Report the length of bothM1 and M2 models (1 point).

```
ATPase model len: 81
GTP model len: 115
```

Report the dimensions of your T,mE and iE for both models (1 point)

```
ATPase T, mE, iE shapes: (9, 82),(20, 82),(20, 82)
GTP T, mE, iE shapes: (9, 116),(20, 116),(20, 116)
```

For position 50 in each model, plot the transition and emission frequencies. (3 points)

I printed position. I am not sure because 49 is more interesting so I attached also this one. There is difference in lists in python: starts from zero, it corresponds to your example. If you wanted position number 50 in R notation it is described as position 49.

```
----------Position 50 of ATPases HMM model.------------
Match emission: A    0.051450
R    0.190833
N    0.080449
D    0.039289
C    0.000935
E    0.191768
Q    0.088868
G    0.018709
H    0.028064
I    0.010290
L    0.049579
K    0.173994
M    0.004677
F    0.000935
P    0.003742
S    0.021515
T    0.021515
W    0.000935
Y    0.001871
V    0.020580
Name: 50, dtype: float64
Insert emission: A    0.05
R    0.05
N    0.05
D    0.05
C    0.05
E    0.05
Q    0.05
G    0.05
H    0.05
I    0.05
L    0.05
K    0.05
M    0.05
F    0.05
P    0.05
S    0.05
T    0.05
W    0.05
Y    0.05
M-M: 0.9980988593155894
M-D: 0.0009505703422053232
M-I: 0.0009505703422053232
D-M: 0.07142857142857142
D-D: 0.8571428571428571
D-I: 0.07142857142857142
I-M: 0.3333333333333333
I-D: 0.3333333333333333
I-I: 0.3333333333333333
```

```
----------Position 50 of GTP binding proteins HMM model.------------
Match emission: A    0.174419
R    0.073643
N    0.023256
D    0.073643
C    0.007752
E    0.158915
Q    0.085271
G    0.050388
H    0.034884
I    0.011628
L    0.027132
K    0.085271
M    0.003876
F    0.015504
P    0.023256
S    0.077519
T    0.019380
W    0.003876
Y    0.003876
V    0.046512
Name: 50, dtype: float64
Insert emission: A    0.05
R    0.05
N    0.05
D    0.05
C    0.05
E    0.05
Q    0.05
G    0.05
H    0.05
I    0.05
L    0.05
K    0.05
M    0.05
F    0.05
P    0.05
S    0.05
T    0.05
W    0.05
Y    0.05
M-M: 0.991701244813278
M-D: 0.004149377593360996
M-I: 0.004149377593360996
D-M: 0.8297872340425532
D-D: 0.14893617021276595
D-I: 0.02127659574468085
I-M: 0.3333333333333333
I-D: 0.3333333333333333
I-I: 0.3333333333333333
```

```
----------Position 49 of ATPases HMM model.------------
Match emission: A     0.162617
R    0.000935
N    0.000935
D    0.000935
C    0.158879
E    0.000935
Q    0.000935
G    0.000935
H    0.000935
I    0.007477
L    0.022430
K    0.000935
M    0.009346
F    0.566355
P    0.000935
S    0.022430
T    0.003738
W    0.000935
Y    0.019626
V    0.017757
Name: 49, dtype: float64
Insert emission: A     0.013158
R    0.078947
N    0.131579
D    0.013158
C    0.013158
E    0.013158
Q    0.315789
G    0.052632
H    0.013158
I    0.013158
L    0.026316
K    0.026316
M    0.039474
F    0.013158
P    0.013158
S    0.039474
T    0.131579
W    0.013158
Y    0.013158
V    0.026316
```

```
----------Position 49 of GTP binding proteins HMM model.------------
Match emission: A     0.056291
R    0.006623
N    0.102649
D    0.566225
C    0.006623
E    0.056291
Q    0.009934
G    0.013245
H    0.052980
I    0.006623
L    0.006623
K    0.013245
M    0.003311
F    0.006623
P    0.003311
S    0.052980
T    0.016556
W    0.006623
Y    0.006623
V    0.006623
Name: 49, dtype: float64
Insert emission: A     0.034483
R    0.006897
N    0.041379
D    0.020690
C    0.020690
E    0.020690
Q    0.027586
G    0.048276
H    0.006897
I    0.020690
L    0.048276
K    0.027586
M    0.006897
F    0.006897
P    0.510345
S    0.075862
T    0.048276
W    0.006897
Y    0.006897
V    0.013793
```

```
M-M: 0.9705603038936372
M-D: 0.001899335232668566
M-I: 0.027540360873694207
D-M: 0.076923076923076923
D-D: 0.8461538461538461
D-I: 0.076923076923076923
I-M: 0.4915254237288136
I-D: 0.01694915254237288
I-I: 0.4915254237288136
```

```
M-M: 0.5263157894736842
M-D: 0.15789473684210525
M-I: 0.3157894736842105
D-M: 0.3333333333333333
D-D: 0.3333333333333333
D-I: 0.3333333333333333
I-M: 0.703125
I-D: 0.0078125
I-I: 0.2890625
```

Plot the values $q(x_i)$ and include this in your report. Which proteins in the list belong to which family? Can you clearly decide for each protein? (5 points)

Family membership
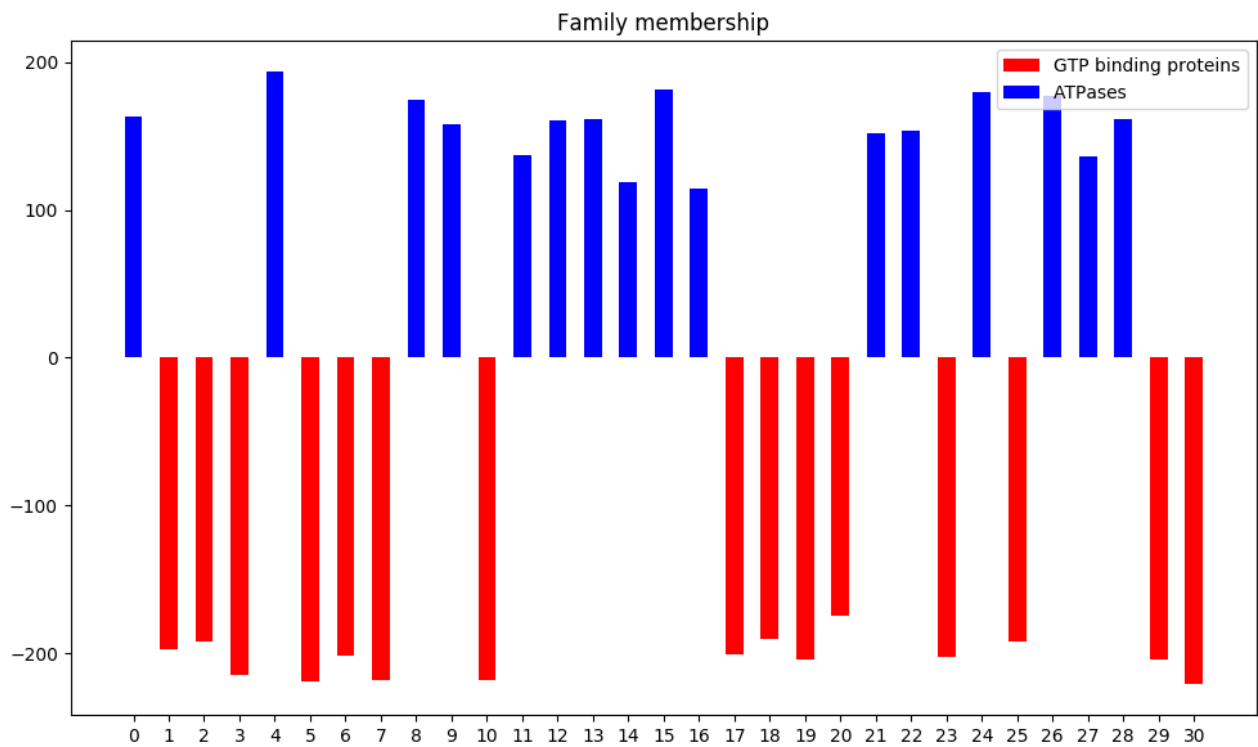
I can really clearly decide to which class given proteins belong.