Mathematica Code for the Singular Value Decomposition (SVD) of the Genetic Code for the 20 Kyte-Doolittle Amino Acid Residue Hydropathy Indices to 12 Nucleotide-Determined Values Without Codon Translation - Replicating Our 1990 Publication

Douglas C. Youvan

doug@youvan.com

January 29, 2023

(* Reaction Centers of Photosynthetic Bacteria (1990) Michel-Beyerle M.ed.pp.209-218,Springer-Verlag Berlin. Genetic Coding Algorithms for Engineering Membrane Proteins. Yang M.M.,Coleman,W.J.,& Youvan,D.C.Massachusetts Institute of Technology,Department of Chemistry,Cambridge 02139. *)

(* Genetic code *)

(* 64 {a,c,g,t} codons in binary in a,c,g,t order as defined by the following *)

```
t=.; g=.;c=.;a=.;
nucgc=Tuples[{a,c,g,t},3];
a={1,0,0,0};
c={0,1,0,0};
g={0,0,1,0};
t={0,0,0,1};

nucgc=Flatten[nucgc];
nucgc=Partition[nucgc,12];


pivns = PseudoInverse[nucgc];  (*corrected*)

(*

codon triplets entered manually in canonical order by alphabet - as above in matrix
```

AAA
AAC
AAG
AAT
ACA
ACC
ACG
ACT
AGA
AGC
AGG
AGT
ATA
ATC
ATG
ATT
CAA
CAC
CAG
CAT
CCA
CCC
CCG
CCT
CGA
CGC
CGG
CGT
CTA
CTC
CTG
CTT
GAA
GAC
GAG
GAT
GCA
GCC
GCG
GCT
GGA
GGC
GGG
GGT
GTA

GTC
GTG
GTT
TAA
TAC
TAG
TAT
TCA
TCC
TCG
TCT
TGA
TGC
TGG
TGT
TTA
TTC
TTG
TTT

manually copy the triplets from above and replace with amino acid single letter code, in identical order, and make matrix 'aagc' for amino acid genetic code, including 3 stop codons

*)

(* Map an attribute of the amino acids onto the triplet codons *)

jjt=.

a=.;
c=.;
d=.;
e=.;f=.;g=.;h=.;i=.;k=.;l=.;m=.;n=.;p=.;q=.;r=.;s=.;t=.;v=.;w=.;x=.;y=.;

jjt=
{{k},{n},{k},{n},{t},{t},{t},{t},{r},{s},{r},{s},{i},{i},{m},{i},{q},{h},{q},{h},{p},{p},{p},{p},{r},{r},{r},{r},{l},{l},{l},{l},{e},{d},{e},{d},{a},{a},{a},{a},{g},{g},{g},{g},{v},{v},{v},{v},{x},{y},{x},{y},{s},{s},{s},{s},{x},{c},{w},{c},{l},{f},{l},{f}};

nuchyd=.;

nuchyd=pivns.jjt;

(* 20 K&D hydropathy values, padded with 0.0 for stop *)
(* checked against http://web.expasy.org/protscale/pscale/Hphob.Doolittle.html *)

```
kd= {1.8,  2.5,  -3.5001,  -3.5002, 2.8,  -0.4, -3.2, 4.5, -3.9, 3.8, 1.9, -3.5003, -1.6, -3.5004, -4.5, -
0.8, -0.7,  4.2, -0.9,0.0,-1.3};


a=kd[[1]];
c=kd[[2]];
d=kd[[3]];
e=kd[[4]] ;
f=kd[[5]]  ;
g=kd[[6]] ;
h=kd[[7]] ;
i=kd[[8]] ;
k=  kd[[9]];
l=kd[[10]] ;
m=kd[[11]] ;
n=kd[[12]] ;
p=kd[[13]] ;
q=kd[[14]] ;
r=kd[[15]] ;
s=kd[[16]] ;
t=kd[[17]]  ;
v=kd[[18]];
w=kd[[19]];
x=kd[[20]]  ;
y=kd[[21]]  ;


nuchyd  (* numeric with above hydropathy values *)

{{-0.638558},{-1.25107},{0.686442},{0.880229},{-2.63865},{-0.163521},{-
1.46977},{3.94898},{-0.169808},{0.117704},{-0.388558},{0.117704}}
(* Reconstruct hydropathy values from nucleotide information using dot product of a weighted
vector for triplets *)
(* use single letter code for new vectors, precede by the letter 'a', because variables are global *)

(* Watch out!  Triplets are in  A,C,G,T for use here *)


aa={0,0,1,0,       0,1,0,0,  .25,.25,.25,.25};
aanuc=Total[aa.nuchyd,2]

ac={0,0,0,1,  0,0,1,0,  0, 0.5, 0, 0.5};
acnuc=Total[ac.nuchyd,2]

ad={0,0,1,0,  1,0,0,0,  0,0.5,0.0,0.5};
adnuc=Total[ad.nuchyd,2]
```

```
ae={0,0,1,0,  1,0,0,0,  0.5,0,0.5,0};
aenuc=Total[ae.nuchyd,2]

af={0,0,0,1,    0,0,0,1,    0,1,0,1};
afnuc=Total[af.nuchyd,2]

ag={0,0,1,0,    0, 0,1,0,     .25,.25,.25,.25};
agnuc=Total[ag.nuchyd,2]

ah={0,1,0,0,    1, 0,0,0,     0,0.5,0,0.5};
ahnuc=Total[ah.nuchyd,2]

ai={1,0,0,0,    0, 0,0,1,     1/3,1/3,0,1/3};
ainuc=Total[ai.nuchyd,2]

ak={1,0,0,0,    1, 0,0,0,     0.5,0,0.5,0};
aknuc=Total[ak.nuchyd,2]

al={0,2/3,0,1/3,    0, 0,0,1,     2/6,1/6,2/6,1/6};
alnuc=Total[al.nuchyd,2]

am={1,0,0,0,    0, 0,0,1,     0,0,1,0};
amnuc=Total[am.nuchyd,2]

an={1,0,0,0,    1, 0,0,0,     0,0.5,0,0.5};
annuc=Total[an.nuchyd,2]

ap={0,1,0,0,    0, 1,0,0,     .25,.25,.25,.25};
apnuc=Total[ap.nuchyd,2]

aq={0,1,0,0,    1, 0,0,0,     0.5,0,0.5,0};
aqnuc=Total[aq.nuchyd,2]

ar={1/3,2/3,0,0,    0, 0,1,0,     1/3,1/6,1/3,1/6};
arnuc=Total[ar.nuchyd,2]

as={1/3,0,0,2/3,    0, 2/3,1/3,0,     1/6,1/3,1/6,1/3};
asnuc=Total[as.nuchyd,2]

at={1,0,0,0,    0, 1,0,0,     .25,.25,.25,.25};
atnuc=Total[at.nuchyd,2]

av={0,0,1,0,    0, 0,0,1,     .25,.25,.25,.25};
avnuc=Total[av.nuchyd,2]
```

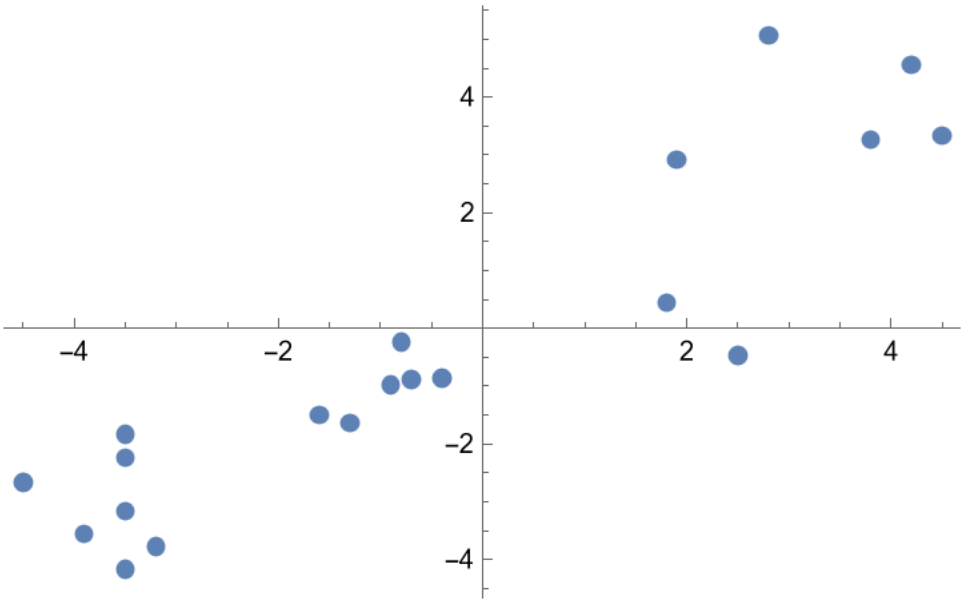
```
aw={0,0,0,1,    0, 0,1,0,     0,0,1,0};
awnuc=Total[aw.nuchyd,2]

ax={0,0,0,1,    2/3, 0,1/3,0,     2/3, 0,1/3,0};
axnuc=Total[ax.nuchyd,2]

ay={0,0,0,1,    1, 0,0,0,     0,1/2,0,1/2};
aynuc=Total[ay.nuchyd,2]
0.442181
-0.471838
-1.8345
-2.23139
5.06462
-0.864069
-3.77201
3.33229
-3.55639
3.26145
2.92186
-3.1595
-1.49533
-4.1689
-2.66356
-0.239563
-0.882819
4.55468
-0.9781
-1.61152
-1.64071
(* Fill empirical and SVP values for hydropathy into a 20 x 2 matrix *)

paired={{a,aanuc},{c,acnuc},{d,adnuc},{e,aenuc},{f,afnuc},{g,agnuc},{h,ahnuc},{i,ainuc},{k,a
knuc},{l,alnuc},{m,amnuc},{n,annuc},{p,apnuc},{q,aqnuc},{r,arnuc},{s,asnuc},{t,atnuc},{v,av
nuc},{w,awnuc},{y, aynuc}};

ListPlot[paired,PlotStyle->PointSize[0.02]]
FindFit[paired,intercept+slope*xaxis,{intercept,slope},xaxis]
Correlation[Take[paired,All,{1}],Take[paired,All,{2}]]
```



{intercept->0.00334568,slope->0.862005}

{{0.914183}}