Analysis of Protein Folding Pathways:
Reame 1

The results from the evaluation (step 2, file "pathways.values") and principal component analysis (step 3, file "pathway.scores") are ready. You need only the last one "pathways.scores" to run your clustering analysis (step 4 and 5). And, the data for the step 6 will available on this Wednesday afternoon.

Summary

The general objective was to establish the relation between conformational and biophysical changes of a protein during folding, and we defined the following specific objectives:

1. Run a cluster analysis on the protein conformations of various PRM simulated pathways using structural information.
2. Evaluate a set of biophysical properties on the protein conformations of simulated pathways.
3. Reduce the set of properties to a more workable set of variables or components.
4. Run a cluster analysis on the protein conformations using the new components.
5. Establish the relation of the resulting cluster between the two cluster analysis.
6. Reduce the dataset of one MD simulated pathway
7. Establish the relation between conformational changes and biophysical properties in the above MD pathway

The following 11 biophysical properties were evaluated on protein structures from multiple pathways:

- native contacts (NC),
- contact order (CO),
- radius of gyration (RG),
- hydrogen bonds (HB),
- accessible surface area (AS),
- root- mean square deviation (RM),
- local root-mean square deviation (LR),
- residues in correct secondary structure elements (SSEs) (RC),
- residues in any SSEs (RA),
- structural score (SS),
- and degrees of freedom (DF).

And the results were saved in the file "**pathways.values**", (see listing below) where the rows correspond to the conformations and the columns to the following properties:

```
"NC" "CO" "RG" "HB" "AS" "RM" "LR" "RC" "RA" "DF" "SS"
"R1008_ARKADIA_120_0_000" 0.009 0.088 0.862 0.069 0.942 0.934 1 0.075 0.235 0.964 0
"R1008_ARKADIA_120_0_001" 0 0.074 0.875 0.069 0.959 0.947 0.995 0.06 0.176 0.955 0
"R1008_ARKADIA_120_0_002" 0.009 0.082 0.883 0 0.953 0.957 0.988 0.03 0.118 0.991 0
"R1008_ARKADIA_120_0_003" 0.009 0.094 0.888 0.034 0.946 0.963 0.98 0.03 0.118 0.973 0
"R1008_ARKADIA_120_0_004" 0.009 0.12 0.894 0.069 0.909 0.967 0.97 0.03 0.118 0.964 0
"R1008_ARKADIA_120_0_005" 0.009 0.172 0.901 0.069 0.83 0.97 0.959 0.03 0.147 1 0
"R1008_ARKADIA_120_0_006" 0.009 0.249 0.911 0.172 0.765 0.972 0.948 0.015 0.118 0.919 0
"R1008_ARKADIA_120_0_007" 0.018 0.319 0.924 0.138 0.723 0.975 0.937 0.015 0.118 0.982 0.009
"R1008_ARKADIA_120_0_008" 0.018 0.363 0.939 0.103 0.693 0.979 0.926 0.03 0.059 0.739 0.009
"R1008_ARKADIA_120_0_009" 0.018 0.37 0.957 0.103 0.685 0.984 0.915 0.03 0.088 0.649 0.009
"R1008_ARKADIA_120_0_010" 0.028 0.361 0.974 0.103 0.689 0.991 0.905 0.045 0.147 0.604 0.009
"R1008_ARKADIA_120_0_011" 0.028 0.336 0.989 0.069 0.706 0.997 0.895 0.015 0.059 0.613 0.017
"R1008_ARKADIA_120_0_012" 0.037 0.3 0.999 0.103 0.718 1 0.885 0 0 0.694 0.026
...
```

Next, a Principal Component Analysis was run on the evaluated protein conformations and it summarized the 11 properties into 3 components: RC1, RC2, and RC3, as it is shown below:

```
Loadings:
    RC1     RC2     RC3
NC   0.706 −0.507 −0.400
CO   0.246 −0.919 −0.194
RG −0.225   0.615   0.678
HB   0.745 −0.335 −0.425
AS −0.408   0.810   0.374
RM −0.264   0.648   0.669
LR −0.436   0.261   0.827
RC   0.766 −0.211 −0.550
RA   0.897 −0.236 −0.191
DF −0.633   0.609   0.363
SS   0.472 −0.302 −0.798

                       RC1   RC2   RC3
SS loadings          3.607 3.299 3.218
Proportion Var       0.328 0.300 0.293
Cumulative Var       0.328 0.628 0.920
```

And, using these loadings we calculated the scores for each protein conformation and resulting values were saved in the file "**pathways.scores**". The first items of this files is shown in the next listing:

```
"RC1" "RC2" "RC3"
"R1008–ARKADIA_120−0−000" −0.770822848910553 0.19864113373733 2.64369649854373
"R1008–ARKADIA_120−0−001" −0.874106341011358 0.230549082535894 2.57828611073212
"R1008–ARKADIA_120−0−002" −1.11658751473656 0.175617778446126 2.4993306140524
"R1008–ARKADIA_120−0−003" −1.05764317146723 0.156452229857772 2.53324238998454
"R1008–ARKADIA_120−0−004" −1.01723868787549 0.0775684956092013 2.5852351568191
"R1008–ARKADIA_120−0−005" −1.00877213288083 −0.0775077164188179 2.67981350080489
"R1008–ARKADIA_120−0−006" −0.926362623292258 −0.346220861249225 2.84270479976395
"R1008–ARKADIA_120−0−007" −1.02962889970688 −0.484770547028307 2.87013518125264
"R1008–ARKADIA_120−0−008" −1.00400104659676 −0.787455878015043 3.02897595644082
"R1008–ARKADIA_120−0−009" −0.863213185300443 −0.843044446606555 3.14829004393121
"R1008–ARKADIA_120−0−010" −0.663351586915068 −0.796178222145434 3.24588925579011
"R1008–ARKADIA_120−0−011" −0.894921688794369 −0.759796079485674 3.09245832899375
"R1008–ARKADIA_120−0−012" −1.01524216326376 −0.613705394694795 2.91603796989101
"R1008–ARKADIA_120−0−013" −0.596128740283448 −0.50633922716996 3.04125269083415
"R1008–ARKADIA_120−0−014" −0.496826621869845 −0.379957255679797 2.96988490428928
"R1008–ARKADIA_120−0−015" −0.64575098001617 −0.415607780041284 2.86981764140085
"R1008–ARKADIA_120−0−016" −0.628410491531592 −0.599578282227478 2.97714945143919
...
```