

Genome wide Analysis of Psychrophiles and Mesophiles to Understand Low Temperature Adaptations of Proteins in Microorganisms

Raghu Prasad Rao Metpally and Boojala Vijay B. Reddy

Bioinformatics and *in silico* Drug Design Lab, Computer Science Department, Queens College, Flushing, NY 11367

And The Graduate Center – Computer Science Department, City University of New York (raghu@cs.cuny.edu, vijay@cs.cuny.edu)



Introduction

- Organisms live at a wide range of temperatures, from below 0°C to above 100°C. Despite the dramatic effects of temperature on the function and stability of their proteins.
- The comparison of microbial genomes and their lifestyles at global sequence levels is a subject of considerable interest, especially with the availability of many completely sequenced genomes.
- Are there any specific relationships associated with their temperature adaptations and proteome sequences?
- An understanding of the factors that enhance the thermal adaptations of proteins in extreme conditions is of particular interest because it raises the possibility of engineering enzymes with enhanced temperature stability and catalytic efficiency for industrial applications
- We performed comparative analysis using pairwise alignments of protein sequences from the six each of mesophilic and psychrophilic proteomes to investigate the general proteome wide characteristics for cold-adaptation among microorganisms

Methods

- We have carried out a comparative proteome analysis of 12 complete proteome sequences of psychrophilic (P), and mesophilic (M) microorganisms
- Secondary Structure Prediction and Principal Component Analysis (PCA)
- Orthologous pairs of all proteins were identified between the proteomes using BLAST search tool with 10^{-5} expectation value and > 40 bit scores.
- We have picked up alignments of the best possible single hit for every protein sequence in a query proteome that has an ortholog in the subject proteome.
- The alignments were parsed to calculate amino acid substitution counts between the two proteins of respective proteomes.
- The substitution counts were further normalized with respect to the composition of total amino acids in their respective proteomes. The resultant frequency of substitutions were further used to calculate two types of likelihood log odd scores *LOS1* and *LOS2*.
- The $F(X \rightarrow Y)$ represent normalized frequency of amino acid X substituted by the amino acid Y in the thermal group *Meso* compared to thermal group *Psycho*. Thus *LOS1* (Table V) scores indicate the pattern of substitutions observed between each pair of thermal groups. On the other hand the *LOS2* (Table V) scores indicate the pattern of substitutions that are purely correlated to their thermal adaptation and not due to speciation.
- Student t-test to evaluate the level of significance of *LOS2* substitution scores.

$$t = \frac{F_{Psycho} - F_{Meso}}{\sqrt{Var_{Psycho}/n_{Psycho} + Var_{Meso}/n_{Meso}}}$$

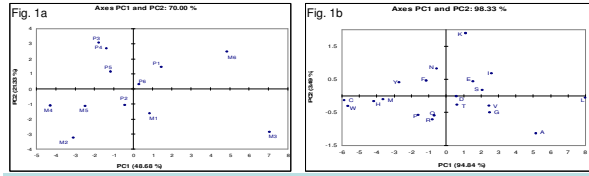
$$LOS1(X_{Meso} \rightarrow Y_{Psycho}) = \log \frac{F(X_{Meso} \rightarrow Y_{Psycho})}{F(Y_{Psycho} \rightarrow X_{Meso})}$$

$$LOS2(X_{Meso} \rightarrow Y_{Psycho}) = \log \frac{F(X_{Meso} \rightarrow Y_{Psycho})}{F(Y_{Meso} \rightarrow X_{Meso})}$$

TABLE I: Values for amino acid substitution preferences between proteins of Mesophilic and Psychrophilic proteomes. The values calculated using *LOS2* substitution scores of mutation frequencies normalized with substitution frequencies within the (a) mesophiles, (m2m) and (b) psychrophiles (p2p)

	Psychrophiles																			
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
a) m2pm2m	0.77	0.11	0.88	-1.97	-1.44	0.33	1.01	0.16	-2.07	-1.21	0.48	-0.85	-1.92	0.07	-0.91	0.78	1.01	-0.88	0.71	-1.56
b) p2mp2m	0.77	0.11	0.88	-1.97	-1.44	0.33	1.01	0.16	-2.07	-1.21	0.48	-0.85	-1.92	0.07	-0.91	0.78	1.01	-0.88	0.71	-1.56

Results



- PCA: Reduced dimensionality plot produces a separation of mesophiles and psychrophiles (Fig. 1a).
- Shows influence of GC pressure on global AA composition (Fig. 1b)

TABLE II: Distribution of amino acids and property group parameters in the predicted

Amino Acid	α-Helix				β-Sheet				Coil				
	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	Avg	SD	
Ala (A)	5.01	0.28	4.79	0.73	0.677	1.16	0.08	1.05	0.12	1.861	2.66	0.22	2.27

Secondary structures: Amino acids in Psychrophiles

- L, E and F show significantly low frequencies in α-helices
- D, T, A, G and S are significantly high in the β-sheets
- S, D, T, A, V and I are significantly high in the coil regions
- Majority of the property groups have significantly low frequencies in the helices
- Tiny, small, hydrophobic, neutral, acidic, aliphatic, and polar groups show significantly high frequencies in the coil regions

are shown in shaded boxes

	Psychrophiles																			
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
C	0.11	0.7																		

To note few substitutions with high +ve scores in psychrophiles: Amino acid A is preferred over E and K; D is preferred over E, F, K and P; G is preferred over K; H is preferred over F, W and Y; I is preferred over E; S is preferred over E, K, W and Y; T is preferred over E and K.

Similarly with high -ve scores E is avoided over A, D, I, S and T; F is avoided over H, K; K is avoided over A, D, G, S, T.

TABLE V: Log odd scores (LOS) of amino acid substitutions calculated using the equation (iii). The -ve values are shown in shaded boxes

	Psychrophiles																			
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
A	1.4	0.2	1.5	-5.3	-4.2	0.5	2.3	0.4	-8.8	-2.9	-1.2	-2.9	-2.8	0.2	-2.4	1.6	1.7	-1.2	-2.2	-4.9

TABLE I: The composition of individual amino acids and property groups in protein sequences of psychrophilic and mesophilic proteomes. The average (Avg) values among each set of proteomes along with their standard deviations (SD) are also given. Significant compositional differences as indicated by t-test parameter are shown in bold.

Amino Acid	Psychrophiles												Mesophiles																			
	Ala	Cys	Dpsy	Parc	Pory	Fatl	Phal	Avl	Hinf	Spce	HypL	Ecol	Leul	Leal	Avg	SD	Ala	Cys	Dpsy	Parc	Pory	Fatl	Phal	Avl	Hinf	Spce	HypL	Ecol	Leul	Leal	Avg	SD
Ala (A)	61	85	94	92	89	95	89	84	83	84	66	95	81	67	81	1.28	61	85	94	92	89	95	89	84	83	84	66	95	81	67	81	1.28

- In Psychrophiles: AA residues such as S, D, T and A are significantly preferred
- E and L are significantly avoided (Table I).
- Tiny, small and neutral groups are significantly preferred
- Aliphatic, basic, aromatic and hydrophilic groups are significantly avoided

TABLE II: Percentage of orthologs (best hits) in Psychrophiles and Mesophiles with respect to total number of sequences (size) in each of organisms considered

Genome	Psychrophiles												Mesophiles																	
	Cpsy	Dpsy	Parc	Pory	Fatl	Phal	Avl	Hinf	Spce	HypL	Ecol	Leul	Leal	Avg	SD	Ala	Cys	Dpsy	Parc	Pory	Fatl	Phal	Avl	Hinf	Spce	HypL	Ecol	Leul	Leal	Avg
Genome	100	125	193	178	147	173	163	211	104	153	122	141	13	144	4910	100	125	193	178	147	173	163	211	104	153	122	141	13	144	4910

Conclusions

Compared to mesophilic proteins, the change of amino acid composition in psychrophilic proteins can be categorized into six distinct aspects:

- Increase in frequency of individual amino acids like S, D, T and A
- Decrease in E and L amino acids
- An increase in small, tiny and neutral group residues which contribute to protein flexibility
- Decrease in charged amino acids, particularly basic residues which contribute to the ionic interactions
- Decrease in aromatic amino acids residues mainly F, Y and W residues which contribute to the cation-π interactions
- Decrease in aliphatic residues which provide good covering and masking to produce hydrophobic pockets that are involved in stabilizing the protein structure

The *LOS2* values are influenced by factors other than temperature might be nullified. Psychrophilic proteins prefer to have amino acids K, E, Y, F, W, P and L, on the other hand they tries to avoid amino acids such as S, D, A, T, C, H, M, and Q as compared to mesophile proteins.

References

Samad Jahandeh, Ebrahim Barzegari Asadabadi, Parviz Abolmaleki, Mina Jahandeh, Somayeh Hoseni. Protein psychrophilicity: Role of residual structural properties in adaptation of proteins to low temperatures Journal of Theoretical Biology 248 (2007) 721–726

McDonald, J.H. (2001) Patterns of temperature adaptation in proteins from the bacteria *Deinococcus radiodurans* and *Thermus thermophilus*. Molecular Biology and Evolution 18:741–749

Steinar Thorsvaldsen, Erik Hjerde, Chris Fenton Nils P. Willason. Molecular characterization of cold adaptation based on ortholog protein sequences from Vibrionaceae species Extremophiles (2007) 11:719–732

- Preferred Class:**
 First: QT, LH, NS, PA, ED, ES and ET (-3 t value)
 Second: IH, RS, PS, EH, HA, ID, VM, RT, PD, VD, ND, YH, IS, SD, QD, LS, DS, CH, LM, DA, WS, EA, QS, QA, HS, YS (2 to 3)
- Third:** VG, MD, WH, HT, CN, QG, WQ, LT, FM, LQ, QC, KD, CQ, FD, DI, SA, TM, RD, NA, WA, TD, TA, FH, TS, RA, IM, VS, PT, GD, QM, IT, FS, NT, WC, VT, AD, KT, KS, DT, VQ (13 to 2)
- Avoiding class:**
 First: DE (-3 to -8)
 Second: SE, ME, VE, IE, HF, HY, HE, VK, HK, GK, TK, EE, EK, TW and AK (-3 to -2)
 Third: LE, AE, DK, GH, SF, YV, GY, GE, QE, EP, VY, AP, PY, NK, MF, LT, FK, RE, DY, EY, CL, DP, EW, GR, PE, MK, IW, PK, AY, HW, GN, MW, MY, NE, ML, QF, AF, FP, PK, IK, GL, WK, DN, YK, LF, SY, VL, QY, KY, DF, SF and VF (-2 to -1.3)