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       | Family     | Host Organization | PDB code | $T_{exp}$ (ºC) | $T_{env}$ (ºC) | $T_{pre1}$ (ºC) | $T_{pre2}$ (ºC) | $T_{pre3}$ (ºC) | $T_{pre4}$ (ºC) | $T_{pre5}$ (ºC) | $T_{pre6}$ (ºC) | $T_{pre7}$ (ºC) | $T_{pre8}$ (ºC) | $T_{pre9}$ (ºC) | $T_{pre10}$ (ºC) | $T_{pre11}$ (ºC) | $T_{pre12}$ (ºC) | $T_{pre13}$ (ºC) | $T_{pre14}$ (ºC) | $T_{pre15}$ (ºC) | $T_{pre16}$ (ºC) | $T_{pre17}$ (ºC) | $T_{pre18}$ (ºC) | $T_{pre19}$ (ºC) | $T_{pre20}$ (ºC) | $T_{pre21}$ (ºC) | $T_{pre22}$ (ºC) | $T_{pre23}$ (ºC) | $T_{pre24}$ (ºC) | $T_{pre25}$ (ºC) | $T_{pre26}$ (ºC) | $T_{pre27}$ (ºC) | $T_{pre28}$ (ºC) | $T_{pre29}$ (ºC) | $T_{pre30}$ (ºC) | $T_{pre31}$ (ºC) | $T_{pre32}$ (ºC) | $T_{pre33}$ (ºC) | $T_{pre34}$ (ºC) | $T_{pre35}$ (ºC) | $T_{pre36}$ (ºC) | $T_{pre37}$ (ºC) | $T_{pre38}$ (ºC) | $T_{pre39}$ (ºC) | $T_{pre40}$ (ºC) |
|--------------|------------|-------------------|----------|---------------|--------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| Protein ID | Experiment | Prediction 1 | Prediction 2 | Prediction 3 | Prediction 4 | Prediction 5 | Prediction 6 | Prediction 7 | Prediction 8 | Prediction 9 | Prediction 10 | Combined Tool | Host Organism | Family |
|------------|------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|---------------|--------------|---------|
| 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 | Sarcaropolyspora 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 |
| 1fit       | 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                      | $T_{\text{m}}^{(2)} \text{est}$   | $T_{\text{env}}$ | $N$ |
|-----------------------------|----------------------------------|------------------|-----|
| α-Amylase                   | 1.23 $T_{\text{env}} + 22.4 \degree \text{C}$ | 4                |
| Lysozyme                    | 3.53 $T_{\text{env}} - 69.9 \degree \text{C}$ | 4                |
| Myoglobin                   | 1.39 $T_{\text{env}} + 28.8 \degree \text{C}$ | 3                |
| β-Lactamase                 | 1.71 $T_{\text{env}} - 6.4 \degree \text{C}$ | 4                |
| α-Lactalbumin               | 15.6 $T_{\text{env}} - 538.8 \degree \text{C}$ | 3                |
| Acylphosphatase             | 1.00 $T_{\text{env}} + 17.5 \degree \text{C}$ | 3                |
| Adenylate kinase            | 0.72 $T_{\text{env}} + 30.2 \degree \text{C}$ | 6                |
| Cell 12A Endoglucanase      | 0.39 $T_{\text{env}} + 45.2 \degree \text{C}$ | 5                |
| Cold shock protein          | 0.59 $T_{\text{env}} + 35.6 \degree \text{C}$ | 3                |
| Cytochrome P450              | 0.83 $T_{\text{env}} + 28.2 \degree \text{C}$ | 5                |
| Ribonuclease                | 1.05 $T_{\text{env}} + 18.1 \degree \text{C}$ | 5                |

**Table S 2.** Family-dependent $T_{\text{m}}$-$T_{\text{env}}$ 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 | $T_{exp}$ (°C) | pH | Protein Name | Res. (Å) | Host Organism | $T_{mwp}$ (°C) | N | Reference |
|---------|----------------|----|--------------|---------|---------------|---------------|---|-----------|
| 1aqh    | 43.7           | 7.2 | α-amylase    | 2.00    | Alteromonas haloplancis | 20.0 | 448 | J Biol Chem 276, 25791 (2001) |
| 1ppi    | 65.6           | 7.2 | α-amylase    | 2.20    | Sus Scrofa     | 39.0 | 496 | J Biol Chem 276, 25791 (2001) |
| 1sae    | 65.9           | 7.2 | α-amylase    | 1.65    | Tenebrio Molitor | 30.0 | 470 | J Biol Chem 276, 25791 (2001) |
| 1smd    | 70.3           | 7.2 | α-amylase    | 1.60    | Homo Sapiens   | 37.0 | 495 | J Biol Chem 276, 25791 (2001) |
| 1am7    | 52.3           | 7.0 | Lysozyme     | 2.30    | Bacteriophage Lambda | 37.0 | 150 | FEBS Lett 460, 442 (1999) |
| 2lzm    | 64.8           | 6.5 | Lysozyme     | 1.70    | Escherichia Coli | 37.0 | 164 | Nature 334,406 (1988), PNAS 85,410 (1988) |
| 1zl1    | 64.9           | 2.7 | Lysozyme     | 1.50    | Homo Sapiens   | 37.0 | 130 | J Mol Biol 254, 62 (1995) |
| 4lyz    | 74.9           | 7.0 | Lysozyme     | 2.00    | Gallus Gallus  | 41.0 | 129 | Proteins 40, 49 (2000) |
| 2fal    | 52.0           | 7.0 | Myoglobin    | 1.80    | Aplysia Limacina | 17.0 | 146 | J Mol Biol 297, 1231 (2000) |
| 1ymb    | 78.3           | 7.0 | Myoglobin    | 1.90    | Equus Caballus | 38.0 | 153 | Biochemistry 4, 5075 (2001) |
| 1bvc    | 82.2           | 9.6 | Myoglobin    | 1.50    | Physter Cadocon | 36.0 | 153 | Protein Sci 2, 1099 (1993) |
| 1blc    | 41.6           | 7.50| β-lactamase  | 2.20    | Staphylococcus Aureus | 32.0 | 257 | Biochemistry 33, 116 (1994) |
| 1ke4    | 54.6           | 6.80| β-lactamase  | 1.72    | Escherichia Coli | 37.0 | 357 | Protein Sci 8, 1816 (1999) |
| 4blm    | 66.0           | 7.00| β-lactamase  | 2.00    | Bacillus Licheniformis | 40.0 | 256 | Biochemistry 29, 5797 (1990) |
| 1bnc    | 51.0           | 7.00| β-lactamase  | 2.50    | Bacillus Cereus | 30.0 | 213 | Biochemistry 61, 6603 (1992) |
| 1hml    | 39.5           | 7.00| α-lactalbumin| 1.70    | Homo Sapiens   | 37.0 | 123 | Biochemistry 28, 8568 (1989) |
| 1hfa    | 56.2           | 7.40| α-lactalbumin| 2.30    | Bos Taurus     | 38.0 | 123 | Protein Eng 12, 581 (1999) |
| 1hmk    | 70.8           | 7.50| α-lactalbumin| 2.00    | Capra Hircus   | 39.0 | 121 | Proteins 60, 118 (2005) |
| 2acy    | 53.8           | 5.50| Acylphosphatase | 1.45  | Bos Taurus     | 38.0 | 90 | Proteins 62, 64 (2006) |
| 2bjd    | 100.8          | 5.50| Acylphosphatase| 1.27  | Sulfolobus Solfataricus | 80.0 | 90 | Proteins 62, 64 (2006) |
| 1v3z    | 111.5          | 7.40| Acylphosphatase| 1.72  | Pyrococcus Horkoshii | 98.0 | 91 | Biochemistry 38, 35 (2005) |
| 1p3j    | 47.6           | 7.40| Adenylate kinase | 1.90  | Bacillus Subtilis | 30.0 | 212 | J Bio Chem 279, 28202 (2004) |
| 3fb4    | 47.6           | 7.40| Adenylate kinase | 2.00  | Jeotgalibacillus Marinus | 18.0 | 216 | J Bio Chem 279, 28202 (2004) |
| 1s3g    | 43.4           | 7.40| Adenylate kinase | 2.25  | Bacillus Globisporus | 15.0 | 217 | J Bio Chem 279, 28202 (2004) |
| 1aky    | 47.7           | 7.50| Adenylate kinase | 1.96  | Saccharomyces Cerevisiae | 27.0 | 218 | Eur J Biochem 231, 405 (1995) |
| 1ank    | 51.8           | 7.20| Adenylate kinase | 2.00  | Escherichia Coli | 37.0 | 214 | J Bio Chem 266, 23654 (1991) |
| 1zip    | 74.8           | 7.20| Adenylate kinase | 1.85  | Bacillus Steathermophilus | 55.0 | 217 | Biochemistry 31, 3038 (1992) |
| 1oa3    | 49.2           | 8.00| Cell 12A     | 1.70    | Hypocrea Schweinitzii | 40.0 | 217 | Protein Sci 12, 848 (2003) |
| 1h8v    | 54.5           | 8.00| Cell 12A     | 1.90    | Trichoderma Reesei | 34.0 | 217 | Protein Sci 12, 2782 (2003) |
| 1oa4    | 66.8           | 8.00| Cell 12A     | 1.50    | Streptomyces sp. 1tag8 | 30.0 | 222 | Protein Sci 12, 848 (2003) |
| PDB  | Score | Mol. Weight | Protein | Reference |
|------|-------|-------------|---------|-----------|
| 1olr | 68.7  | 8.00        | Cell 12A| 1.20      |
| 1lcc | 70.4  | 7.20        | Cell 12A| 2.15      |
| 1esp | 53.8  | 7.00        | Cold Shock Protein | 2.50 |
| 1mjc | 57.0  | 7.00        | Cold Shock Protein | 2.00 |
| 1c8o | 76.9  | 7.00        | Cold Shock Protein | 1.17 |
| 1bu7 | 47.0  | 7.40        | Cytochrome p450 | 1.65 |
| 1oxa | 55.1  | 7.40        | Cytochrome p450 | 2.10 |
| 1akd | 56.0  | 7.40        | Cytochrome p450 | 1.80 |
| 1n9t | 88.5  | 7.40        | Cytochrome p450 | 1.80 |
| 1f4t | 91.0  | 7.40        | Cytochrome p450 | 1.93 |
| 1rgg | 49.3  | 7.00        | Ribonuclease | 1.40 |
| 9rnt | 50.9  | 7.00        | Ribonuclease | 1.50 |
| 1rhb | 53.2  | 5.50        | Ribonuclease | 2.00 |
| 1rbn | 63.6  | 7.10        | Ribonuclease | 2.10 |
| 2ehg | 102.0 | 7.00        | Ribonuclease | 1.60 |

**Protein Sci 12, 2872 (2003)**

**Biophys Chem 132, 229 (2002)**

**J Mol Bio 347, 1063 (2005)**

**Protein Sci 9, 387 (2000), Protein Sci 10, 2028 (2001)**

**J Mol Biol 319, 541 (2002)**

**Nat Struct Biol 7, 380 (2000)**

**J Biol Chem 278, 608 (2003)**

**J Biol Chem 278, 608 (2003)**

**J Biol Chem 278, 608 (2003)**

**J Biol Chem 278, 608 (2003)**

**J Biol Chem 278, 608 (2003)**

**Biochemistry 37, 16192 (1998), Protein Sci 8, 1843 (1999)**

**J Mol Biol 279, 31790 (2003), J Mol Biol 354, 967 (2005)**

**J Bio Chem 264, 11621 (1989), Biochemistry 33, 3312 (1994)**

**Eur J Biochem 220, 663 (1994)**

**Biochemistry 45, 10795 (2009)**

**PLoS ONE 6, e16226 (2011)**