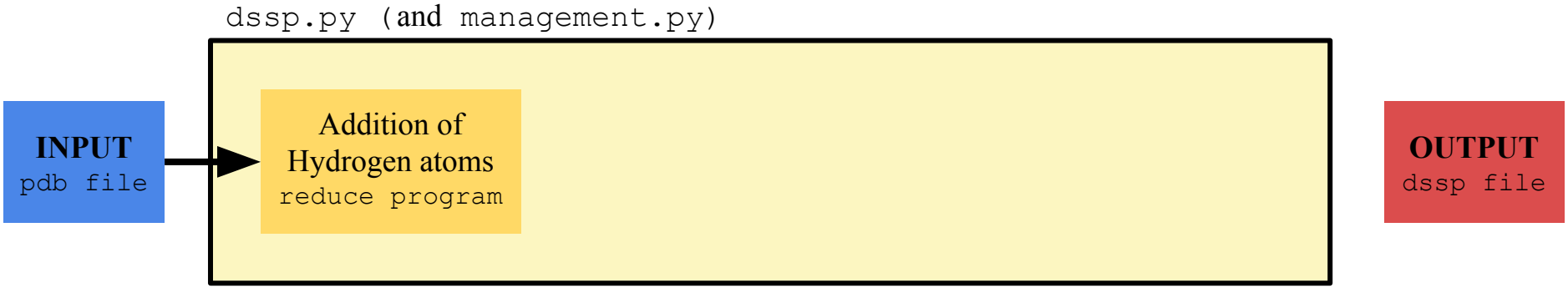
INPUT
pdb file

OUTPUT
dssp file

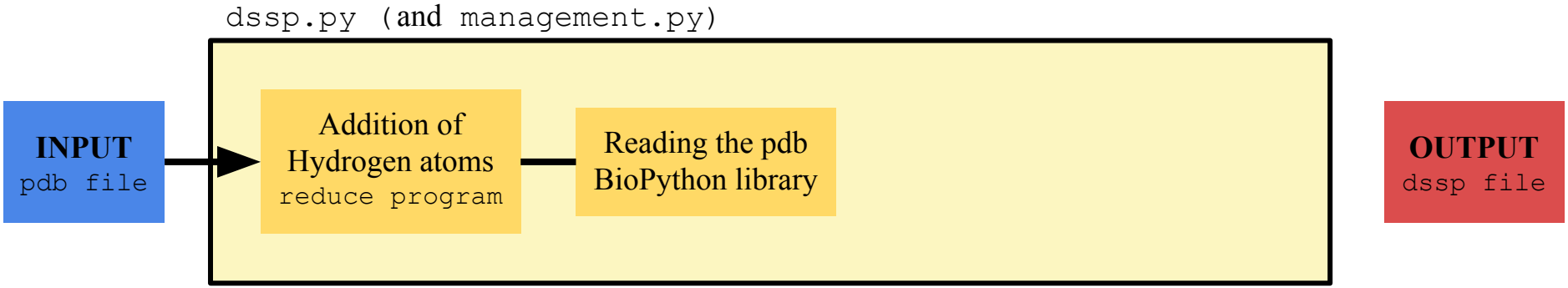
```
ATOM      1  N   LYS A   1      -8.655   5.770   8.371  1.00  1.40           N
ATOM      2  CA  LYS A   1      -7.542   5.187   9.163  1.00  0.52           C
ATOM      3  C   LYS A   1      -6.210   5.619   8.561  1.00  0.39           C
ATOM      4  O   LYS A   1      -6.156   6.468   7.693  1.00  0.33           O
ATOM      5  CB  LYS A   1      -7.641   3.666   9.159  1.00  1.53           C
ATOM      6  CG  LYS A   1      -7.750   3.179  10.602  1.00  2.38           C
ATOM      7  CD  LYS A   1      -9.158   3.466  11.126  1.00  3.11           C
ATOM      8  CE  LYS A   1      -9.145   4.744  11.965  1.00  3.58           C
ATOM      9  NZ  LYS A   1      -9.497   4.416  13.375  1.00  4.21           N
ATOM     10  H1  LYS A   1      -8.409   6.745   8.105  1.00  1.94           H
ATOM     11  H2  LYS A   1      -8.809   5.202   7.514  1.00  2.04           H
ATOM     12  H3  LYS A   1      -9.523   5.775   8.942  1.00  1.89           H
ATOM     13  HA  LYS A   1      -7.604   5.545  10.180  1.00  1.19           H
ATOM     14  HB2 LYS A   1      -8.516   3.361   8.603  1.00  1.84           H
ATOM     15  HB3 LYS A   1      -6.756   3.246   8.705  1.00  2.02           H
ATOM     16  HG2 LYS A   1      -7.558   2.117  10.640  1.00  2.57           H
ATOM     17  HG3 LYS A   1      -7.025   3.703  11.212  1.00  2.64           H
ATOM     18  HD2 LYS A   1      -9.833   3.592  10.290  1.00  3.38           H
ATOM     19  HD3 LYS A   1      -9.491   2.640  11.735  1.00  3.41           H
ATOM     20  HE2 LYS A   1      -8.160   5.186  11.934  1.00  3.84           H
ATOM     21  HE3 LYS A   1      -9.866   5.442  11.568  1.00  3.71           H
ATOM     22  HZ1 LYS A   1      -9.394   3.393  13.530  1.00  4.40           H
ATOM     23  HZ2 LYS A   1      -8.864   4.932  14.019  1.00  4.64           H
ATOM     24  HZ3 LYS A   1     -10.481   4.694  13.559  1.00  4.45           H
ATOM     25  N   LYS A   2      -5.128   5.066   9.028  1.00  0.51           N
ATOM     26  CA  LYS A   2      -3.813   5.485   8.490  1.00  0.44           C
ATOM     27  C   LYS A   2      -2.931   4.277   8.172  1.00  0.41           C
ATOM     28  O   LYS A   2      -3.040   3.218   8.757  1.00  0.58           O
ATOM     29  CB  LYS A   2      -3.138   6.383   9.534  1.00  0.63           C
```

```
#  RESIDUE AA STRUCTURE BP1 BP2    TCO  KAPPA ALPHA  PHI   PSI    X-CA   Y-CA   Z-CA
  1    1 A K              0   0  0.000 360.0 360.0 360.0-171.7   -7.5    5.2    9.2
  2    2 A K  E     -a    50   0 -0.957 360.0-169.5-133.8 153.7   -3.8    5.5    8.5
  3    3 A A  E     -a    51   0 -0.933   5.4-159.9-134.7 157.3   -1.1    3.4    6.8
  4    4 A V  E     -a    52   0 -0.926   5.9-160.1-144.5 118.6    2.7    3.8    6.8
  5    5 A I  E     -a    53   0 -0.740  11.7-170.1-100.1  93.0    5.1    2.2    4.2
  6    6 A N  E >>  -a    54   0 -0.684   4.7-172.9 -83.5 104.2    8.5    2.2    5.9
  7    7 A G  B 34 S+B     9   0  0.931  79.7  70.5 -64.8 -43.3   11.0    1.3    3.1
  8    8 A E  T 34 S+      0   0  0.812 111.2  36.3 -44.5 -25.9   13.9    1.0    5.5
  9    9 A Q  B <4 S+B     7   0  0.868  92.5  99.6 -94.9 -47.2   12.1   -2.1    6.7
 10   10 A I    <  +      0   0 -0.075  41.7 176.6 -41.3 133.7   10.6   -3.4    3.4
 11   11 A R              0   0  0.480  64.2 -14.5-119.2 -11.8   12.8   -6.3    2.2
 12   12 A S  S  > S-     0   0 -0.982  82.9 -77.3-174.2 175.9   10.7   -7.3   -0.9
 13   13 A I  H  > S+     0   0  0.951 127.9  57.2 -56.6 -45.7    7.3   -6.8   -2.5
 14   14 A S  H  > S+     0   0  0.940 105.2  51.7 -50.8 -45.4    5.9   -9.3   -0.0
 15   15 A D  H  > S+     0   0  0.933 106.8  54.1 -57.9 -43.3    7.1   -7.0    2.7
 16   16 A L  H  X S+     0   0  0.978 112.3  41.4 -55.2 -56.5    5.3   -4.1    1.0
 17   17 A H  H  X S+     0   0  0.740 110.3  64.2 -64.5 -18.6    2.0   -5.9    0.9
 18   18 A Q  H  X S+     0   0  0.979 107.7  36.0 -70.2 -55.2    2.9   -7.0    4.5
 19   19 A T  H  X S+     0   0  0.935 117.5  53.5 -65.2 -41.3    2.9   -3.6    6.0
 20   20 A L  H  X S+     0   0  0.903 108.1  52.9 -59.0 -35.8   -0.0   -2.6    3.8
 21   21 A K  H  X>S+     0   0  0.915 111.4  44.2 -66.9 -42.1   -1.8   -5.7    5.1
 22   22 A K  H  <5S+     0   0  0.993 120.2  39.8 -67.3 -57.4   -1.2   -4.6    8.7
 23   23 A E  H  <5S+     0   0  0.986 124.9  37.7 -52.3 -66.1   -2.2   -1.0    8.2
 24   24 A L  H  <5S-     0   0  0.680 104.4-138.4 -61.2 -14.8   -5.1   -1.9    5.9
 25   25 A A  T  <5 -     0   0  0.947  32.7-172.8  55.0  47.6   -5.7   -4.9    8.2
```

Code Structure

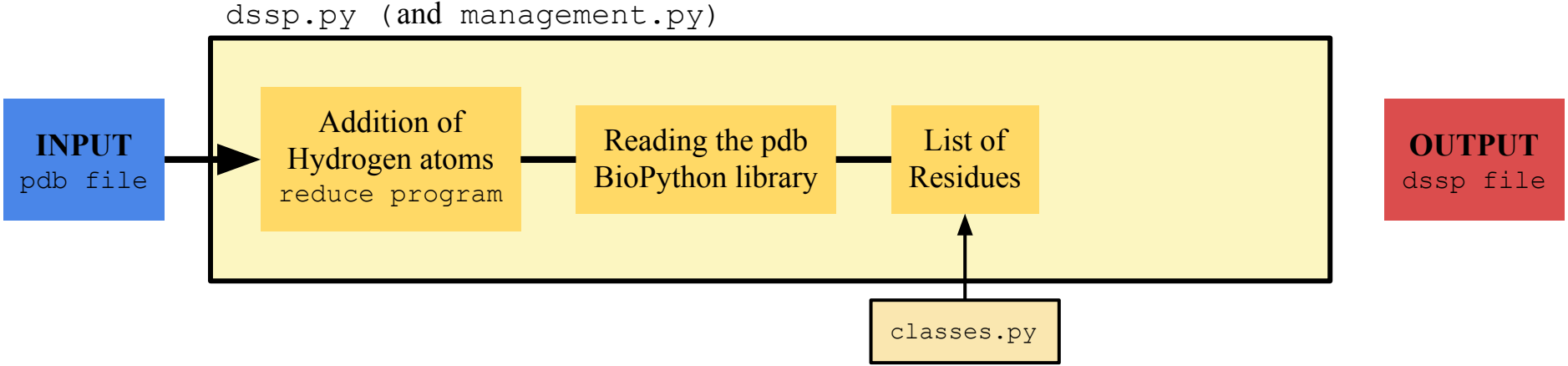


Code Structure

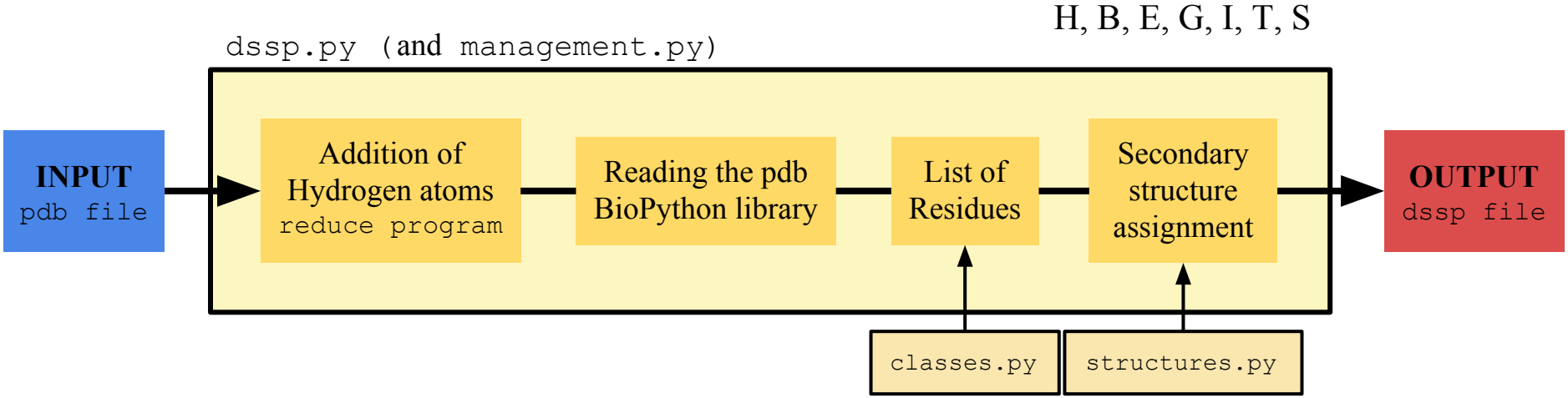


Code Structure

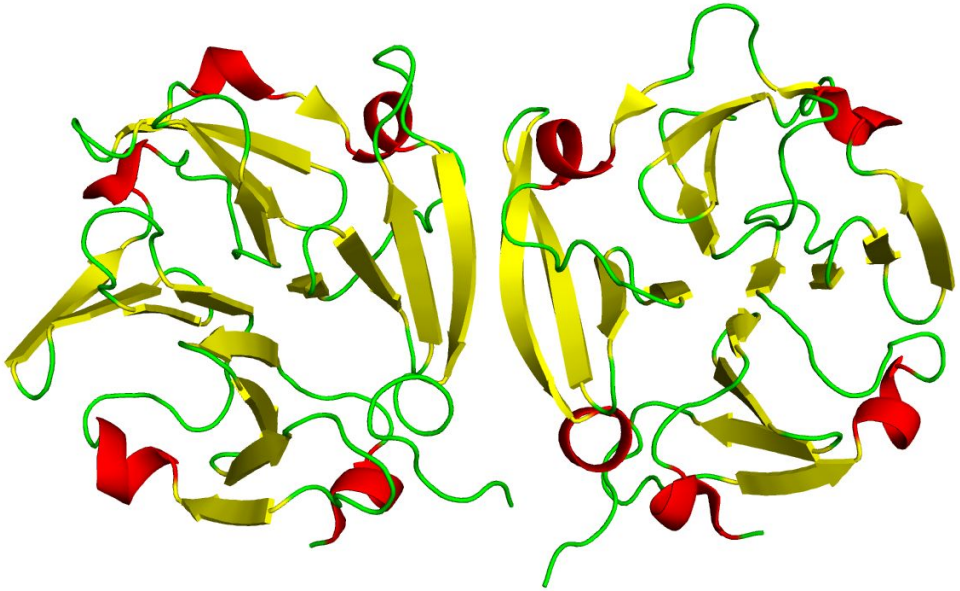
Cosine of angle **tco**
Bend angle **kappa** (→bend)
Dihedral angle **α** (→chirality)
Dihedral angles **Φ, Ψ**



Code Structure



Code Structure



Dimeric form of the haemoplexin domain of MMP9

Cha Hyunju et al. *Journal of molecular biology* 320.5 (2002): 1065-1079.

Output Example : 1ITV

université PARIS DIDEROT PARIS 7

Main SSE in chain A :

G : 3-5

E : 11-15, 18-23, 26-30, 42-43

H : 44-47

E : 57-60, 67-71, 74-79, 82-84, 87-88

G : 90-92

E : 111-116, 119-124, 129-130

G : 132-134

H : 138-141

E : 151-155, 158-163, 166-171, 178-184

```
==== Secondary Structure Assignment using DSSP method ====
DATE    2018-09-19
REFERENCE W. KABSCH AND C.SANDER, BIOPOLYMERS 22 (1983) 2577-2637
HEADER    HYDROLASE                               2002-02-11
COMPND    MOL_ID: 1; MOLECULE: MMP9; CHAIN: A, B; FRAGMENT: HAEMOPEXIN-LIKE DOMAIN; SYNONYM: 92
KDA TYPE IV COLLAGENASE; EC_NUMBER: 3.4.24.35; EC: 3.4.24.35; ENGINEERED: YES;
SOURCE    MOL_ID: 1; ORGANISM_SCIENTIFIC: HOMO SAPIENS; ORGANISM_COMMON: HUMAN; ORGANISM_TAXID:
9606; EXPRESSION_SYSTEM: ESCHERICHIA COLI; EXPRESSION_SYSTEM_TAXID: 562;
AUTHOR    H.CHA,E.KOPETZKI,R.HUBER,M.LANZENDOERFER,H.BRANDSTETTER
  #  RESIDUE AA STRUCTURE BP1 BP2    TCO  KAPPA ALPHA  PHI   PSI    X-CA   Y-CA   Z-CA
    1    1 A D              0   0    0.000 360.0 360.0 360.0 -77.0  -51.7  -49.4  -11.2
    2    2 A D    >   -     0   0   -0.840 360.0-143.6 -97.2 124.7  -48.4  -50.0   -9.2
    3    3 A A  G >  S+     0   0    0.793  99.9  65.5 -55.5 -27.9  -46.3  -46.9   -8.5
    4    4 A C  G 3  S+     0   0    0.655  97.0  53.9 -70.2 -16.4  -43.3  -49.1   -9.0
    5    5 A N  G <  S+     0   0    0.514  85.0 103.3 -95.3  -5.9  -44.1  -49.7  -12.7
    6    6 A V    <  -     0   0   -0.621  48.0-172.5 -79.5 133.8  -44.3  -45.9  -13.5
    7    7 A N  S    S+     0   0    0.867  75.5  23.3 -91.5 -45.6  -41.3  -44.5  -15.3
    8    8 A I  S    S-     0   0   -0.924  71.4-154.8-129.6 107.8  -42.1  -40.8  -15.3
    9    9 A F       -      0   0   -0.575  19.7-132.5 -77.3 140.8  -44.4  -39.4  -12.6
   10   10 A D  S    S-     0   0    0.916  84.2 -10.1 -60.1 -41.2  -46.1  -36.2  -13.7
   11   11 A A  E     -A   22   0   -0.995  57.8-160.1-157.7 155.5  -45.2  -34.6  -10.4
   12   12 A I  E     +A   21   0   -0.999  20.0 157.9-138.9 138.5  -43.8  -35.5   -6.9
   13   13 A A  E     -A   20   0   -0.980  37.9-120.5-160.4 147.2  -44.0  -33.5   -3.7
   14   14 A E  E     -A   19   0   -0.810  28.5-175.5 -94.7 126.7  -43.8  -34.0    0.0
   15   15 A I  E >   -A   18   0   -0.987  61.1 -39.9-121.6 119.2  -47.0  -33.1    1.9
   16   16 A G  T 3  S-     0   0   -0.489 125.3 -23.6  65.9-125.0  -46.8  -33.3    5.7
   17   17 A N  T 3  S+     0   0   -0.067 125.7  75.4-110.3  32.2  -44.9  -36.4    6.6
   18   18 A Q  E <  S-A   15   0   -0.966  73.3-125.2-141.7 157.7  -45.5  -38.2    3.3
   19   19 A L  E     -AB  14  30   -0.902  25.0-152.9-106.1 125.3  -44.3  -38.1   -0.3
   20   20 A Y  E     -AB  13  29   -0.840   6.2-163.8-100.8 136.9  -47.1  -37.8   -2.9
   21   21 A L  E     -AB  12  28   -0.974   6.4-161.5-120.0 130.8  -46.7  -39.1   -6.4
   22   22 A F  E     +AB  11  27   -0.919  18.0 156.7-118.7 141.8  -48.9  -38.0   -9.3
   23   23 A K  E >   - B   0  26   -0.840  67.1 -21.9-163.5 120.6  -49.6  -39.6  -12.7
   24   24 A D  T 3  S-     0   0    0.910 127.6 -41.3  44.0  63.5  -52.5  -39.4  -15.1
   25   25 A G  T 3  S+     0   0    0.717 123.8 102.8  62.3  20.4  -55.2  -38.3  -12.7
   26   26 A K  E <   -BC  23  43   -0.922  48.1-172.2-131.0 156.8  -53.8  -40.7  -10.1
   27   27 A Y  E     -BC  22  42   -0.993   7.5-155.9-150.1 156.2  -51.7  -40.3   -7.0
   28   28 A W  E     -B   21   0   -0.845  11.1-139.3-127.3 164.8  -49.8  -42.3   -4.4
   29   29 A R  B     -BC  20  39   -0.995  17.7-169.5-129.2 125.8  -48.6  -41.8   -0.8
   30   30 A F  E     -B   19   0   -0.963   7.3-176.2-122.2 132.7  -45.3  -43.0    0.4
   31   31 A S        -      0   0   -0.993  30.4-117.9-126.6 124.1  -44.0  -43.1    4.0
   32   32 A E        +      0   0    0.041  57.7 108.9 -52.1 166.1  -40.4  -44.3    4.7
   33   33 A G  S    S-     0   0   -0.777  76.3 -61.0 144.7 171.8  -39.8  -47.4    6.9
   34   34 A R  S    S-     0   0   -0.668 124.4  -4.0 -90.3  86.3  -38.7  -51.0    6.7
   35   35 A G  S    S+     0   0    0.375  89.0 145.7 114.4  -1.4  -41.2  -52.5    4.4
```

Output Example : 1ITV

université PARIS DIDEROT PARIS 7

Successful assignment of all the secondary structures described by the dictionary protein of secondary structures.

Some differences between the output of dssp 3.0.0.

université PARIS DIDEROT PARIS 7

Thank you for your attention !