# Superior Protein Thermophilicity Prediction With Protein Language Model Embeddings

Supplementary Materials

# Florian Haselbeck<sup>1,2</sup>, Maura John<sup>1,2</sup>, Yuqi Zhang<sup>1</sup>, Jonathan Pirnay<sup>1,2</sup>, Juan Pablo Fuenzalida-Werner<sup>3</sup>, Rubén D. Costa<sup>3</sup> and Dominik G. Grimm<sup>1,2,4,\*</sup>

<sup>1</sup>Technical University of Munich, Campus Straubing for Biotechnology and Sustainability, Bioinformatics, 94315 Straubing, Germany, <sup>2</sup>Weihenstephan-Triesdorf University of Applied Sciences, Bioinformatics, 94315 Straubing, Germany, <sup>3</sup>Technical University of Munich, Campus Straubing for Biotechnology and Sustainability, Chair of Biogenic Functional Materials, 94315 Straubing, Germany, <sup>4</sup>Technical University of Munich, TUM School of Computation, Information and Technology (CIT), 85748 Garching, Germany \*Corresponding author: dominik.grimm@hswt.de

Received YYYY-MM-DD; Revised YYYY-MM-DD; Accepted YYYY-MM-DD

# SUPPLEMENTARY METHODS

# **Basic Descriptors**

The *weight* descriptor is defined as the difference of the sum of the molecular weight of each amino acid in the protein sequence and the molecular weight of a water molecule for each peptide bound in grams per mol. The *charge* descriptor consists of three features: the net charge, the number of amino acids with positively charged side chains and the number of amino acids carrying negatively charged side chains. Here the amino acids H, K and R are specified to carry positive charges; D and E to carry negative charges; and the remaining 15 are defined as uncharged. For *polarity* we used the number of amino acids with polar side chains and the number of amino acids with non-polar side chains as features, where A, F, G, I, L, M, P, V and W were considered to be non-polar and the remaining 11 residues to be polar. Additionally, we considered the number of *aromatic* amino acids (F, W, Y) in a peptide. For the *mean hydrophobicity* of a peptide we used the hydrophobicity scale PRAM900101 (1). The *mean vdW volume* is defined to be the sum of the normalized van der Waals volume of the residues averaged over the number of amino acids in the peptide.

# **Amino Acid Composition**

The *amino acid composition* (AAC) describes the frequencies of each of the 20 residues in a protein sequence. Thus, for a protein sequence  $\rho$  and an amino acid  $a \in \{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y\}$  the AAC feature is given as

$$AAC(\rho, a) = \frac{l(a)}{l_{\rho}} \in \mathbb{R}$$
(1)

where  $l_{\rho}$  is the length of the sequence  $\rho$  and l(a) denotes the number of residues of type a in  $\rho$ .

#### **Dipeptide Composition**

The *dipeptide composition* (DPC) measures the frequencies of each of the 400 possible contiguous dipeptides in a protein sequence. Hence, for a protein sequence  $\rho$  and two amino acids  $a, b \in \{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y\}$  the DPC feature is defined as

$$\mathrm{DPC}(\rho,(a,b)) = \frac{l(a,b)}{l_{\rho}-1} \in \mathbb{R}$$
<sup>(2)</sup>

where  $l_{\rho}$  denotes again the length of  $\rho$  and l(a,b) is the number of contiguous dipeptides of type (a,b) in  $\rho$ .

# Physicochemical Composition

These features describe the distribution patterns of specific physicochemical and structural properties along the protein sequence. We considered seven of those properties: hydrophobicity, where we use the scale PRAM900101 (1), normalized van der Waals

© 2023 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/ by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

volume, polarity, polarizability, charge, secondary structure and solvent accessibility. For each of these properties, the 20 amino acids are distributed among three categories based on (2, 3). The different groups are listed in Table S1.

Table S1. Distribution of amino acids among different physicochemical properties: The first column gives the seven different physicochemical attributes. In the remaining three columns the amino acids are given in parenthesis, sorted into different groups based on their properties.

Attribute		Categories	
Hydrophobicity	Polar (RKEDQN)	Neutral (GASTPHY)	Hydrophobicity (CLVIMFW)
Normalized vdW Volume	0 - 2.78 (GASCTPD)	2.95 - 4.0 (NVEQIL)	4.43 - 8.08 (MHKFRYW)
Polarity	4.9 - 6.2 (LIFWCMVY)	8.0 - 9.2 (PATGS)	10.4 - 13.0 (HQRKNED)
Polarizability	0 - 0.108 (GASDT)	0.128 - 0.186 (CPNVEQIL)	0.219 - 0.409 (KMHFRYW)
Charge	Positive (KR)	Neutral (ANCQGHILMFPSTWYV)	Negative (DE)
Secondary Structure	Helix (EALMQKRH)	Strand (VIYCWFT)	Coil (GNPSD)
Solvent Accessibility	Buried (ALFCGIVW)	Exposed (PKQEND)	Intermediate (MPSTHY)

*Composition* The *composition* (CTDC) descriptor measures for each property and each category in Table S1 the corresponding fraction of residues in the sequence, leading to 21 features in total. Hence, for a protein sequence  $\rho$ , a physicochemical property q and a corresponding group r the CTDC feature is defined as

$$\operatorname{CTDC}(\rho, q, r) = \frac{l(r \mid q)}{l_{\rho}} \in \mathbb{R}$$
(3)

where l(r|q) denotes the number of residues of group r and property q in the sequence.

*Transition* The *transition* (CTDT) descriptor measures for each property in Table S1 the fraction of dipeptides in the sequence where the two contiguous amino acids belong to two different groups, e.g. where a residue of group 1 is followed by a residue of group 2 or the other way round. This leads to 21 features in total. For a protein sequence  $\rho$ , a physicochemical property q and two corresponding groups r and s the CTDT feature is defined as

$$\operatorname{CTDT}(\rho, q, \{r, s\}) = \frac{l(\{r, s\} \mid q)}{l_{\rho} - 1} \in \mathbb{R}$$

$$\tag{4}$$

where  $l(\{r,s\}|q)$  is the number of dipeptides (a,b) in  $\rho$ , where either a belongs in group r and b belongs in group s or a belongs in group s and b belongs in group r.

*Distribution* The *distribution* (CTDD) descriptor consists of five values for each property and each category in Table S1, leading to 105 features in total. These values are given by the fractions of the entire sequence where the first amino acid of the corresponding group is located and where 25 %, 50 %, 75 % and 100 % of residues within a group are contained. For a protein sequence  $\rho$ , a physicochemical property q, a corresponding group r and a characteristic c (corresponding to the first amino acid or to 25 %, 50 %, 75 % or 100 % of residues) the CTDD feature is defined as

$$\operatorname{CTDD}(\rho, q, r, c) = \frac{i_{(r|q)}^{c}}{l_{\rho}} \in \mathbb{R}$$
(5)

with  $i_{(r|q)}^c$  denoting the position in the protein sequence where the portion of residues of property q and group r corresponding to c occurred.

#### **Pseudo Amino Acid Composition**

The *pseudo amino acid composition* (PAAC) includes information on the amino acid composition as well as additional discrete values that reflect the sequence order effect (4). Given a protein sequence  $\rho$  of length  $l_{\rho}$ , let  $a_i$  denote the  $i^{\text{th}}$  amino acid in  $\rho$  for  $i \in \{1, ..., l_{\rho}\}$ . Furthermore, denote by  $\vartheta_1(a_i)$ ,  $\vartheta_2(a_i)$  and  $\vartheta_3(a_i)$  the hydrophobicity value, hydrophilicity value and side-chain mass of  $a_i$  after a standard conversion as described in (4), respectively. Define the correlation function as

$$\Psi(a_i, a_{i'}) = \frac{1}{3} \left( (\vartheta_1(a_{i'}) - \vartheta_1(a_i))^2 + (\vartheta_2(a_{i'}) - \vartheta_2(a_i))^2 + (\vartheta_3(a_{i'}) - \vartheta_3(a_i))^2 \right)$$
(6)

Then for  $\lambda \in \mathbb{N}$  with  $\lambda < l_p$ , the sequence order effect can be approximated via the following correlation factors:

$$\tau_1 = \frac{1}{l_{\rho} - 1} \sum_{i=1}^{l_{\rho} - 1} \Psi(a_i, a_{i+1}) \tag{7}$$

$$\tau_2 = \frac{1}{l_{\rho} - 2} \sum_{i=1}^{l_{\rho} - 2} \Psi(a_i, a_{i+2}) \tag{8}$$

$$\tau_{\lambda} = \frac{1}{l_{\rho} - \lambda} \sum_{i=1}^{l_{\rho} - \lambda} \Psi(a_i, a_{i+\lambda})$$
(10)

(11)

where  $\tau_k$  is called the  $k^{\text{th}}$ -tier correlation factor for  $k \in \{1, ..., \lambda\}$ . Now, without loss of generality, denote the 20 amino acids by their index  $j \in \{1, ..., 20\}$ , when sorted alphabetically according to their single-letter codes. Then the PAAC features are defined as

$$PAAC(\rho) = (\rho_u)_{u \in \{1, \dots, 20+\lambda\}} \in \mathbb{R}^{20+\lambda}$$
(12)

with

$$\rho_{u} = \begin{cases} \frac{f_{u}}{\sum_{j=1}^{20} f_{j} + \omega \sum_{k=1}^{\lambda} \tau_{k}}, & \text{if } 1 \le u \le 20\\ \frac{\omega \tau_{u} - 20}{\sum_{j=1}^{20} f_{j} + \omega \sum_{k=1}^{\lambda} \tau_{k}}, & \text{if } 20 + 1 \le u \le 20 + \lambda \end{cases}$$
(13)

where  $f_j$  for  $j \in \{1, ..., 20\}$ , denotes the normalized occurrence frequency of amino acid j in the peptide  $\rho$  and  $\omega$  is a weight factor. In our study we used  $\omega = 0.05$  and  $\lambda = 3$ .

# SUPPLEMENTARY TABLES

Table S2. List of thermophilic species and number of proteins in our full dataset before filtering for sequence length and CD-HIT. Species marked with an asterisk were used for our independent test set.

Actidiums cambination of the second	Species	Number of proteins	Species	Number of proteins
Acidans ambivalens *12Rhodohermus marinus *9Acidaliprofundan boonei *1Rubobacter sylonophius2Acceptorum pernix50Saccharablus solfanaricus195Alicyclobacillus acidocaldarius5Spirochaeta thermophila *2Aquifes acidocaldarius51Sulphylohterms marinus2Aquifes acidocaldarius61Sattaphylophaeta tokodaii52Caldanaerobacter subpranews6Sulphylophaeta tokodaii52Caldanaerobacter subpranews6Symbiobacterium thermophilun3Caldalaerobacter subpranews6Symbiobacterium thermophilun3Caldalaerobacter subpranews1Thermoanaerobacter italicus *1Deferribacter deulfuricans *1Thermoanaerobacter italicus *1Deferribacter deulfuricans *3Thermoanaerobacter italicus *1Deferribacter deulfuricans *1Thermoanaerobacter italicus *1Geobacillus kustophilus *3Thermoanaerobacter italicus *1Geobacillus kustophilus *3Thermoacoccus finciolans *2Hydrogenobacter thrum politicurans *1Thermoacoccus italicians *1Internooccus substraines1Thermoacoccus finciolans *2Hydrogenobacter thrum politicus *1Thermoacoccus italicians *1Internooccus substraines1Thermoacoccus italicians *1Internooccus substraines1Thermoacoccus italicians *1Internooccus substraines1 <td< td=""><td>Acetivibrio thermocellus</td><td>33</td><td>Pyrococcus woesei</td><td>1</td></td<>	Acetivibrio thermocellus	33	Pyrococcus woesei	1
Actiduity1Rubrobacter xylanophilus2Acregorum pernix50Saccharolbus solidarcias195Alexyclobacillus acidocaldarius5Spirocharet thermophila +2Aquifex acidocaldarius5Spirocharet thermophila +2Aquifex pyrophilus *3Sulfolobus acidocaldarius87Archaeoglobus fulgidus14Sulfolobus acidocaldarius87Archaeoglobus profundus *8Sulfolobus acidocaldarius87Caldadactibaceter subterraneus6Symbobacterium thermophilum3Caldadactobaceter subterraneus6Symbobacterium thermophilum3Caldacter subterraneus6Symbobacterium thermophilum3Deferribacter desulfuricans *1Thermoanaerobacter induces *1Deferribacter desulfuricans *1Thermoanaerobacter induces *4Fervidabcacterium pennivorans *1Thermoanaerobacterium schnolyticum *4Geobacillus utermodeniurificans11Thermoaccocus barnophilus *1Geobacillus utermodeniurificans *1Thermoaccocus loginoloans *1Inguiecocus houghalas *1Thermoaccocus loginoloans *1Inguiecocus houghalas *1Thermoaccocus loginoloans *1Inguiecocus houghalas *1Thermoaccocus loginolans *1Inguiecocus houghalas *1Thermoaccocus loginolans *1Inguiecocus houghalas *1Thermoaccocus loginolans *1Inguiecocus houghalas *1Thermoaccocus	Acidianus ambivalens *	12	Rhodothermus marinus *	9
Arenzymm pernix50Saccharolabus solfaaricus195Alkryclobactikus acidocaldarius2Aquifex acolicus111Staphylthermus marinus2Aquifex pyrophilus *3Sulfolobus acidocaldarius87Arthacoglobus fulgidus154Sulfolobus cididarius6Archacoglobus profindus *8Sulfolobus cididarius6Caldalachibus profindus *8Sulfolobus cididarius52Caldalachibus profindus *6Symbiobaccerius thermophilum3Caldicellubority por saccharopsicus *1Thermoanerobacter ethanolicus *1Deferrihacter desulfaricus *1Thermoanerobacter infusion1Dictycyglomus thermophilum3Thermoanerobacter infusion1Dictycyglomus thermophilum *38Thermoanerobacterium pennivorans *1Geobacillus kustophilus *38Thermoaccus barphilus *2Geoglobus activorans *1Thermoaccus comphilus *1Intermooccus barphilus *3Thermoaccus comphilus *1Intermooccus barphilus *1Thermooccus functions *1Intermooccus barphilus *1Thermooccus barphilus *1Intermooteus barphicus *1Thermoot	Aciduliprofundum boonei *	1	Rubrobacter xylanophilus	2
Alicyclobacillus acidocaldarius5Spirochaeta thermophila *2Aquifex pyrophilus *11Staplothermus marinus2Aquifex pyrophilus *3Sulfolobus acidocaldarius87Archaeoglobus profundus *8Sulfolobus sciendicus *1Caldakalibacillus thermarum *1Sulforbabeccare sultigit *1Caldakalibacitlus stermarum *2Thermoonaerobaccer influences *1Caldacetocare suberraneus6Symboloacceinu thermophilum3Caldicellulasiruptor saccharolyticus *2Thermoonaerobaccer influences *1Deferribacter desulfuricans *1Thermoonaerobacter influences *1Deferribacter desulfuricans *1Thermoonaerobacterium thermophilum *4Geobacillus kaustophilus *38Thermooaerobacterium stennyophilus *2Geobacillus kaustophilus *1Thermooaccus famicalans *1Hydregenbus aceitoronus *1Thermooaccus famicalans *2Hydregenbus aceitoronus *1Thermooaccus famicalans *1Hydregenbus aceitoronus	Aeropyrum pernix	50	Saccharolobus solfataricus	195
Aquifex acolicus111Staphylohermus marinus2Aquifex pyrophilus *3Sulfolohus saidacadarirus87Archaeoglohus profindus *85ulfolohus silandicus *6Archaeoglohus profindus *8Sulfolohus concatalityii *1Caldakalibacilis hermarum *1Sulfolohus concatilityii *1Caldakalibacilis hermarum *6Symbiobaccerium thermophilum3Caldicellus inprove acchoropyticus *1Thermooanerobacter ethanolicus *1Caldiculus inprove acchoropyticus *1Thermooanerobacter italicus *1Deferribacter desulfuricans *1Thermooanerobacter inscharolyticum *4Fervidobacterium pennivorans *1Thermooacuro bacterium sennosulfurigenes *4Geobacillus kastophilus *3Thermooacuro bacterium sennosulfurigenes *1Geobacillus kastophilus *9Thermooacurs deposition *1Intermooccus gamatolerans *1Thermooccus gamatolerans *1Iprincoccus hospitalis *1Thermooccus gamatolerans *1Iprincoccus hospitalis *1Thermooccus gamatolerans *1Iprincoccus hospitalis *1Thermooccus gamatolerans *1Iprincoccus hospitalis *1Thermooccus funcioans *1Iprincoccus hospitalis *1Thermooccus gamatolerans *1Iprincoccus hospitalis *1Thermooccus gamatolerans *1Iprincoccus hospitalis *1Thermooccus funcioans *1Iprincoccus ho	Alicyclobacillus acidocaldarius	5	Spirochaeta thermophila *	2
Applie's prophilus *3Sulfolobus acidocalarius87Archaceglobus profundus *154Sulfolobus caidocalians6Archaceglobus profundus *1Sulforbos caccus zilligii *1Caldalacillus clilus thermanum *1Sulfurisphaera tokodaii3Caldacector suberraneus6Symbiobacterium thermophilum3Caldicellulosiruptor saccharolyticus *11Deferibacter desulfuricans *1Thermoanaerobacter indicus *1Deferibacter desulfuricans *1Thermoanaerobacter indicus *4Deferibacter desulfuricans *1Thermoanaerobacter indicus *4Deferibacter desulfuricans *1Thermoanaerobacterium thermosulfurigenes *4Geobacillus kaustophilus3Thermococcus barophilus *2Geobacillus cativorans *1Thermococcus generatives *1Hydrogenobacter thermophilus *9Thermococcus generatives *1Iguicoccus hospitalis *1Thermococcus korophilus *1Iguisphaera agergans *1Thermococcus korophines *1Iguisphaera genua1Thermococcus korophines *1Metallosphaera prune1Thermococcus korophines *1Metallosphaera prune1Thermococcus subrohilus *1Metallosphaera sedula *1Thermococcus subrohilus *1Metallosphaera prune1Thermococcus subrohilus *1Metallosphaera sedula *1Thermococcus subrohilus *1	Aquifex aeolicus	111	Staphylothermus marinus	2
Archaceglobus fulgidus154Sulfolobus islandicus6Archaceglobus profundus *8Sulforuboccecus silligit *1Caldalkalibacillus thermarum *1Sulfurisphaera tokodait52Caldancerbacter subterraneus6Symbiobacterium thermophilum3Caldicelluosingtor saccharolyticus *2Thermoanaerobacter ethanolicus *1Carboxydothermus hydrogenoformans4Thermoanaerobacter kivai1Dictyoglomus thermophilum3Thermoanaerobacter kivai1Dictyoglomus thermophilum *3Thermoanaerobacter kivai1Geobacillus kaustophilus *1Thermoanaerobacter intermosulfurigenes *4Geobacillus kaustophilus *3Thermococcus cleftensis *2Geobacillus kaustophilus *3Thermococcus cleftensis *2If procescus harphilus *3Thermococcus gammatolerans *1If procescus harphilus *3Thermococcus infuncionars *2If ydrogenobacter thermaphilus *3Thermococcus gammatolerans *1Ignicoccus hapitalis *1Thermococcus infuncionars *2Ignicoccus hapitalis *1Thermococcus infuncionars *1Ignicoccus hapitalis *1Thermococcus infuncionars *1Ignicocus hapitalis *1Thermococcus infuncionars *1Ignicocus hapitalis *1Thermococcus infuncionars *1Ignicocus hapitalis *1Thermococcus infuncionars *1Ignicocus hapitalis *1<	Aquifex pyrophilus *	3	Sulfolobus acidocaldarius	87
Archaeoglobus profinadas *8Sulfgriphobacccus III[i]*1Caldalacibacillus thermarum *1Sulfurisphaera tokodaii52Caldacellulosiruptor succharolyticus *2Thermoanaerobacter influence1Carboxydohermus hydrogenoformans4Thermoanaerobacter influence1Deferribacter desulfuricans *1Thermoanaerobacter influences *4Deferribacter desulfuricans *1Thermoanaerobacter influence4Fervidobacterium pentivorans *1Thermoanaerobacter influences4Geobacillus kaustophilus3Thermoanaerobacter influences2Geobacillus kentophilus *38Thermococcus barophilus *2Geobacillus kentophilus *1Thermococcus funcionans *2Hydrogenobacter thermophilus *9Thermococcus funcionans *1Ignisoccus hosphilais *1Thermococcus funcionans *1Ignisphera agergans *1Thermococcus fordicans *1Ignisphera agergans *1Thermococcus inductans *1Ignisphera agergans *1Thermococcus inductans *1Metallosphaera cuprina1Thermococcus inductans *1Metallosphaera sedula *12Thermococcus inductans *1Metallosphaera sedula *12Thermococcus inductans *1Metallosphaera sedula *12Thermococcus inductans *1Metallosphaera sedula *12Thermococcus inductans *1Metallosphaera sedula *12Thermococc	Archaeoglobus fulgidus	154	Sulfolobus islandicus	6
Caldialithalibacillus thermarum *1Sulfurispharer ackadii52Caldacerbaciter subterraneus6Symbiobacterium thermophilum3Caldacellusiruptor saccharolyticus *11Carboxydohermus hydrogenoformans4Thermoanaerobacter indicus *1Deferribacter desulfuricans *1Thermoanaerobacter indicus *1Dictyoglonus thermophilum3Thermoanaerobacter kivai1Dictyoglonus thermophilus *1Thermoanaerobacterium saccharolyticum *4Geobacillus kaustophilus *38Thermoanaerobacterium saccharolyticum *4Geobacillus kaustophilus *38Thermoacoccus barophilus *1Geolobacillus thermodenitrificans11Thermoacoccus functolorus *1Geolobus activorans *1Thermoacoccus functolorus *1Ignicoccus hospitalis *9Thermoacoccus gorgonarius1Ignicoccus hydrothermalis1Thermoacoccus functolorus *1Ignicoccus hydrothermalis1Thermoacoccus functorus *1Ignicoccus hydrothermalis1Thermoacoccus functorus *1Ignicoccus hydrothermalis1Thermoacocus briftmais *19Metallosphaera acgurgans *1Thermoacocus profinadus *1Idenaocidococcus iferus *1Thermoacocus profinadus *1Idenaocidococcus iferus *1Thermoacocus profinadus *1Idenaocidococcus iferus *1Thermoacocus profinadus *1Methanocaldococcus iferus *1 <td>Archaeoglobus profundus *</td> <td>8</td> <td>Sulfophobococcus zilligii *</td> <td>1</td>	Archaeoglobus profundus *	8	Sulfophobococcus zilligii *	1
Caldacenerbacter subterraneus6Symbiobacterium thermophilum3Caldicellulosiruptor saccharolyticus *11Carboxydohernus hydrogenoformans4Thermoanaerobacter italicus *1Deferribacter desulfiricans *1Thermoanaerobacter italicus *1Dictyoglonus thermophilum3Thermoanaerobacterium saccharolyticum *4Ferridacter desulfiricans *1Thermoanaerobacterium thermosulfurigenes *4Geobacillus kaustophilus *38Thermoanaerobacterium thermosulfurigenes *1Geobacillus thermophilus *3Thermoanaerobacterium thermosulfurigenes *1Geobacillus thermophilus *1Thermoaccus suprophilus *2Hydrogenobacter thermophilus *9Thermoaccus suprophilus *1Ignisphare agergans *1Thermoaccus gorgonarius1Ignisphare agergans *1Thermoaccus barophilus *1Ignispharen agergans *1Thermoaccus barophilus *1Metallosphaera cruprina1Thermoaccus barophilus *1Metallosphaera gergans *1Thermoaccus barophilus *1Metallosphaera sedula *1Thermoacus undrineus *6Metallosphaera sedula *1Thermoacus undrineus *1Metallosphaera sedula *2Thermooleus undrineus *1Metallosphaera sedula *1Thermooleus undrineus *1Metallosphaera sedula *2Thermooleus undrineus *1Metallosphaera sedula *2Thermoole	Caldalkalibacillus thermarum *	1	Sulfurisphaera tokodaii	52
Caldicellulosiruptor saccharolyticus *1Carboxydothermus hydrogenoformans4Thermoanaerobacter italicus *1Deferribacter desulfuricans *1Thermoanaerobacter italicus *1Dictyoglomus thermophilum3Thermoanaerobacterium saccharolyticum *4Fervidobacterium penivorans *1Thermoanaerobacterium saccharolyticum *4Geobacillus kaustophilus *38Thermococcus barophilus *2Geobacillus thermodenitrificans11Thermococcus funicolans *2Hydrogenobacter thermophilus *9Thermococcus funicolans *2Hydrogenobacter thermophilus *9Thermococcus forgonarius1Ignicoccus hospitalis *1Thermococcus forgonarius1Ignicoccus hospitalis *1Thermococcus fordakarensis *138Metallosphaera agergeans *1Thermococcus fordakarensis *1Metallosphaera cuprina1Thermococcus fordakarensis *1Metallosphaera sedula *2Thermococcus fordakarensis *1Metanocaldococcus forvers *1Thermococcus fordakarensis *1Methanocaldococcus forvers *1Thermococcus fordakarensis *2Methanocald	Caldanaerobacter subterraneus	6	Symbiobacterium thermophilum	3
Carboxydothermus hydrogenoformans4Thermoanaerobacter italicus *1Deferribacter desulfuricans *1Thermoanaerobacter ikviui1Dictyoglonus thermophilum3Thermoanaerobacterium saccharolyticum *4Ferridobacterium pennivorans *1Thermoanaerobacterium saccharolyticum *4Geobacillus kaustophilus *38Thermococcus barophilus *2Geobacillus thermodenitrificans11Thermococcus functionals *2Hydrogenobacter thermophilus *9Thermococcus gammatolerans *1Hyperthermus butylicus *3Thermococcus gammatolerans *1Ignisophaera ageregans *1Thermococcus functionals *2Hydrogenobacter ageregans *1Thermococcus functionals *1Ignisophaera ageregans *1Thermococcus functionals *1Ignisophaera ageregans *1Thermococcus functionals *1Metallosphaera aprina1Thermococcus profundus *1Metallosphaera ageregans *1Thermococcus profundus *1Metallosphaera sedula *12Thermococcus giligit *2Methanocaldococcus infernus *1Thermococcus gama olocanium3Methanoblermobacter marburgensis5Thermoplasma acidephilus *1Methanocaldococcus infernus *1Thermococcus giligit *2Methanocaldococcus infernus *1Thermococcus giligit *2Methanoblermobacter marburgensis5Thermoplasma acidephilum3Metha	Caldicellulosiruptor saccharolyticus *	2	Thermoanaerobacter ethanolicus *	1
Deferribacter desulfuricans *1Thermoanaerobacter kivui1Dictyoglomus thermophilum3Thermoanaerobacterium saccharolyticum *4Geobacillus kaustophilus *1Thermoanaerobacterium thermosulfurigenes *4Geobacillus kaustophilus *38Thermococcus cleftensis *2Geobacillus kaustophilus *1Thermococcus cleftensis *2Geobacillus kaustophilus *9Thermococcus gomatolerans *1Hydrogenobacter thermophilus *9Thermococcus gorgonarius1Ignicoccus hospitalis *1Thermococcus gorgonarius1Ignicoccus hospitalis *1Thermococcus gorgonarius1Ignisphaera aggregans *1Thermococcus phydrothermalis1Metallosphaera cuprina1Thermococcus profundus *6Metallosphaera cuprina1Thermococcus profundus *1Metallosphaera sedula *1Thermococcus profundus *1Methanocaldococcus fervens *1Thermococcus profundus *1Methanocaldococcus fervens *1Thermolosuifobacterium geofontis *1Methanocaldococcus fervens *16Thermoplasma acidaphilum31Methanothermobacter thermativersis59Thermoosufidibacter cueani *1Methanothermobacter thermativersis6Thermosyladificater cueani *2Methanothermobacter thermativersis59Thermopatisticati *2Methanothermobacter thermativersis6Thermosyladificateri cukaii *2M	Carboxydothermus hydrogenoformans	4	Thermoanaerobacter italicus *	1
Dictyoglomus thermophilum3Thermoanaerobacterium saccharolyticum *4Fervidobacterium pennivorans *1Thermoanaerobacterium thermosulfurigenes *4Geobacillus kuustophilus *38Thermococcus barophilus *1Geobacillus thermodenitrificans11Thermococcus functionas *2Geoglobus acetivorans *1Thermococcus ginnicolans *2Hydrogenobacter thermophilus *9Thermococcus gorgonarius1Hyperthermus butylicus *3Thermococcus gorgonarius1Ignicocus hospitalis *1Thermococcus hydrothermalis1Ignisphera aggregans *1Thermococcus hydrothermalis1Metallosphaera cuprina1Thermococcus onnurineus *6Metallosphaera prunae1Thermococcus profindus *1Metallosphaera sedula *2Thermococcus profindus *1Methanocaldococcus inferus *2Thermologung acidophilum31Methanocaldococcus inferus *2Thermoplasma volcanium1Methanocaldococcus inferus *117Thermoseitylinitacter oceani *1Methanothermobacter marburgensis59Thermoosphilum albam1Methanothermobacter thermolitorophicus *6Thermosiph adricus2Methanothermobacter thermolitorophicus *1Thermoseitylinitacter oceani *1Methanothermobacter thermolitorophicus *1Thermoseitylinitacter oceani *1Methanothermobacter thermolitorophicus *2Thermolitacter akaii *	Deferribacter desulfuricans *	1	Thermoanaerobacter kivui	1
Fervidobacterium pennivorans *1Thermoanaerobacterium thermosulfurigenes *4Geobacillus kaustophilus *38Thermococcus barophilus *1Geobacillus kaustophilus *1Thermococcus barophilus *2Geobaciter thermophilus *9Thermococcus gamnatolerans *1Hydrogenobacter thermophilus *9Thermococcus gamnatolerans *1Ignicoccus hospitalis *1Thermococcus gamnatolerans *1Ignicoccus hospitalis *5Thermococcus kodkarensis *18Ignicoccus hospitalis *1Thermococcus kodkarensis *19Metallosphaera aggregans *1Thermococcus brotilis *19Metallosphaera cuprina1Thermococcus profundus *1Metallosphaera sedula *12Thermococcus profundus *1Methanocaldococcus iferrus *2Thermococcus profundus *1Methanocaldococcus infernus *2Thermoplilum album1Methanocaldococcus infernus *16Thermoplilum album1Methanochermobacter marburgensis59Thermospinolitanta1Methanothermobacter internuotorphicus *6Thermospinolitanta1Methanothermobacter internuotorphicus *1Thermospinolitata *2Methanothermobacter internuotorphicus *6Thermospinolitata *2Methanocaldococcus internuitis117Thermospinolitata *2Methanothermobacter internuotorphicus *111Methanothermobacter internuotorphicus * <td< td=""><td>Dictyoglomus thermophilum</td><td>3</td><td>Thermoanaerobacterium saccharolyticum *</td><td>4</td></td<>	Dictyoglomus thermophilum	3	Thermoanaerobacterium saccharolyticum *	4
Geobacillus kaustophilus *38Thermococcus barophilus *1Geobacillus thermodenitrificans11Thermococcus barophilus *2Geoglobus acctivorans *1Thermococcus funicolans *2Hydrogenobacter thermophilus *9Thermococcus gammatolerans *1Hyperhermus buspitalis *5Thermococcus gogonarius1Ignicocus hospitalis *1Thermococcus hospitalis *1Ignisphaera aggregans *1Thermococcus lioralis *18Metallosphaera cuprina1Thermococcus lioralis *19Metallosphaera runae1Thermococcus profundus *1Metallosphaera sedula *12Thermococcus profundus *1Metallosphaera sedula *12Thermococcus gilligii *2Methanocaldococcus infernus *2Thermologenui me sedontis *1Methanocaldococcus infernus *2Thermologium album1Methanocaldococcus infernus *311Methanocaldococcus infernus *6Thermoplasma ocionium13Methanochermobacter marburgensis59Thermoplasma ocionium13Methanothermobacter thermalitorophicus *6Thermostipol gircianus2Methanothermobacter thermolithorophicus *1Thermostipol gircianus2Methanothermobacter thermalitorophicus *6Thermostipol gircianus2Methanothermobacter thermalitorophicus *1Thermostipol gircianus2Methanothermobacter thermolithorophicus *11	Fervidobacterium pennivorans *	1	Thermoanaerobacterium thermosulfurigenes *	4
Geobacillus thermodenitrificans11Thermococcus cleftensis *2Geoglobus acetivorans *1Thermococcus funicolans *2Hydrogenobacter thermophilus *9Thermococcus gammatolerans *1Hyperthermus butylicus *3Thermococcus gorgonarius1Ignicoccus hospitalis *5Thermococcus hydrothermalis1Ignisphaera aggregans *1Thermococcus hydrothermalis1Ignisphaera aggregans *1Thermococcus hydrothermalis1Metallosphaera cuprina1Thermococcus nutrineus *6Metallosphaera ruprina1Thermococcus profindus *1Metallosphaera sedula *12Thermococcus profindus *1Methanocaldococcus farens *1Thermoleophilum album1Methanocaldococcus infernus *2Thermoleophilum album31Methanocaldococcus infernus *16Thermoproteus teax *19Methanochermobacter hermautorophicus117Thermosedininibacter oceni *1Methanothermobacter hermautorophicus *6Thermoproteus teax *2Methanothermobacter hermolyticus *6Thermosedininibacter oceni *1Methanothermobacter hermolyticus *17Thermocage periodina *2Methanothermobacter hermolyticus *1Thermocage periodina *2Methanothermobacter hermolyticus *1Thermocage periodina *2Methanothermobacter hermolyticus *1Thermocage periodina *2Methanothermobacter hermolyt	Geobacillus kaustophilus *	38	Thermococcus barophilus *	1
Geoglobus acetivorans *1Thermococcus funicolans *2Hydrogenobacter thermophilus *9Thermococcus gammanolerans *1Hyperthermus butylicus *3Thermococcus gorgonarius1Ignicoccus hospitalis *5Thermococcus hydrothermalis1Ignicoccus hospitalis *1Thermococcus hydrothermalis1Ignicoccus hospitalis *1Thermococcus hydrothermalis1Ignicoccus hospitalis *1Thermococcus hydrothermalis1Metallosphaera aggregans *1Thermococcus onnurineus *6Metallosphaera prunae1Thermococcus onnurineus *1Metallosphaera agunae1Thermococcus profundus *1Methanocaldococcus fervens *1Thermodesulfbacterium geofontis *1Methanocaldococcus infernus *2Thermolesphilum album1Methanocaldococcus jannaschii315Thermoplasma acidophilum31Methanothermobacter marburgensis59Thermoproteus tenax *19Methanothermobacter thermautotrophicus117Thermosediminibacter oceani *1Methanothermosoccus thermolithotrophicus *6Thermosolig neapolitana20Nanaarchaeun equitans3Thermotoga maritima260Nanaarchaeun equitans1Thermotoga maritima260Nanaarchaeun equitans3Thermotoga neapolitana10Parageobacillus thermoglucisaisus *1Thermotoga neapolitana10Parageobacillus utermoglusisaisus *10	Geobacillus thermodenitrificans	11	Thermococcus cleftensis *	2
Hydrogenobacter thermophilus *9Thermococcus gammatolerans *1Hyperhermus butylicus *3Thermococcus gorgonarius1Ignicoccus hospitalis *5Thermococcus hydrothermalis1Ignicoccus hospitalis *1Thermococcus kodakarensis *138Meiothermus ruber *1Thermococcus litoralis *19Metallosphaera cuprina1Thermococcus profundus *6Metallosphaera prunae1Thermococcus profundus *1Metallosphaera sedula *12Thermococcus grofinatis *1Metallosphaera sedula *12Thermococcus grofinatis *1Methanocaldococcus infernus *2Thermolecophilum album1Methanocaldococcus infernus *2Thermolesphilum album1Methanocaldococcus infernus *16Thermoplasma volcanium13Methanothermobacter marburgensis59Thermoplasma volcanium1Methanothermobacter thermautotrophicus117Thermosiph africanus2Methanothermus fervidus *8Thermotoga maritina26Nanoarchaeum equitans3Thermotoga maritina26Nanoarchaeum equitans *10Thermotoga neapolitans1Probaculum actifiqintis *10Thermosy anonophila2Pyrobaculum actifiqintis *10Thermosy anonophila2Pyrobaculum actifiqintis *10Thermosy anonophila2Pyrobaculum actifiqintis *10Thermosy anonophila2Pyrobaculum actifig	Geoglobus acetivorans *	1	Thermococcus fumicolans *	2
Hyperhermus butylicus *3Thermococcus gorgonarius1Ignicoccus hospitalis *5Thermococcus hydrohermalis1Ignisphaera aggregans *1Thermococcus hydrohermalis *138Meiothermus ruber *1Thermococcus kodakarensis *199Metallosphaera cuprina1Thermococcus profundus *6Metallosphaera prunae1Thermococcus profundus *1Metallosphaera sedula *12Thermococcus profundus *1Methanocaldococcus fervens *1Thermococcus gorgonarium1Methanocaldococcus fervens *1Thermococcus gorgonarium1Methanocaldococcus fervens *1Thermoleophilum album1Methanocaldococcus infernus *2Thermoleophilum album1Methanochermobacter narburgensis59Thermoplasma acidophilum13Methanothermobacter thermautotrophicus117Thermosediminibacter oceani *1Methanothermobacter thermautotrophicus *6Thermosediminibacter oceani *2Methanothermus fervidus *8Thermosogia paritima260Nanoarchaeum equitans3Thermologa nagatitima2Persephonella marina *1Thermologa nagatitima2Pyrobaculum calidifichtis *10Thermus gauaticus *2Pyrobaculum calidifichtis *10Thermus gauaticus *3Pyrobaculum calidifichtis *10Thermus gauaticus *3Pyrobaculum calidifichtis *10Thermus gauaticus *3<	Hydrogenobacter thermophilus *	9	Thermococcus gammatolerans *	1
Ignicoccus hospitalis *5Thermococcus hydrothermalis1Ignicoccus hospitalis *1Thermococcus kodakarensis *138Meiothermus ruber *1Thermococcus kodakarensis *19Metallosphaera cyrina1Thermococcus onnurineus *6Metallosphaera prunae1Thermococcus onnurineus *1Metallosphaera sedula *12Thermococcus profundus *1Metallosphaera sedula *1Thermococcus profundus *1Methanocaldococcus forenus *1Thermolesuffobacterium geofontis *1Methanocaldococcus infernus *2Thermolesuffobacterium geofontis *1Methanocaldococcus infernus *16Thermoplasma acidophilum album1Methanochdercoccus infernus *16Thermoplasma volcanium13Methanothermobacter marburgensis59Thermoproteus tenax *19Methanothermobacter thermautorophicus *6Thermospino africanus2Methanothermobacter thermautorophicus *6Thermospino africanus2Methanothermos fervidus *2Thermotoga maritima2Methanothermus fervidus *3Thermotoga neriphila2Persephonella marina *2Thermotoga pertophila2Persephonella marina *10Thermus aquaticus *3Pyrobaculum carophilum21Thermus aquaticus *2Pyrobaculum acidifiotis *10Thermus aquaticus *3Pyrobaculum islandicum3Thermus aquaticus *3	Hyperthermus butylicus *	3	Thermococcus gorgonarius	1
Ignisphaera aggregans *1Thermococcus kodakarensis *138Meiothermus ruber *1Thermococcus litoralis *19Metallosphaera cuprina1Thermococcus onnurineus *6Metallosphaera prunae1Thermococcus profundus *1Metallosphaera sedula *12Thermococcus profundus *1Metallosphaera sedula *12Thermococcus cilligii *2Methanocaldococcus forvens *1Thermoleophilum album1Methanocaldococcus infernus *2Thermoleophilum album31Methanocaldococcus jamaschii315Thermoleophilum album31Methanocaldococcus infernus *2Thermologasma acidophilum31Methanocaldococcus infernus *16Thermoplasma acidophilum31Methanothermobacter marburgensis59Thermoproteus tenax *19Methanothermobacter thermautotrophicus6Thermosylidibacter cecani *1Methanothermos fervidus *8Thermosylidibacter takaii *2Methanothermos fervidus *1Thermologa maritima260Nanoarchaeum equitans3Thermologa neapolitana10Parageobacillus thermoglucosidasius *1Thermologa petrophila2Prophaculum aerophilum21Thermus gliformis *2Pyrobaculum aerophilum3Thermus fliformis *3Pyrobaculum aerophilum3Thermus fliformis *2Pyrobaculum asidadicum3Thermus fliformis *2Pyrobaculum a	Ignicoccus hospitalis *	5	Thermococcus hydrothermalis	1
Meiothermus ruber *1Thermococcus litoralis *19Metallosphaera cuprina1Thermococcus onnurineus *6Metallosphaera sedula *1Thermococcus onnurineus *6Metallosphaera sedula *12Thermococcus profindus *1Methanocaldococcus fervens *1Thermodesulfobacterium geofontis *1Methanocaldococcus fervens *1Thermodesulfobacterium geofontis *1Methanocaldococcus fervens *2Thermoleophilum album1Methanocaldococcus fervens *16Thermoplasma acidophilum31Methanothermobacter marburgensis59Thermoproteus tenax *19Methanothermobacter thermautotrophicus117Thermosediminibacter oceani *1Methanothermobacter thermautotrophicus *6Thermoslyfidibacter takaii *2Methanothermus fervidus *8Thermotoga maritima260Nanoarchaeum equitans3Thermotoga nearolitana10Parageobacillus thermoglucosidasius *1Thermotoga nearolitana2Pyrobaculum aerophilum21Thermus filformis *2Pyrobaculum caldifontis *10Thermus sectoductus *3Pyrobaculum islandicum3Thermus fulformis *2Pyrobaculum islandicum3Thermus fulformis *1Pyrococcus furiosus75Ureibacillus thermosphaericus *3Pyrococcus furiosus10Thermus fulformis *2Pyrobaculum caldifontis *10Thermus fulformis *3	Ignisphaera aggregans *	1	Thermococcus kodakarensis *	138
Metallosphaera cuprina1Thermococcus onnurineus *6Metallosphaera prunae1Thermococcus profundus *1Metallosphaera sedula *12Thermococcus zilligii *2Methanocaldococcus ifervens *1Thermodesulfobacterium geofontis *1Methanocaldococcus infernus *2Thermoleophilum album1Methanocaldococcus infernus *16Thermoplasma acidophilum31Methanocaldococcus jannaschii315Thermoplasma acidophilum31Methanothermobacter marburgensis59Thermoplasma volcanium13Methanothermobacter thermautotrophicus117Thermosediminibacter oceani *1Methanothermobacter thermautotrophicus *6Thermosulfidibacter takaii *2Methanothermoscocus thermolithotrophicus *8Thermosulfidibacter takaii *260Nanoarchaeum equitans3Thermotoga maritima260Nanoarchaeum equitans3Thermotoga petrophila2Persephonella marina *2Thermovibrio ammonificans *1Picrophilus torridus *10Thermus aquaticus *2Pyrobaculum calidifontis *10Thermus pliformis *2Pyrobaculum calidifontis *10Thermus pliformis *3Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum calidifontis *10Thermus philus455Pyrococcus furiosus3Thermus thermophil	Meiothermus ruber *	1	Thermococcus litoralis *	19
Metallosphaera prunae1Thermococcus profundus *1Metallosphaera sedula *12Thermococcus cilligii *2Methanocaldococcus fervens *1Thermoleosulfobacterium geofontis *1Methanocaldococcus infernus *2Thermoleophilum album1Methanocaldococcus jannaschii315Thermoplasma acidophilum31Methanopyrus kandleri16Thermoplasma acidophilum13Methanothermobacter marburgensis59Thermoproteus tenax *19Methanothermobacter thermautotrophicus6Thermosulfidibacter to cecani *1Methanothermobacter thermolithotrophicus *6Thermosulfidibacter takaii *2Methanothermobacter thermolithorophicus *8Thermosulfidibacter takaii *2Methanothermosulta fervidus *2Thermosulfidibacter takaii *2Methanothermosulta fervidus *3Thermosulfidibacter takaii *2Methanothermosulta fervidus *2Thermotoga maritima260Nanoarchaeum equitans3Thermotoga neapolitana10Parageobacillus thermoglucosidasius *1Thermotoga petrophila2Persephonella marina *2Thermosulfidifornis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrococcus furiosus3Thermosphaericus *1Pyrococcus furiosus3Thermus furthermosphaericus *3Pyrococcus furiosus3	Metallosphaera cuprina	1	Thermococcus onnurineus *	6
Metallosphaera sedula *12Thermococcus zilligii *2Methanocaldococcus fervens *1Thermodesulfobacterium geofontis *1Methanocaldococcus infernus *2Thermoleophilum album1Methanocaldococcus jannaschii315Thermoleophilum acidophilum31Methanotaldococcus jannaschii315Thermoplasma acidophilum31Methanothermobacter marburgensis59Thermoproteus tenax *19Methanothermobacter thermautotrophicus117Thermosediminibacter oceani *1Methanothermosoccus thermolithotrophicus *6Thermosulfidibacter takaii *2Methanothermus fervidus *8Thermotoga maritima260Nanoarchaeum equitans3Thermotoga neapolitana10Parageobacillus thermoglucosidasius *1Thermotoga neapolitana10Persephonella marina *2Thermotypic aquaticus *2Pyrobaculum aerophilum21Thermus filiformis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum islandicum3Thermus thermophilus *3Pyrococcus abyssi75Ureibacillus thermophilus *455Pyrococcus horikoshii213Vulcanisaeta distributa *1Pyrococcus horikoshii147Thermus thermophilus *1	Metallosphaera prunae	1	Thermococcus profundus *	1
Methanocaldococcus fervens *1Thermodesulfobacterium geofontis *1Methanocaldococcus infernus *2Thermoleophilum album1Methanocaldococcus jannaschii315Thermoplasma acidophilum31Methanopyrus kandleri16Thermoplasma volcanium13Methanopyrus kandleri16Thermoplasma volcanium13Methanothermobacter marburgensis59Thermoproteus tenax *19Methanothermobacter thermautotrophicus117Thermosediminibacter oceani *1Methanothermosoccus thermolithotrophicus *6Thermosulfidibacter takaii *2Methanothermus fervidus *8Thermotoga maritima260Nanoarchaeum equitans3Thermotoga neapolitana10Parageobacillus thermoglucosidasius *1Thermotoga neapolitana1Picrophilus torridus *10Thermus quaticus *2Pyrobaculum aerophilum21Thermus scotoductus *2Pyrobaculum islandicum3Thermus thermophilus *3Pyrococcus abyssi75Ureibacillus thermophilus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii147141	Metallosphaera sedula *	12	Thermococcus zilligii *	2
Methanocaldococcus infernus *2Thermoleophilum album1Methanocaldococcus jannaschii315Thermoplasma acidophilum31Methanopyrus kandleri16Thermoplasma volcanium13Methanothermobacter marburgensis59Thermoproteus tenax *19Methanothermobacter thermautotrophicus117Thermosediminibacter oceani *1Methanothermococcus thermolithotrophicus *6Thermosulfaibacter takaii *2Methanothermus fervidus *8Thermosulfaibacter takaii *2Methanotorris igneus *2Thermotoga maritima260Nanoarchaeum equitans3Thermotoga neapolitana10Parageobacillus thermoglucosidasius *1Thermotoga petrophila2Persephonella marina *2Thermosulficrmis *1Pyrobaculum calidifontis *10Thermus glucicus *2Pyrobaculum silandicum3Thermus filformis *2Pyrococcus alyssi75Ureibacillus thermosphaericus *3Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii147147	Methanocaldococcus fervens *	1	Thermodesulfobacterium geofontis *	1
Methanocaldococcus jannaschii315Thermoplasma acidophilum31Methanopyrus kandleri16Thermoplasma volcanium13Methanothermobacter marburgensis59Thermoproteus tenax *19Methanothermobacter thermautotrophicus117Thermosediminibacter oceani *1Methanothermobacter thermautotrophicus *6Thermosipho africanus2Methanothermus fervidus *8Thermosulfidibacter takaii *2Methanotorris igneus *2Thermotoga neapolitana10Nanoarchaeum equitans3Thermotoga petrophila2Parageobacillus thermoglucosidasius *1Thermosiproi ammonificans *1Picrophilus torridus *10Thermus gauticus *25Pyrobaculum aerophilum21Thermus filformis *2Pyrobaculum islandicum3Thermus thermophilus455Pyrococcus abyssi75Ureibacillus thermophilus *1Pyrococcus horikoshii213Vulcanisaeta distributa *1Pyrococcus horikoshii1471471	Methanocaldococcus infernus *	2	Thermoleophilum album	1
Methanopyrus kandleri16Thermoplasma volcanium13Methanothermobacter marburgensis59Thermoproteus tenax *19Methanothermobacter thermautotrophicus117Thermosediminibacter oceani *1Methanothermococcus thermolithotrophicus *6Thermosipho africanus2Methanothermus fervidus *8Thermosulfidibacter takaii *2Methanotorris igneus *2Thermotoga maritima260Nanoarchaeum equitans3Thermotoga neapolitana10Parageobacillus thermoglucosidasius *1Thermotoga petrophila2Persephonella marina *2Thermotoga petrophila2Pyrobaculum aerophilum21Thermus filifornis *2Pyrobaculum islandicum3Thermus scotoductus *3Pyrococcus abyssi75Ureibacillus thermosphaericus *455Pyrococcus horikoshii213Vulcanisaeta distributa *1Pyrococcus horikoshii1471471	Methanocaldococcus jannaschii	315	Thermoplasma acidophilum	31
Methanothermobacter marburgensis59Thermoproteus tenax *19Methanothermobacter thermautotrophicus117Thermosediminibacter oceani *1Methanothermococcus thermolithotrophicus *6Thermosipho africanus2Methanothermus fervidus *8Thermosulfidibacter takaii *2Methanotorris igneus *2Thermosulfidibacter takaii *2Methanotorris igneus *2Thermotoga maritima260Nanoarchaeum equitans3Thermotoga neapolitana10Parageobacillus thermoglucosidasius *1Thermotoga petrophila2Persephonella marina *2Thermovibrio ammonificans *1Picrophilus torridus *10Thermus aquaticus *25Pyrobaculum aerophilum21Thermus filiformis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrococcus abyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii147141	Methanopyrus kandleri	16	Thermoplasma volcanium	13
Methanothermobacter thermautotrophicus117Thermosediminibacter oceani *1Methanothermococcus thermolithotrophicus *6Thermosipho africanus2Methanothermus fervidus *8Thermosulfidibacter takaii *2Methanotorris igneus *2Thermotoga maritima260Nanoarchaeum equitans3Thermotoga neapolitana10Parageobacillus thermoglucosidasius *1Thermotoga petrophila2Persephonella marina *2Thermotoga petrophila2Picrophilus torridus *10Thermus aquaticus *25Pyrobaculum aerophilum21Thermus filiformis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum islandicum3Thermus thermophilus455Pyrococcus dyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii1471471	Methanothermobacter marburgensis	59	Thermoproteus tenax *	19
Methanothermococcus thermolithotrophicus *6Thermosipho africanus2Methanothermus fervidus *8Thermosulfidibacter takaii *2Methanotorris igneus *2Thermotoga maritima260Nanoarchaeum equitans3Thermotoga neapolitana10Parageobacillus thermoglucosidasius *1Thermotoga petrophila2Persephonella marina *2Thermovibrio ammonificans *1Picrophilus torridus *10Thermus aquaticus *25Pyrobaculum aerophilum21Thermus filiformis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum islandicum3Thermus thermophilus thermosphaericus *1Pyrococcus dayssi75Ureibacillus thermosphaericus *1Pyrococcus horikoshii1471471	Methanothermobacter thermautotrophicus	117	Thermosediminibacter oceani *	1
Methanothermus fervidus *8Thermosulfidibacter takaii *2Methanotorris igneus *2Thermotoga maritima260Nanoarchaeum equitans3Thermotoga neapolitana10Parageobacillus thermoglucosidasius *1Thermotoga petrophila2Persephonella marina *2Thermovibrio ammonificans *1Picrophilus torridus *10Thermus aquaticus *25Pyrobaculum aerophilum21Thermus filiformis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum islandicum3Thermus thermophilus455Pyrococcus dyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii1471471	Methanothermococcus thermolithotrophicus *	6	Thermosipho africanus	2
Methanotorris igneus *2Thermotoga maritima260Nanoarchaeum equitans3Thermotoga neapolitana10Parageobacillus thermoglucosidasius *1Thermotoga petrophila2Persephonella marina *2Thermovibrio ammonificans *1Picrophilus torridus *10Thermus aquaticus *25Pyrobaculum aerophilum21Thermus filiformis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum islandicum3Thermus thermophilus455Pyrococcus abyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii1471471	Methanothermus fervidus *	8	Thermosulfidibacter takaii *	2
Nanoarchaeum equitans3Thermotoga neapolitana10Parageobacillus thermoglucosidasius *1Thermotoga petrophila2Persephonella marina *2Thermovibrio ammonificans *1Picrophilus torridus *10Thermus aquaticus *25Pyrobaculum aerophilum21Thermus filiformis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum islandicum3Thermus thermophilus455Pyrococcus abyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii1471471	Methanotorris igneus *	2	Thermotoga maritima	260
Parageobacillus thermoglucosidasius *1Thermotoga petrophila2Persephonella marina *2Thermovibrio ammonificans *1Picrophilus torridus *10Thermus aquaticus *25Pyrobaculum aerophilum21Thermus filiformis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum islandicum3Thermus thermophilus455Pyrococcus abyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii1471471	Nanoarchaeum equitans	3	Thermotoga neapolitana	10
Persephonella marina *2Thermovibrio ammonificans *1Picrophilus torridus *10Thermus aquaticus *25Pyrobaculum aerophilum21Thermus filiformis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum islandicum3Thermus thermophilus455Pyrococcus abyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii1471471	Parageobacillus thermoglucosidasius *	1	Thermotoga petrophila	2
Picrophilus torridus *10Thermus aquaticus *25Pyrobaculum aerophilum21Thermus filiformis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum islandicum3Thermus thermophilus455Pyrococcus abyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii1471471	Persephonella marina *	2	Thermovibrio ammonificans *	1
Pyrobaculum aerophilum21Thermus filiformis *2Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum islandicum3Thermus thermophilus455Pyrococcus abyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii1471471	Picrophilus torridus *	10	Thermus aquaticus *	25
Pyrobaculum calidifontis *10Thermus scotoductus *3Pyrobaculum islandicum3Thermus thermophilus455Pyrococcus abyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii1471471	Pyrobaculum aerophilum	21	Thermus filiformis *	2
Pyrobaculum islandicum3Thermus thermophilus455Pyrococcus abyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii14711	Pyrobaculum calidifontis *	10	Thermus scotoductus *	3
Pyrococcus abyssi75Ureibacillus thermosphaericus *1Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii147	Pyrobaculum islandicum	3	Thermus thermophilus	455
Pyrococcus furiosus213Vulcanisaeta distributa *1Pyrococcus horikoshii147	Pyrococcus abyssi	75	Ureibacillus thermosphaericus *	1
Pyrococcus horikoshii 147	Pyrococcus furiosus	213	Vulcanisaeta distributa *	1
	Pyrococcus horikoshii	147		

Species	Number of proteins	Species	Number of proteins
Acaryochloris marina	3	Formosa agariphila *	36
Acidithiobacillus ferridurans *	2	Geobacter metallireducens	3
Acidithiobacillus thiooxidans *	2	Geobacter sulfurreducens	10
Actinoplanes missouriensis *	3	Gillisia limnaea *	1
Aeromonas hydrophila	11	Gluconacetobacter diazotrophicus	2
Agrobacterium fabrum	52	Gluconobacter oxydans	9
Agrobacterium radiobacter	5	Gluconobacter thailandicus	2
Agrobacterium tumefaciens	5	Halalkalibacterium halodurans	20
Albidiferax ferrireducens	1	Haliscomenobacter hydrossis *	1
Alcanivorax borkumensis *	4	Halobacillus andaensis *	1
Aliivibrio fischeri	10	Halomonas halodenitrificans *	3
Aquaspirillum arcticum *	1	Halothiobacillus neapolitanus *	15
Aquincola tertiaricarbonis *	4	Herminiimonas arsenicoxydans	3
Aromatoleum aromaticum	4	Hyphomicrobium methylovorum *	3
Asticcacaulis excentricus *	4	Hyphomicrobium sulfonivorans *	1
Bacillus atrophaeus *	3	Ideonella sakaiensis *	2
Bacillus mojavensis	3	Ilyobacter polytropus *	2
Bacillus subtilis	1951	Lacinutrix mariniflava *	1
Bartonella bacilliformis	1	Lactococcus lactis	75
Bradyrhizobium diazoefficiens	17	Leifsonia aquatica *	1
Brevibacillus laterosporus	1	Leptospira interrogans	10
Carnobacterium maltaromaticum *	7	Leuconostoc mesenteroides	15
Catenulispora acidiphila *	2	Leucothrix mucor	1
Cellvibrio japonicus	17	Lewinella persica *	1
Chitinophaga pinensis *	3	Lysobacter antibioticus *	3
Chlamydia trachomatis	41	Marivirga tractuosa *	1
Chondromyces crocatus *	2	Martelella endophytica *	1
Chromobacterium violaceum	10	Mesorhizobium japonicum *	16
Clavibacter michiganensis	1	Methanococcoides burtonii *	3
Clostridium botulinum	15	Methanococcus vannielii	6
Colwellia psychrerythraea	5	Methanosphaerula palustris *	1
Corynebacterium ammoniagenes *	2	Methylorubrum extorquens	20
Cupriavidus metallidurans	22	Moritella abyssi *	2
Cupriavidus necator	40	Moritella profunda *	1
Cupriavidus pinatubonensis	5	Mycoplasma genitalium	7
Cyclobacterium marinum *	1	Mycoplasmopsis agalactiae	1
Cytophaga hutchinsonii	2	Myxococcus fulvus *	2
Dehalococcoides mccartyi *	6	Myxococcus xanthus	45
Deinococcus radiodurans	70	Nonlabens dokdonensis *	1
Delftia acidovorans	6	Nonlabens ulvanivorans *	4
Desulfotalea psychrophila *	1	Oceanicola granulosus *	2
Escherichia coli	259	Oceanobacillus iheyensis	1
Flavobacterium frigidimaris *	1	Oenococcus oeni	3
Flavobacterium johnsoniae	6	Paenarthrobacter aurescens	1
Flavobacterium psychrophilum	1	Paenibacillus amylolyticus *	1

Table S3. List of non-thermophilic species and number of proteins in our full dataset before filtering for sequence length and CD-HIT. Species marked with an asterisk were used for our independent test set.

Species	Number of proteins	Species	Number of proteins
Paenibacillus lemnae *	1	Saccharopolyspora erythraea	19
Paenibacillus thiaminolyticus *	1	Salinispora arenicola *	1
Paludibacter propionicigenes	1	Salinispora tropica	4
Paraburkholderia phytofirmans	1	Shewanella colwelliana	1
Paraburkholderia xenovorans	9	Shewanella frigidimarina	3
Paracoccus denitrificans	48	Shewanella halifaxensis	1
Pectobacterium atrosepticum	17	Shewanella oneidensis *	34
Pelagibacterium halotolerans *	1	Shewanella pealeana *	1
Photobacterium phosphoreum *	7	Shigella flexneri	55
Photobacterium profundum	3	Singulisphaera acidiphila *	1
Polaromonas naphthalenivorans	1	Sinorhizobium medicae	2
Polaromonas sp.	3	Sodalis glossinidius	1
Prosthecochloris aestuarii	3	Sorangium cellulosum *	4
Pseudoalteromonas atlantica *	5	Sphingosinicella xenopeptidilytica *	1
Pseudoalteromonas carrageenovora *	2	Staphylococcus xylosus *	7
Pseudoalteromonas haloplanktis	4	Starkeya novella *	7
Pseudoalteromonas piscicida *	2	Stigmatella aurantiaca *	7
Pseudoalteromonas translucida	3	Streptomyces avermitilis	16
Pseudomonas aeruginosa	424	Streptomyces cyslabdanicus *	2
Pseudomonas entomophila	3	Streptomyces muensis *	1
Pseudomonas fluorescens	37	Streptomyces tsukubensis *	1
Pseudomonas marginalis *	3	Sulfurimonas autotrophica *	1
Pseudooceanicola batsensis *	1	Sulfurospirillum multivorans *	2
Psychrobacter arcticus	3	Synechocystis sp.	142
Psychrobacter cryohalolentis	2	Thalassotalea agarivorans	1
Psychrobacter immobilis *	2	Thiothrix nivea *	1
Psychroflexus torquis *	1	Vibrio campbellii	3
Psychromonas ingrahamii	2	Vibrio cholerae serotype	147
Renibacterium salmoninarum	1	Vibrio harveyi	13
Rhizobium leguminosarum	24	Vibrio metoecus *	1
Rhizobium meliloti	54	Vibrio parahaemolyticus	26
Rhizobium radiobacter	28	Vibrio vulnificus	14
Rhodococcus erythropolis	23	Xanthomonas axonopodis	3
Rhodopirellula baltica	1	Xanthomonas campestris	40
Rhodopseudomonas palustris	81	Xanthomonas citri	3
Rhodospirillum rubrum	27	Xylella fastidiosa	9
Rickettsia prowazekii	11	Yersinia enterocolitica	47
Roseivirga ehrenbergii *	2	Yersinia pestis	77
Roseovarius nubinhibens *	3	Yersinia pseudotuberculosis	25
Runella zeae *	2	Zunongwangia profunda *	1
Saccharibacillus brassicae *	1	Zymomonas mobilis	19
Saccharophagus degradans	1	-	

Table S4. List of non-thermophilic species and number of proteins in our full dataset before filtering for sequence length and CD-HIT. Species marked with an asterisk were used for our independent test set.

**Table S5.** Hyperparameters and ranges optimized for ProLaTherm: Numbers reflect a list of potential values (curly brackets) or the lower respective upper bound (square brackets), with a step size of 1 for integer values as default. In some cases, specific step sizes  $\Delta$  were used instead of a continuous search space.

Hyperparameter	Values	Notes
	ProLaTherm	
dropout learning_rate early_stopping_patience batch_size n_init_units_factor	$ \begin{array}{l} [0,0.5] \text{ with } \Delta \!=\! 0.1 \\ \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\} \\ [0,10] \text{ with } \Delta \!=\! 5 \\ \{64, 128, 256\} \\ [0.5,1] \text{ with } \Delta \!=\! 0.05 \end{array} $	dropout rate for dropout layers learning rate of the Adam optimizer epochs without improvement needed for early stopping batch size for training number of neurons in the first fully-connected layer in relation to the number of input features

**Table S6.** Hyperparameters and ranges optimized for the feature-based comparison partners Elastic Net, SVM, Random Forest, XGBoost and MLP: Numbers reflect a list of potential values (curly brackets) or the lower respective upper bound (square brackets), with a step size of 1 for integer values as default. In some cases, specific step sizes  $\Delta$  were used instead of a continuous search space.

Hyperparameter	Values	Notes
	Elastic Net	
С	$[10^{-3}, 10^3]$	weighting factor of the regularization terms
11_ratio	$[0.05, 0.95]$ with $\Delta\!=\!0.05$	trade off between L1- and L2-regularization
	SVM	
kernel	{'linear', 'poly', 'rbf'}	kernel function to use
С	$[10^{-3}, 10^3]$	regularization factor
degree	[1,5]	polynomial degree of kernel function (if kernel is 'poly')
gamma	$[10^{-3}, 10^3]$	kernel coefficient (if kernel is 'rbf' or 'poly')
	Random Forest	
n estimators	$\{50, 100, 250, 500, 750, 1000, 1250, 1500, 1750, 2000,$	number of trace in the encemble
n_estimators	$2250, 2500, 2750, 3000, 3500, 4000, 4500, 5000\}$	number of trees in the ensemble
min_samples_split	$[0.005, 0.9]$ with $\Delta \!=\! 0.005$	minimum ratio of the number of samples to split a node
max_depth	$[2,50]$ with $\Delta = 2$	maximum depth of a tree
min_samples_leaf	$[0.005, 0.5]$ with $\Delta \!=\! 0.005$	minimum ratio of the number of samples at a leaf node
max_features	{'sqrt', 'log2'}	number of features to consider at determining best split
	XGBoost	
n_estimators	$\{50, 100, 250, 500, 750, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750, 2000\}$	number of trees in the ensemble
max depth	[2.20]	maximum depth of a tree
learning rate	[2,20] [0.025.0.5] with $\Lambda = 0.025$	hoosting learning rate
gamma	$[0.025, 0.5]$ with $\Delta = 0.025$	minimum loss reduction for a further partition on a leaf node
subsample	$[0, 10]$ with $\Delta = 0.1$ [0.05, 0.95] with $\Delta = 0.05$	subsample ratio of training instances for tree construction
colsample bytree	$[0.05, 0.95]$ with $\Delta = 0.05$	ratio of features to use for each tree
reg alpha	$[0.00, 0.00]$ with $\Delta = 0.00$	I 1-regularization term on weights
	MLP	
dropout	$[0, 0.5]$ with $\Delta = 0.1$	dropout rate for dropout layers
act_function	$\{\text{relu}', \text{tanh}'\}$	activation function to use
learning_rate	$\{10^{\circ}, 10^{\circ}, 10^{\circ}, 10^{\circ}, 10^{\circ}, 10^{\circ}\}$	learning rate of the Adam optimizer
early_stopping_patience	$[0, 10]$ with $\Delta = 5$	epochs without improvement needed for early stopping
batch_size	$\{64, 128, 256\}$	batch size for training
n_layers	[1,5]	number of blocks consisting of a fully-connected, batch
n_init_units_factor	$[0.1, 0.95]$ with $\Delta\!=\!0.05$	normalization and dropout layer number of neurons in the first fully-connected layer in relation to the number of input features
perc_dec	$[0.1, 0.5]$ with $\Delta\!=\!0.05$	percentage decrease of number of neurons per building block

**Table S7.** Hyperparameters and ranges optimized for the hybrid sequence-based comparison partners LSTM\_BasicDesc and Bi-LSTM\_BasicDesc as well as the purely sequence-based models MLP\_Embedding, LSTM and Bi-LSTM: Numbers reflect a list of potential values (curly brackets) or the lower respective upper bound (square brackets), with a step size of 1 for integer values as default. In some cases, specific step sizes  $\Delta$  were used instead of a continuous search space.

Hyperparameter	Values		Notes
		LSTM_BasicDesc	
dropout	$[0,0.5]$ with $\Delta = 0.1$	Lo Ini-DusieDese	dropout rate for dropout layers
act_function	{'relu', 'tanh'}		activation function to use
learning_rate	$\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$		learning rate of the Adam optimizer
early_stopping_patience	$[0,10]$ with $\Delta = 5$		epochs without improvement needed for early stopping
batch_size	{64,128,256}		batch size for training
n_lstm_layers	[1, 3]		number of LSTM layers
hidden_size_exp	[3, 8]		dimensionality of hidden states as exponent with base 2
n_init_units_factor	$[0.5,1]$ with $\Delta\!=\!0.05$		number of neurons in the first fully-connected layer in
			relation to the number of input features
latent_dim_exp	[3, 8]		final dimensionality of latent representation after linear layers
n_lin_layer	[0, 3]		number of linear layers to reach latent_dim_exp, with 0
			reflecting no dimensionality increase
		Bi-LSTM_BasicDesc	
dropout	$[0, 0.5]$ with $\Delta \!=\! 0.1$		dropout rate for dropout layers
act_function	{'relu', 'tanh'}		activation function to use
learning_rate	$\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$		learning rate of the Adam optimizer
early_stopping_patience	$[0,10]$ with $\Delta = 5$		epochs without improvement needed for early stopping
batch_size	$\{64, 128, 256\}$		batch size for training
n_lstm_layers	[1, 3]		number of LSTM layers
hidden_size_exp	[3, 8]		dimensionality of hidden states as exponent with base 2
n_init_units_factor	$[0.5,1]$ with $\Delta\!=\!0.05$		number of neurons in the first fully-connected layer in
			relation to the number of input features
latent_dim_exp	[3, 8]		final dimensionality of latent representation after linear layers
n_lin_layer	[0, 3]		number of linear layers to reach latent_dim_exp, with 0
			reflecting no dimensionality increase
		MLP_Embedding	
dropout	$[0,0.5]$ with $\Delta\!=\!0.1$		dropout rate for dropout layers
act_function	{'relu', 'tanh'}		activation function to use
learning_rate	$\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$		learning rate of the Adam optimizer
early_stopping_patience	$[0,10]$ with $\Delta = 5$		epochs without improvement needed for early stopping
batch_size	$\{64, 128, 256\}$		batch size for training
two_layer_head_class	{False, True}		use first fully-connected layer of head classifier architecture
n_init_units_factor	$[0.5,1]$ with $\Delta\!=\!0.05$		number of neurons in the first fully-connected layer in
			relation to the number of input features
embedding_dim	1024		dimensionality of embedding layer, here set to 1024
		LSTM	
dropout	$[0,0.5]$ with $\Delta\!=\!0.1$		dropout rate for dropout layers
act_function	{'relu', 'tanh'}		activation function to use
learning_rate	$\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$		learning rate of the Adam optimizer
early_stopping_patience	$[0,10]$ with $\Delta = 5$		epochs without improvement needed for early stopping
batch_size	$\{64, 128, 256\}$		batch size for training
n_lstm_layers	[1, 3]		number of LSTM layers
hidden_size_exp	[3, 8]		dimensionality of hidden states as exponent with base 2
n_init_units_factor	$[0.5,1]$ with $\Delta\!=\!0.05$		number of neurons in the first fully-connected layer in
			relation to the number of input features
embedding_dim	1024		dimensionality of embedding layer, here set to 1024
		Bi-LSTM	
dropout	$[0,0.5]$ with $\Delta = 0.1$		dropout rate for dropout layers
act_function	{'relu', 'tanh'}		activation function to use
learning_rate	$\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$		learning rate of the Adam optimizer
early_stopping_patience	$[0,10]$ with $\Delta = 5$		epochs without improvement needed for early stopping
batch_size	$\{64, 128, 256\}$		batch size for training
n_lstm_layers	[1, 3]		number of LSTM layers
hidden_size_exp	[3, 8]		dimensionality of hidden states as exponent with base 2
n_init_units_factor	$[0.5,1]$ with $\Delta \!=\! 0.05$		number of neurons in the first fully-connected layer in
			relation to the number of input features
embedding_dim_exp	[4, 8]		dimensionality of embedding layer as exponent with base 2

**Table S8.** Hyperparameters and ranges optimized for the purely sequence-based comparison partners vanilla-Transformer and BigBird: Numbers reflect a list of potential values (curly brackets) or the lower respective upper bound (square brackets), with a step size of 1 for integer values as default. In some cases, specific step sizes  $\Delta$  were used instead of a continuous search space.

Hyperparameter	Values		Notes
		vanilla-Transformer	
dropout	$[0.1, 0.5]$ with $\Delta = 0.1$		dropout rate for dropout layers
learning_rate	$\{10^{-5}, 10^{-4}, 10^{-3}\}$		learning rate of the Adam optimizer
early_stopping_patience	50		epochs without improvement needed for early stopping, set to 50
n_minibatches	{2, 4, 8}		number of minibatches with similar sequence length within each batch of size 1024
embedding_dim_exp	[4, 6]		dimensionality of embeddings as exponent with base 2
kernel_size_avgpool	2, 3, 5		kernel size of average pooling layer
n_heads	$[2,6]$ with $\Delta = 2$		number of heads in each Transformer layer
n_transformer_blocks	$[2,6]$ with $\Delta = 2$		number of Transformer layers
factor_hidden_dim_mlp	[2, 4]		factor of number of neurons in feedforward part of
			Transformer block in relation to input dimensionality
two_layer_head_class	{False, True}		use first fully-connected layer of head classifier architecture
label_smoothing	$\{0.0, 0.1\}$		label smoothing used for loss calculation
weight_decay	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$		weight decay for Adam optimizer
		BigBird	
dropout	$[0.1, 0.5]$ with $\Delta \!=\! 0.1$		dropout rate for dropout layers
learning_rate	$\{10^{-5}, 10^{-4}, 10^{-3}\}$		learning rate of the Adam optimizer
early_stopping_patience	50		epochs without improvement needed for early stopping, set to 50
n_minibatches	{4, 8}		number of minibatches with similar sequence length within each batch of size 1024
embedding_dim_exp	[4, 6]		dimensionality of embeddings as exponent with base 2
n_heads	$[2,6]$ with $\Delta = 2$		number of heads in each Transformer layer
n_transformer_blocks	$[2,6]$ with $\Delta = 2$		number of Transformer layers
fact_num_global_tokens	{2,3}		factor of number of global tokens in relation to block size
block_size	$\{8, 16, 32\}$		size of each block for sparse self-attention
two_layer_head_class	{False, True}		use first fully-connected layer of head classifier architecture
label_smoothing	$\{0.0, 0.1\}$		label smoothing used for loss calculation
weight_decay	$\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$		weight decay for Adam optimizer

**Table S9.** Overview of the test and validation results with no overlap between the species in the test and cross-validation data: For each prediction model, we show the given evaluation metric on the full test set (*test set 1*) in the first line as well as the mean and standard deviation on the validation sets of the five-fold cross-validation in the second line. The prediction models are grouped as feature-based as well as hybrid and purely sequence-based. With respect to the comparison partners from the literature, we only report the evaluation metrics on the test data as these models were cross-validated using another data split with a potential overlap between the authors' training data and our validation sets. Due to the composition of our test data, we ensure a fair comparison, see Experimental Settings. The best result for each evaluation metric is highlighted in bold - both for the test and validation data. Results for *test set 2* consisting of evolutionary less related proteins can be found in Table 4 of the main paper.

Prediction model	Accuracy	F1-Score	Precision	Recall	Specificity	BACC	MCC
Feature-based models							
Electic Not	0.814	0.822	0.976	0.710	0.973	0.842	0.672
Elastic Net	$(0.911 \pm 0.009)$	$(0.855 \pm 0.015)$	$(0.893 \pm 0.015)$	$(0.821 \pm 0.025)$	$(0.954 \pm 0.007)$	$(0.887 \pm 0.013)$	$(0.793 \pm 0.021)$
SVM	0.817	0.827	0.969	0.722	0.964	0.843	0.673
5 111	$(0.911 \pm 0.009)$	$(0.858 \pm 0.015)$	$(0.879 \pm 0.015)$	$(0.838 \pm 0.019)$	$(0.946 \pm 0.007)$	$(0.892 \pm 0.012)$	$(0.794 \pm 0.022)$
Random Forest	0.712	0.695	0.969	0.542	0.973	0.758	0.532
Randolli I olest	$(0.872 \pm 0.008)$	$(0.771 \pm 0.017)$	$(0.904 \pm 0.016)$	$(0.673 \pm 0.025)$	$(0.966 \pm 0.006)$	$(0.820 \pm 0.012)$	$(0.700 \pm 0.019)$
XGBoost	0.840	0.852	0.974	0.757	0.969	0.863	0.710
AGDoost	$(0.921 \pm 0.008)$	$(0.872 \pm 0.014)$	$(0.905 \pm 0.014)$	$(0.843 \pm 0.024)$	$(0.958 \pm 0.007)$	$(0.900 \pm 0.011)$	$(0.817 \pm 0.019)$
MLP	0.844	0.856	0.971	0.765	0.964	0.865	0.714
	$(0.915 \pm 0.008)$	$(0.863 \pm 0.014)$	$(0.890 \pm 0.005)$	$(0.839 \pm 0.025)$	$(0.951 \pm 0.003)$	$(0.895 \pm 0.012)$	$(0.803 \pm 0.019)$
			Hybrid sequence-	based models			
I STM PasiaDasa	0.837	0.854	0.934	0.786	0.915	0.850	0.685
LSTWI_BasicDesc	$(0.903 \pm 0.008)$	$(0.842 \pm 0.015)$	$(0.877 \pm 0.011)$	$(0.811 \pm 0.032)$	$(0.946 \pm 0.007)$	$(0.878 \pm 0.014)$	$(0.774 \pm 0.019)$
Bi I STM BasicDesc	0.779	0.781	0.974	0.652	0.973	0.813	0.622
DI-LSTWI_DasieDese	$(0.908 \pm 0.014)$	$(0.855 \pm 0.019)$	$(0.869 \pm 0.042)$	$(0.842 \pm 0.009)$	$(0.939 \pm 0.023)$	$(0.891 \pm 0.010)$	$(0.788 \pm 0.031)$
Purely sequence-based models							
	0.819	0.827	0.984	0.713	0.982	0.848	0.684
MLP_Embedding	$(0.892 \pm 0.016)$	$(0.829 \pm 0.017)$	$(0.855 \pm 0.060)$	$(0.811 \pm 0.040)$	$(0.931 \pm 0.039)$	$(0.871 \pm 0.009)$	$(0.754 \pm 0.029)$
LOTM	0.837	0.851	0.953	0.768	0.942	0.855	0.694
LSIM	$(0.901 \pm 0.008)$	$(0.842 \pm 0.013)$	$(0.868 \pm 0.028)$	$(0.819 \pm 0.025)$	$(0.940 \pm 0.016)$	$(0.880 \pm 0.010)$	$(0.772 \pm 0.019)$
D: I CTM	0.807	0.818	0.950	0.719	0.942	0.830	0.648
DI-LSTW	$(0.909 \pm 0.006)$	$(0.853 \pm 0.009)$	$(0.888 \pm 0.033)$	$(0.823 \pm 0.032)$	$(0.950 \pm 0.017)$	$(0.886 \pm 0.009)$	$(0.790 \pm 0.012)$
vanilla Transformar	0.803	0.812	0.964	0.701	0.960	0.831	0.651
valima- mansformer	$(0.898 \pm 0.007)$	$(0.833 \pm 0.012)$	$(0.880 \pm 0.022)$	$(0.791 \pm 0.019)$	$(0.948 \pm 0.011)$	$(0.870 \pm 0.009)$	$(0.762 \pm 0.018)$
BigBird	0.814	0.821	0.984	0.704	0.982	0.843	0.677
Digbild	$(0.894 \pm 0.009)$	$(0.825 \pm 0.016)$	$(0.880 \pm 0.016)$	$(0.776 \pm 0.021)$	$(0.950 \pm 0.007)$	$(0.863 \pm 0.012)$	$(0.752 \pm 0.022)$
ProLaTherm	0.919	0.929	0.997	0.870	0.996	0.933	0.847
riobaineim	$(0.978 \pm 0.005)$	$(0.966 \pm 0.009)$	$(0.980 \pm 0.005)$	$(0.952 \pm 0.015)$	$(0.991 \pm 0.002)$	$(0.971 \pm 0.008)$	$(0.950\pm0.012)$
Comparison partners from literature							
ThermoPred(5)	0.817	0.840	0.895	0.791	0.857	0.824	0.635
SCMTPP (6)	0.807	0.821	0.937	0.730	0.924	0.827	0.641
iThermo(7)	0.819	0.842	0.893	0.797	0.853	0.825	0.637
SAPPHIRE (8)	0.870	0.884	0.966	0.814	0.955	0.885	0.752
DeepTP* (9)	0.888	0.903	0.925	0.882	0.897	0.889	0.772
BertThermo(10)	0.880	0.898	0.931	0.867	0.902	0.884	0.757

\* 25 proteins excluded from evaluation metric calculation due to overlap with the comparison partner's training data

Table S10. Mean BLAST identity of thermophilic species from *test set 1*: The total number of proteins per thermophilic organism used in *test set 1* is given, as well as the number of correctly classified proteins. For each species we show the average BLAST sequence identity of the best hit among the thermophilic and non-thermophilic training data.

Species	# proteins	# true positives	BLAST identity thermo.	BLAST identity non-thermo.
Acidianus ambivalens	10	8	0.190	0.114
Aciduliprofundum boonei	1	1	0.568	0.559
Aquifex pyrophilus	3	3	0.348	0.352
Archaeoglobus profundus	8	8	0.480	0.147
Caldalkalibacillus thermarum	1	0	0.265	0.235
Caldicellulosiruptor saccharolyticus	1	1	0.047	0.447
Deferribacter desulfuricans	1	1	0.385	0.424
Fervidobacterium pennivorans	1	1	0.020	0.143
Geobacillus kaustophilus	37	11	0.313	0.499
Geoglobus acetivorans	1	1	0.404	0.323
Hydrogenobacter thermophilus	9	8	0.231	0.239
Hyperthermus butylicus	3	3	0.538	0.222
Ignicoccus hospitalis	5	4	0.493	0.176
Ignisphaera aggregans	1	1	0.540	0.540
Meiothermus ruber	1	0	0.041	0.461
Metallosphaera sedula	10	10	0.490	0.288
Methanocaldococcus fervens	1	1	0.196	0.034
Methanocaldococcus infernus	2	2	0.063	0.094
Methanothermococcus thermolithotrophicus	5	5	0.516	0.275
Methanothermus fervidus	6	6	0.425	0.180
Methanotorris igneus	2	2	0.542	0.073
Parageobacillus thermoglucosidasius	-	-	0.146	0.600
Persephonella marina	2	2	0.195	0.059
Picrophilus torridus	10	-	0.322	0.187
Pyrobaculum calidifontis	10	10	0.271	0.205
Rhodothermus marinus	8	6	0.163	0.179
Spirochaeta thermophila	2	1	0.126	0.189
Sulfonhobococcus zilligii	-	1	0.619	0.088
Thermoanaerobacter ethanolicus	1	1	0.301	0.286
Thermoanaerobacter italicus	1	1	0.267	0 534
Thermoanaerobacterium saccharolyticum	3	2	0.334	0.222
Thermoanaerobacterium thermosulfurigenes	3	2	0.297	0.099
Thermococcus barophilus	1	-	0.299	0.304
Thermococcus cleftensis	2	2	0.210	0.132
Thermococcus funicolans	1	-	0.776	0.115
Thermococcus gammatolerans	1	1	0.831	0.153
Thermococcus kodakarensis	118	118	0.582	0.196
Thermococcus litoralis	16	16	0.519	0.244
Thermococcus onnurineus	5	5	0.489	0.119
Thermococcus profundus	1	1	0.847	0.525
Thermococcus zilligii	1	1	0.534	0.065
Thermodesulfobacterium geofontis	1	1	0.099	0 291
Thermoproteus tenax	18	18	0.280	0.158
Thermosediminibacter oceani	1	1	0.050	0.624
Thermosulfidibacter takaii	2	2	0.275	0.355
Thermovibrio ammonificans	- 1	- 1	0.042	0.050
Thermus aquaticus	19	16	0.601	0.329
Thermus filiformis	1	1	0.891	0.028
Thermus scotoductus	3	3	0.394	0.174
Ureibacillus thermosphaericus	1	0	0.647	0.052
Vulcanisaeta distributa	1	1	0.303	0.087
	-	-		

#### REFERENCES

- 1. Prabhakaran, M. (1990) The distribution of physical, chemical and conformational properties in signal and nascent peptides. *The Biochemical journal*, **269**(3), 691–696.
- 2. Tomii, K. and Kanehisa, M. (1996) Analysis of amino acid indices and mutation matrices for sequence comparison and structure prediction of proteins. *Protein engineering*, **9**(1), 27–36.
- 3. Dubchak, I., Muchnik, I., Mayor, C., Dralyuk, I., and Kim, S.-H. (1999) Recognition of a protein fold in the context of the SCOP classification. *Proteins: Structure, Function, and Bioinformatics*, **35**(4), 401–407.
- 4. Chou, K. C. (2001) Prediction of protein cellular attributes using pseudo-amino acid composition. Proteins, 43(3), 246-255.
- 5. Lin, H. and Chen, W. (2011) Prediction of thermophilic proteins using feature selection technique. Journal of microbiological methods, 84(1), 67-70.
- 6. Charoenkwan, P., Chotpatiwetchkul, W., Lee, V. S., Nantasenamat, C., and Shoombuatong, W. (2021) A novel sequence-based predictor for identifying and characterizing thermophilic proteins using estimated propensity scores of dipeptides. *Scientific reports*, **11**(1), 23782.
- 7. Ahmed, Z., Zulfiqar, H., Khan, A. A., Gul, I., Dao, F.-Y., Zhang, Z.-Y., Yu, X.-L., and Tang, L. (2022) iThermo: A Sequence-Based Model for Identifying Thermophilic Proteins Using a Multi-Feature Fusion Strategy. *Frontiers in microbiology*, **13**, 790063.
- Charoenkwan, P., Schaduangrat, N., Moni, M. A., Lio', P., Manavalan, B., and Shoombuatong, W. (2022) SAPPHIRE: A stacking-based ensemble learning framework for accurate prediction of thermophilic proteins. *Computers in biology and medicine*, 146, 105704.
- 9. Zhao, J., Yan, W., and Yang, Y. (2023) DeepTP: A Deep Learning Model for Thermophilic Protein Prediction. *International Journal of Molecular Sciences*, 24(3).
- Pei, H., Li, J., Ma, S., Jiang, J., Li, M., Zou, Q., and Lv, Z. (2023) Identification of Thermophilic Proteins Based on Sequence-Based Bidirectional Representations from Transformer-Embedding Features. *Applied Sciences*, 13(5).