Supplementary Material

Towards an accurate prediction of the thermal stability of homologous proteins

Fabrizio Pucci and Marianne Rooman

- Table S1. Predicted and Experimental values of the melting temperature for the set of 45 proteins belonging to the 11 homologous families.
- Table S2. Family-dependent T_m - T_{env} regression lines.
- Table S3. List of 45 proteins with known melting temperature analyzed in this study and their characteristics.

Protein PDB code	T_m^{exp} (°C)	$T_m^{\text{pre}}(1)$ (°C)	$T_m^{\text{pre}}(2)$ (°C)	$T_m^{\text{pre}}(3)$ (°C)	$T_m^{\text{pre}}(4)$ (°C)	$T_m^{\text{pre}}(5)$ (°C)	$T_m^{\text{pre}}(6)$ (°C)	$T_m^{\text{pre}}(7)$ (°C)	$T_m^{\text{pre}}(8)$ (°C)	$T_m^{\text{pre}}(T)$ (°C)	T_m^{env} (°C)	Family	Host Organism
1aqh	43.7	55.3	65.7	65.6	60.9	52.9	57.8	59.3	59.8	49.0	20.0	α -amylase	Alteromonas haloplanctis
1ppi	65.6	67.8	76.0	67.0	62.6	55.0	71.0	59.0	67.4	57.4	39.0	α -amylase	Porcine (Sus Scrofa)
1jae	65.9	61.5	57.3	70.3	63.9	71.9	55.2	58.6	62.1	64.8	30.0	α -amylase	Tenebrio Molitor
1smd	70.3	65.8	67.0	65.9	61.1	67.8	68.0	58.6	62.6	60.9	37.0	α -amylase	Homo Sapiens
1an7	52.3	65.8	64.9	74.8	60.3	55.7	59.1	63.4	53.5	52.4	37.0	Lysozyme	Bacteriophage Lambda
2lzm	64.8	65.8	58.6	74.8	79.6	81.2	60.1	63.4	62.7	53.4	37.0	Lysozyme	E. Coli
1lz1	64.9	65.8	58.6	52.3	60.5	55.9	69.9	63.7	49.3	61.1	37.0	Lysozyme	Homo Sapiens
4lyz	74.8	68.3	67.1	64.9	56.9	65.6	57.5	63.7	75.5	69.1	41.0	Lysozyme	Gallus Gallus
2fal	52.0	53.4	119.2	82.2	80.7	65.6	75.7	62.9	59.4	49.7	17.0	Myoglobin	Aplysia limacina
1ymb	78.3	66.5	85.4	82.2	59.7	56.6	67.1	62.9	63.8	64.6	38.0	Myoglobin	Equus Caballus
1bvc	82.2	65.2	75.8	78.3	66.7	109.0	74.3	62.9	69.6	74.1	36.0	Myoglobin	Cetecea Sperm
1blc	41.6	62.7	52.3	65.0	58.2	59.7	71.6	63.5	71.6	58.3	32.0	β -lactamase	Staphylococcus aureus
1ke4	54.6	65.8	58.0	41.6	56.3	42.8	69.2	62.8	68.5	51.5	37.0	β -lactamase	E.Coli
4blm	66.0	67.7	55.0	41.6	47.9	52.7	59.7	63.5	65.4	68.5	40.0	β -lactamase	Bacillus Licheniformis
1bmc	51.0	61.5	35.4	41.6	51.0	47.9	63.1	63.8	61.4	52.5	30.0	β -lactamase	Becillus Cereus
1hml	39.5	65.8	41.6	70.8	70.8	52.0	75.5	64.9	57.9	58.2	37.0	α -lactalbumin	Homo Sapiens
1hfx	56.2	66.5	55.1	70.8	51.5	76.0	68.4	64.9	51.9	69.2	38.0	α -lactalbumin	Bos Taurus
1hmk	70.8	67.1	72.9	56.2	45.9	48.5	65.8	64.9	62.2	65.2	39.0	α -lactalbumin	Capra Hircus
2acy	53.8	66.5	75.8	111.5	78.2	55.1	59.5	60.2	40.8	52.8	38.0	Acylphosphatase	Bos Taurus
2bjd	100.8	92.5	94.2	111.5	94.5	92.1	91.3	60.2	73.8	94.3	80.0	Acylphosphatase	Sulfolobus Solfataricus
1v3z	111.5	103.7	120.9	100.8	97.9	123.2	79.8	60.2	71.7	114.7	98.0	Acylphosphatase	Pyrococcus Horikoshii
1p3j	47.6	61.5	52.7	74.8	52.8	55.2	71.6	64.7	68.2	52.3	30.0	Adenylate kinase	Bacillus Subtilis
3fb4	47.6	54.1	41.2	74.8	56.1	52.6	60.4	64.7	65.0	43.3	18.0	Adenylate kinase	Jeotgalibacillus marinus
1s3g	43.4	52.2	39.6	51.8	54.5	52.3	58.4	64.6	68.0	41.5	15.0	Adenylate kinase	Bacillus Globisporus
1aky	47.7	59.6	50.1	51.8	51.9	52.9	61.4	64.6	43.1	51.1	27.0	Adenylate kinase	Saccharomyces Cerevisiae
1ank	51.8	65.8	58.2	47.6	51.0	50.0	75.6	64.6	58.0	50.4	37.0	Adenylate kinase	E.Coli
1zip	74.8	77.0	56.3	47.6	48.0	50.6	65.3	64.6	47.6	74.0	55.0	Adenylate kinase	Bacillus Stearothermophilus
1oa3	49.2	67.7	64.0	54.5	54.2	52.2	53.9	63.0	71.4	61.7	40.0	Cell 12A	Hypocrea Schweinitzii
1h8v	54.5	64.6	60.4	49.2	52.7	61.5	52.6	63.0	74.3	59.4	34.0	Cell 12A	Trichoderma Reesei
1oa4	66.8	61.5	48.0	49.2	54.9	78.6	62.6	63.0	71.4	56.7	30.0	Cell 12A	Streptomyces sp. 1lag8
1olr	68.7	73.9	63.1	49.2	47.6	59.2	65.3	62.9	71.1	67.0	50.0	Cell 12A	Humicola Grisea
1cec	70.4	80.1	64.6	54.5	112.7	54.2	77.8	61.6	56.4	73.3	60.0	Cell 12A	Clostridium Thermocellum

1csp	53.8	61.5	52.4	76.9	57.2	60.3	77.1	64.8	64.9	47.6	30.0	Cold Shock Protein	Bacillus Subtilis
1mjc	57.0	65.8	57.8	53.8	62.8	71.4	60.8	64.8	44.5	65.0	37.0	Cold Shock Protein	E. Coli
1c9o	76.9	86.3	70.4	53.8	65.2	50.9	65.8	64.9	58.6	79.8	70.0	Cold Shock Protein	Bacillus Caldolyticus
1bu7	47.0	61.5	56.2	88.0	79.9	70.3	66.3	57.1	55.4	55.2	30.0	Cytochrome p450	Bacillus Megaterium
1oxa	55.1	62.1	53.6	91.0	75.7	72.2	67.9	58.1	56.8	57.3	31.0	Cytochrome p450	Saccaropolyspora Erythraea
1akd	56.0	60.3	49.1	88.0	78.3	53.3	66.1	58.1	68.5	48.0	28.0	Cytochrome p450	Pseudomonas putida
1n97	88.0	85.1	82.9	55.1	63.7	64.6	62.1	58.5	63.7	91.5	68.0	Cytochrome p450	Thermus thermophilus
1f4t	91.0	91.3	96.9	55.1	56.8	77.6	67.7	58.8	68.8	83.8	78.0	Cytochrome p450	Solfolobus Solfataricus
1rgg	49.3	59.6	45.2	102.0	67.6	48.9	60.0	64.6	62.3	52.4	27.0	Ribonuclease	Streptomyces aureofaciens
9rnt	50.9	64.6	56.0	63.6	70.8	53.7	67.1	64.5	65.4	44.6	35.0	Ribonuclease	Aspergillus Oryzae
1rnh	53.2	65.8	57.9	50.9	59.1	50.3	71.6	64.5	61.1	53.9	37.0	Ribonuclease	E. Coli
1rbn	63.6	66.5	56.4	102.0	58.2	52.3	68.0	63.9	66.1	65.6	38.0	Ribonuclease	Bos Taurus
2ehg	102.0	92.5	97.1	63.6	59.4	123.4	74.2	63.9	70.4	105.6	80.0	Ribonuclease	Sulfolobus tokodaii

TABLE S1. Predicted and Experimental values of the melting temperature for the set of 45 proteins belonging to the 11 homologous families. In column 2 the experimental melting temperatures are listed. In columns 3-10 the predicted T_m 's with the different methods are reported. In column 11 there are the results obtained from the combined tool. Finally the environmental temperature of the host organism, the family to which proteins belong and the name of the host organism are reported in column 12,13,14 respectively.

Family	$T_m^{(2) \text{ est}}$	N
α -Amylase	$1.23 T_{env} + 22.4 \text{ }^\circ\text{C}$	4
Lysozyme	$3.53 T_{env} - 69.9^\circ\text{C}$	4
Myoglobin	$1.39 T_{env} + 28.8 \text{ }^\circ\text{C}$	3
β -Lactamase	$1.71 T_{env} - 6.4 \text{ }^\circ\text{C}$	4
α -Lactalbumin	$15.6 T_{env} - 538.8 \text{ }^\circ\text{C}$	3
Acylphosphatase	$1.00 T_{env} + 17.5 \text{ }^\circ\text{C}$	3
Adenylate kinase	$0.72 T_{env} + 30.2 \text{ }^\circ\text{C}$	6
Cell 12A Endoglucanase	$0.39 T_{env} + 45.2 \text{ }^\circ\text{C}$	5
Cold shock protein	$0.59 T_{env} + 35.6 \text{ }^\circ\text{C}$	3
Cytochrome P450	$0.83 T_{env} + 28.2 \text{ }^\circ\text{C}$	5
Ribonuclease	$1.05 T_{env} + 18.1 \text{ }^\circ\text{C}$	5

Table S 2. Family-dependent T_m - T_{env} regression lines. N is the number of proteins in the family.

Table S 3: List of 45 proteins with known melting temperature analyzed in this study and their characteristics.

Protein PDB code	T_m^{exp} (°C)	pH	Protein Name	Res. (Å)	Host Organism	T_{env} (°C)	N	Reference
1aqh	43.7	7.2	α -amylase	2.00	<i>Alteromonas haloplantis</i>	20.0	448	J Biol Chem 276, 25791 (2001)
1ppi	65.6	7.2	α -amylase	2.20	<i>Sus Scrofa</i>	39.0	496	J Biol Chem 276, 25791 (2001)
1jae	65.9	7.2	α -amylase	1.65	<i>Tenebrio Molitor</i>	30.0	470	J Biol Chem 276, 25791 (2001)
1smd	70.3	7.2	α -amylase	1.60	<i>Homo Sapiens</i>	37.0	495	J Biol Chem 276, 25791 (2001)
1am7	52.3	7.0	Lysozyme	2.30	<i>Bacteriophage Lambda</i>	37.0	150	FEBS Lett 460, 442 (1999)
2lzm	64.8	6.5	Lysozyme	1.70	<i>Escherichia Coli</i>	37.0	164	Nature 334,406 (1988), PNAS 85,401 (1988)
1lz1	64.9	2.7	Lysozyme	1.50	<i>Homo Sapiens</i>	37.0	130	J Mol Biol 254, 62 (1995)
4lyz	74.8	7.0	Lysozyme	2.00	<i>Gallus Gallus</i>	41.0	129	Proteins 40, 49 (2000)
2fal	52.0	7.0	Myoglobin	1.80	<i>Aplysia Limacina</i>	17.0	146	J Mol Biol 297, 1231 (2000)
1ymb	78.3	7.0	Myoglobin	1.90	<i>Equus Caballus</i>	38.0	153	Biochemistry 4, 5075 (2001)
1bvc	82.2	9.6	Myoglobin	1.50	<i>Physeter Catodon</i>	36.0	153	Protein Sci 2, 1099 (1993)
1blc	41.6	7.50	β -lactamase	2.20	<i>Staphylococcus Aureus</i>	32.0	257	Biochemistry 33, 116 (1994)
1ke4	54.6	6.80	β -lactamase	1.72	<i>Escherichia Coli</i>	37.0	357	Protein Sci 8, 1816 (1999)
4blm	66.0	7.00	β -lactamase	2.00	<i>Bacillus Licheniformis</i>	40.0	256	Biochemistry 29, 5797 (1990)
1bmc	51.0	7.00	β -lactamase	2.50	<i>Bacillus Cereus</i>	30.0	213	Biochemistry 61, 6603 (1992)
1hml	39.5	7.00	α -lactalbumin	1.70	<i>Homo Sapiens</i>	37.0	123	Biochemistry 28, 8568 (1989)
1hfx	56.2	7.40	α -lactalbumin	2.30	<i>Bos Taurus</i>	38.0	123	Protein Eng 12, 581 (1999)
1hmk	70.8	7.50	α -lactalbumin	2.00	<i>Capra Hircus</i>	39.0	121	Proteins 60, 118 (2005)
2acy	53.8	5.50	Acylphosphatase	1.45	<i>Bos Taurus</i>	38.0	99	Proteins 62, 64 (2006)
2bjd	100.8	5.50	Acylphosphatase	1.27	<i>Sulfolobus Solfataricus</i>	80.0	90	Proteins 62, 64 (2006)
1v3z	111.5	7.40	Acylphosphatase	1.72	<i>Pyrococcus Horikoshii</i>	98.0	91	Biochemistry 38, 33 (2005)
1p3j	47.6	7.40	Adenylate kinase	1.90	<i>Bacillus Subtilis</i>	30.0	212	J Bio Chem 279, 28202 (2004)
3fb4	47.6	7.40	Adenylate kinase	2.00	<i>Jeotgalibacillus Marinus</i>	18.0	216	J Bio Chem 279, 28202 (2004)
1s3g	43.4	7.40	Adenylate kinase	2.25	<i>Bacillus Globisporus</i>	15.0	217	J Bio Chem 279, 28202 (2004)
1aky	47.7	7.50	Adenylate kinase	1.96	<i>Saccharomyces Cerevisiae</i>	27.0	218	Eur J Biochem 231, 405 (1995)
1ank	51.8	7.20	Adenylate kinase	2.00	<i>Escherichia Coli</i>	37.0	214	J Biol Chem 266, 23654 (1991)
1zip	74.8	7.20	Adenylate kinase	1.85	<i>Bacillus Stearothermophilus</i>	55.0	217	Biochemistry 31, 3038 (1992)
1oa3	49.2	8.00	Cell 12A	1.70	<i>Hypocrea Schweinitzii</i>	40.0	217	Protein Sci 12, 848 (2003)
1h8v	54.5	8.00	Cell 12A	1.90	<i>Trichoderma Reesei</i>	34.0	217	Protein Sci 12, 2782 (2003)
1oa4	66.8	8.00	Cell 12A	1.50	<i>Streptomyces sp. 11ag8</i>	30.0	222	Protein Sci 12, 848 (2003)

1olr	68.7	8.00	Cell 12A	1.20	<i>Humicola Grisea</i>	50.0	223	Protein Sci 12, 2872 (2003)
1cec	70.4	7.20	Cell 12A	2.15	<i>Clostridium Thermocellum</i>	60.0	331	Biophys Chem 132, 229 (2002)
1csp	53.8	7.00	Cold Shock Protein	2.50	<i>Bacillus Subtilis</i>	30.0	67	J Mol Bio 347, 1063 (2005)
1mjc	57.0	7.00	Cold Shock Protein	2.00	<i>Escherichia Coli</i>	37.0	69	Protein Sci 9, 387 (2000), Protein Sci 10, 2028 (2001)
1c9o	76.9	7.00	Cold Shock Protein	1.17	<i>Bacillus Caldolyticus</i>	70.0	66	Nat Struct Biol 7, 380 (2000), J Mol Biol 319, 541 (2002)
1bu7	47.0	7.40	Cytochrome p450	1.65	<i>Bacillus Megaterium</i>	30.0	455	J Biol Chem 278, 608 (2003)
1oxa	55.1	7.40	Cytochrome p450	2.10	<i>Saccaropolyspora Erythraea</i>	31.0	403	J Biol Chem 278, 608 (2003)
1akd	56.0	7.40	Cytochrome p450	1.80	<i>Pseudomonas putida</i>	28.0	405	J Biol Chem 278, 608 (2003)
1n97	88.5	7.40	Cytochrome p450	1.80	<i>Thermus thermophilus</i>	68.0	385	J Biol Chem 278, 608 (2003)
1f4t	91.0	7.40	Cytochrome p450	1.93	<i>Solfolobus Solfataricus</i>	78.0	367	J Biol Chem 278, 608 (2003)
1rgg	49.3	7.00	Ribonuclease	1.40	<i>Streptomyces aureofaciens</i>	27.0	96	Biochemistry 37, 16192 (1998), Protein Sci 8, 1843 (1999)
9rnt	50.9	7.00	Ribonuclease	1.50	<i>Aspergillus Oryzae</i>	35.0	104	J Mol Biol 279, 31790 (2003), J Mol Biol 354, 967 (2005)
1rnh	53.2	5.50	Ribonuclease	2.00	<i>Escherichia Coli</i>	37.0	155	J Bio Chem 264, 11621 (1989), Biochemistry 33, 3312 (1994)
1rbn	63.6	7.10	Ribonuclease	2.10	<i>Bos Taurus</i>	38.0	123	Eur J Biochem 220, 663 (1994)
2ehg	102.0	7.00	Ribonuclease	1.60	<i>Sulfolobus tokodaii</i>	80.0	149	Biochemistry 45, 10795 (2009)
								PLoS ONE 6, e16226 (2011)