Making sense of the past: hyperstability of ancestral thioredoxins explained by free energy simulations

Álvaro Cortés Cabrera, Pedro A. Sánchez-Murcia and Federico Gago

Área de Farmacología, Departamento de Ciencias Biomédicas. Facultad de Medicina y Ciencias de la Salud. Universidad de Alcalá. 28871 Alcalá de Henares, Madrid

Supplementary material

Supplementary Table 1. Wild-type dataset from Ingles-Prieto *et al.*[1]. Mutations were introduced on the wild-type E. coli Trx sequence. The value 0 or 1 denotes the absence or the presence, respectively, of a particular mutation. T_m is the melting temperature expressed in Celsius degrees and $\Delta T_{m,WT}$ is the difference between the melting temperatures of variant and wild-type enzymes.

Variant	G21K	A22P	E30P	G65N	P68A	G74S	S95P	G97E	T_m (ºC)	$\Delta T_{m,WT}$
wild-type	0	0	0	0	0	0	0	0	87.3	0
var1	1	1	1	0	1	1	1	0	94.9	7.6
var2	1	1	0	0	1	0	1	1	97.3	10
var5	1	0	0	0	0	1	0	0	89.3	2
var7	1	0	0	0	1	0	0	0	94.5	7.2
var8	1	0	1	1	1	0	1	1	93.1	5.8
var12	0	0	1	0	0	0	0	1	89.7	2.4
var13	0	0	0	1	1	0	1	0	91.2	3.9
var14	0	0	1	1	0	0	1	1	85.9	-1.4
var15	1	0	1	1	1	0	1	0	88.0	0.7
var16	0	1	0	0	1	0	0	1	99.1	11.8
var17	1	1	1	1	1	0	1	1	96.9	9.6
var19	1	1	0	0	0	0	1	1	88.7	1.4
var22	1	1	1	1	0	0	1	1	89.5	2.2
var23	0	0	0	1	0	0	0	0	87.6	0.3
var24	0	1	1	1	1	0	1	0	91.9	4.6
var25	1	0	1	0	0	0	1	1	83.8	-3.5
var27	0	0	1	1	0	1	0	1	94.7	7.4
var32	0	0	0	0	0	0	1	1	85.4	-1.9
var36	0	1	0	1	0	0	1	0	85.3	-2
var39	1	0	1	1	0	0	0	1	90.8	3.5
var40	0	0	0	1	1	0	0	1	99.8	12.5
var41	1	1	1	0	1	0	1	0	91.5	4.2
var42	0	1	0	0	1	1	0	1	102.5	15.2
var43	1	0	1	0	0	1	1	0	81.3	-6
var22b	1	1	1	1	1	1	0	1	104.4	17.1

var31	1	1	0	1	1	1	0	1	105.0	17.7
var32b	0	1	1	1	1	1	0	1	102.3	15
var41b	0	1	0	1	1	1	0	1	103.4	16.1
var33	1	1	1	0	1	1	0	1	102.7	15.4
var42b	1	1	0	0	1	1	0	1	103.7	16.4
var23b	1	1	0	1	1	1	1	1	100.9	13.6
var11	1	1	1	1	1	0	0	1	101.1	13.8
var21	1	1	0	1	1	0	0	1	102.2	14.9
var44	0	1	1	0	1	1	0	1	101.0	13.7
var12b	1	1	1	1	1	1	1	1	99.9	12.6

Supplementary Table 2. Var31K90L dataset from Ingles-Prieto *et al.*[1]. Amino acid replacements were introduced in the var31+K90L *E. coli* Trx sequence. The value 0 or 1 denotes the absence or the presence, respectively, of the mutation. T_m is the melting temperature expressed in Celsius degrees and $\Delta T_{m,v31}$ is the difference between the variant melting temperature and the variant31+K90L.

Variant	A87V	A88D	T89R	L94Q	Q98A	F102R	L103I	N106H	T_m (°C)	$\Delta T_{m,v31}$
var31K90L	0	0	0	0	0	0	0	0	110.3	0
subvar2	1	1	0	1	1	0	0	1	106.4	-3.9
subvar3	0	1	0	1	0	1	1	0	99.1	-11.2
subvar4	0	1	0	0	1	0	1	1	104.8	-5.5
subvar5	1	0	0	0	0	1	0	0	106.5	-3.8
subvar6	1	0	1	0	0	1	0	0	107.2	-3.1
subvar7	1	0	0	1	1	0	0	1	111.7	1.4
subvar8	0	1	0	1	1	0	0	0	101.4	-8.9
subvar9	0	1	0	1	0	0	0	1	103.5	-6.8
subvar10	1	1	1	0	1	1	0	0	106.6	-3.7
subvar11	0	1	0	0	0	1	0	1	104.9	-5.4
subvar12	1	1	0	1	1	1	0	1	104.6	-5.7
subvar13	0	1	0	0	1	0	0	1	102.3	-8
subvar14	0	1	0	1	1	0	0	1	103.2	-7.1
subvar15	0	0	0	0	0	0	1	1	110.5	0.2
subvar16	1	0	0	1	1	1	1	0	103.7	-6.6
subvar17	1	0	1	1	0	0	1	1	112.2	1.9
subvar18	0	0	0	0	0	1	0	1	105.1	-5.2
subvar19	1	1	1	1	0	1	1	1	106.0	-4.3
subvar20	0	0	1	1	0	0	0	0	108.7	-1.6
subvar21	1	0	1	0	1	0	0	1	116.1	5.8
subvar22	0	1	0	1	1	1	0	1	101.4	-8.9
subvar23	0	1	0	0	1	1	0	1	104.8	-5.5
subvar24	1	1	0	0	0	0	0	0	106.7	-3.6
subvar25	1	0	0	0	0	0	0	0	112.7	2.4
subvar26	1	1	1	0	0	1	1	0	107.8	-2.5
subvar27	1	1	1	0	1	0	0	0	108.1	-2.2
subvar28	0	0	0	1	1	1	0	0	99.84	-10.46
subvar29	0	1	1	0	0	1	1	1	106.4	-3.9
subvar30	1	0	0	1	1	0	1	1	111.6	1.3
subvar31	1	0	0	0	0	0	1	1	113.9	3.6
subvar32	1	1	1	1	1	1	1	0	103.7	-6.6

subvar33	0	1	0	1	1	0	1	1	103.2	-7.1
subvar34	1	1	0	1	1	1	0	0	101.9	-8.4
subvar35	0	1	1	1	1	0	1	0	104.2	-6.1

Supplementary Table 3. Frequency of each amino acid at position 87 of *E. coli* Trx in the thioredoxin protein family.

Residue	Frequency (%)
V	38.51
I	10.70
L	8.94
A	8.01
K	5.76
E	4.35
T	3.27
S	3.24
P	3.21
Q	2.49
R	2.44
M	1.85
F	1.63
Y	1.60
D	1.49
N	0.82
H	0.68
G	0.59
C	0.34
W	0.08

Supplementary Table 4. Frequency of each amino acid at position 87 of *E. coli* Trx in thermophilic thioredoxins. Information about growing conditions was extracted from the PGTdb database[2]

Residue	Frequency (%)
V	52.83
I	13.21
L	11.32
A	9.43
K	5.03
Q	3.14
F	1.89
D	0.63
M	0.63
R	0.63
T	0.63
Y	0.63

1. Á. I. Prieto, Universidad de Granada, 2012.
2. S.-L. Huang, L.-C. Wu, H.-K. Liang, K.-T. Pan, J.-T. Horng and M.-T. Ko, *Bioinformatics*, 2004, **20**, 276-278.