Supporting Information

Akanuma et al. 10.1073/pnas.1308215110

SI Results and Discussion

Structural information is required to investigate the features by which proteins adapt to their environments. A crystal structure of a resurrected ancestral protein for vertebrate hormone receptors that might have existed ∼450 million years ago has been solved (1). However, to date, no structure has been available for any of the most deeply rooted archaeal and bacterial proteins. We successfully solved the crystal structures for the ancestral sequences for Archaea (Arc1) and Bacteria (Bac1) predicted by CODEML to 2.4-Å resolution. The Arc1 and Bac1 protomers self-associate as hexameric structures, a quaternary structure found for many extant nucleoside diphosphate kinases (NDKs) (2–4), and the relationship among the protomers of Arc1 and Bac1 is identical to that of all known hexameric NDKs (Fig. S6*A*). Superdex 200 chromatography performed for purification of the other ancestral NDKs indicated that they also are stable hexamers. The Arc1 and Bac1 promoters each contain a βαββαβ supersecondary structure, which also has been found in the structures of extant NDKs (5). The structures of Arc1 and Bac1 are virtually superimposable on that of *Thermus thermophilus* NDK (137 aligned residues; rmsd for Arc1 and *T. thermophilus* NDK C_αs is 0.53 Å, and the rmsd for Bac1 and *T. thermophilus* NDK C_αs is 0.68 Å; Fig. S6*B*). Subtle deviations in the backbone C_α positions are found only in the loops and both termini. The rmsd for the active-site C_αs of *T. thermophilus* NDK and Arc1 is 0.87 Å and for *T. thermophilus* NDK and Bac1 is 0.86 Å (Fig. S6*C*). These very similar active-site structures are consistent with the observation that all three enzymes exhibit similar kinetic parameters (Table 2).

All ancestral NDKs have extremely high thermal stability, suggesting that their conserved residues are responsible for their stability. Therefore, to identify structural features that might be responsible for the thermostability, we compared the Arc1 and Bac1 structures with that of the NDK from the mesophilic *Dictyostelium discoideum* (4). The C_αs (of the 139 aligned residues) of the reconstructed and *D. discoideum* NDKs superimposed with an rmsd of 0.81 Å (Arc1 vs. *D. discoideum* NDK) and 0.82 Å (Bac1 vs. *D. discoideum* NDK) (Fig. S6*D*), indicating that the backbone structure is not responsible for the differences in stability. However, other structural features are thought to influence thermal stability. Because exposed nonpolar residues interact unfavorably with the solvent, reduction of the nonpolar solvent-accessible surface area contributes to protein stability. Indeed, naturally occurring hyperthermophilic proteins often have substantially reduced nonpolar surfaces and increased polar surfaces (6). Nonpolar residues occupy 52% and 51% of the Arc1 and Bac1 surface areas, respectively (Table S1). Conversely, 59% of the *D. discoideum* NDK surface area is nonpolar. Therefore, the reduced nonpolar solvent-accessible surface areas of the ancestral NDKs are likely to be at least partially responsible for their high thermal stability.

Thermophilic proteins often have increased numbers of charged residues on their surfaces as compared with the surfaces of their mesophilic homologs (7). However, because the difference in the numbers of intraprotomer ion pairs for the ancestral NDKs and *D. discoideum* NDK is marginal (Table S1), this type of ion pairing does not seem to contribute substantially to the ancestral proteins' thermostability. More important are the differences in the number of ion pairs and nonpolar interactions found at the interfaces of the reconstructed NDKs and the mesophilic protein protomers. Thermostable oligomeric proteins often have increased numbers of improved ion-pair interactions (8–10) and/or nonpolar interactions (11) at their protomer interfaces as compared with less stable, homologous proteins. Each hexameric NDK structure can be viewed as a trimer of three dimers or as a dimer of two trimers. The dimer interface is formed by an α-helix and a β-strand from each protomer. Compared with those found at the interfaces of the mesophilic NDK, the reconstructed protein interfaces have increased numbers of interprotomer nonpolar interactions that likely contribute to their stability (Table S1). The reconstructed proteins also have increased numbers of electrostatic interactions at the dimer interfaces compared with the interfaces of *D. discoideum* NDK. For Arc1 and Bac1, 14 and 11 interprotomer hydrogen bonds, respectively, are present at each dimer interface, whereas only eight such hydrogen bonds are found in *D. discoideum* NDK. In addition, the dimer interfaces of Arc1 and Bac1 have two and three interprotomer ion pairs, respectively (Fig. S7*A*), but no such interprotomer ion pairs are found in *D. discoideum* NDK. Therefore, the additional interactions across the dimer interface probably contribute substantially to the stability of the reconstructed NDKs.

For the trimer interface interactions, fewer interprotomer nonpolar contacts exist in the reconstructed NDKs than in *D. discoideum* NDK. The trimer interface of *D. discoideum* NDK contains 12 interprotomer nonpolar contacts, whereas there are only six and three in the trimer interfaces of Arc1 and Bac1, respectively. Conversely, the reconstructed proteins have increased numbers of interprotomer ion pairs in the trimer interface (Table S1). The NDK trimers contain a cavity parallel to the threefold axis (Fig. S7*B*). One positively and two negatively charged side chains are directed toward the center of the cavity and toward each other. Specifically, the side chain of Arg27 from one Bac1 protomer forms two ion pairs with the side chains of Glu20 (4.23-Å separation) and Asp104 (4.12-Å separation) from a neighboring protomer. In Arc1, the side chain of Arg27 forms an ion pair with the side chain of Glu20 (3.96-Å separation) from an adjacent protomer. In *D. discoideum* NDK, a lysine (Lys34) is found in the same position as Arc1 Arg27; however, the side-chain amino group of Lys34 is not within ion-pairing distance of a negatively charged side chain (the Glu27 carboxyl is removed by 6.48 Å and the Asp111 carboxyl by 7.59 Å) (Fig. S7*B*). Possibly, therefore, the enhanced stability of the reconstructed NDKs is partly a consequence of the interprotomer ion pairs within the central cavity.

SI Methods

Crystallization. Arc1 and Bac1 were crystallized at 20 °C using the sitting-drop vapor-diffusion method. Each purified protein in 20 mM Tris·HCl, pH 7.4, and 20 mM KCl was added to an equal volume of reservoir solution [0.1 M Hepes, pH 7.5, 0.8 M potassium sodium tartrate tetrahydrate, 3% (wt/vol) (±)-2-methyl-2,4-pentanediol (for Arc1) or 0.1 M sodium citrate, pH 6.5, 1.4 M sodium acetate trihydrate (for Bac1)]. The Arc1 and Bac1 crystals had final dimensions of 0.1 × 0.1 × 0.1 mm and 0.2 × 0.2 × 0.2 mm, respectively.

Data Collection and X-Ray Crystallography of Arc1 and Bac1. The crystals were cooled in a stream of nitrogen at 95 K after being rapidly transferred individually into separate reservoir solutions containing 25% (wt/vol) ethylene glycol. Diffraction data were collected on the BL5A beamline at the Photon Factory and were processed using HKL2000 (12). Evaluation of the Matthews coefficient (13) indicated that each crystal contained two molecules

per asymmetric unit: $V_M = 3.58$ Å^3·Da^{-1} (Arc1), and $V_M = 3.49$ Å^3·Da^{-1} (Bac1). The structures were solved by molecular replacement using MOLREP (14) in the CCP4 suite (15), and the coordinates of *Pyrococcus horikosii* NDK [Protein Data Bank (PDB) code, 2cwk] and *T. thermophilus* NDK (PDB code, 1wkl) as the initial search models for Arc1 and Bac1, respectively. Model building was performed with XFIT in the XtalView package (16), and refinement was performed with Refmac5 (17) and CNS (18). Data collection and refinement statistics are summarized in Table S2. In the Ramachandran plots, 100% of the residue ϕ-φ values are in allowed conformational space for both Arc1 and Bac1. Although the crystallographic asymmetric units of Arc1 and Bac1 each contain two NDK molecules, hexameric structures could be generated by rotation around a crystallographic threefold axis. All images of the protein structures were drawn using PyMOL (www.pymol.org). The crystal structures of the resurrected ancestral enzymes have been deposited in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (www.rcsb.org) under the accession numbers 3VVT (Arc1) and 3VVU (Bac1).

1. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW (2007) Crystal structure of an ancient protein: Evolution by conformational epistasis. *Science* 317(5844):1544–1548.
2. Misra G, et al. (2009) Crystal structure of the *Bacillus anthracis* nucleoside diphosphate kinase and its characterization reveals an enzyme adapted to perform under stress conditions. *Proteins* 76(2):496–506.
3. Webb PA, Perisic O, Mendola CE, Backer JM, Williams RL (1995) The crystal structure of a human nucleoside diphosphate kinase, NM23-H2. *J Mol Biol* 251(4):574–587.
4. Moréra S, et al. (1994) Refined X-ray structure of *Dictyostelium discoideum* nucleoside diphosphate kinase at 1.8 A resolution. *J Mol Biol* 243(5):873–890.
5. Janin J, et al. (2000) Three-dimensional structure of nucleoside diphosphate kinase. *J Bioenerg Biomembr* 32(3):215–225.
6. Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* 65(1):1–43.
7. Cambillau C, Claverie JM (2000) Structural and genomic correlates of hyperthermostability. *J Biol Chem* 275(42):32383–32386.
8. Cheung YY, et al. (2005) Crystal structure of a hyperthermophilic archaeal acylphosphatase from *Pyrococcus horikoshii*—structural insights into enzymatic catalysis, thermostability, and dimerization. *Biochemistry* 44(12):4601–4611.
9. Dalhus B, et al. (2002) Structural basis for thermophilic protein stability: Structures of thermophilic and mesophilic malate dehydrogenases. *J Mol Biol* 318(3):707–721.
10. Arnott MA, Michael RA, Thompson CR, Hough DW, Danson MJ (2000) Thermostability and thermoactivity of citrate synthases from the thermophilic and hyperthermophilic archaea, *Thermoplasma acidophilum* and *Pyrococcus furiosus*. *J Mol Biol* 304(4): 657–668.
11. Kirino H, et al. (1994) Hydrophobic interaction at the subunit interface contributes to the thermostability of 3-isopropylmalate dehydrogenase from an extreme thermophile, *Thermus thermophilus*. *Eur J Biochem* 220(1):275–281.
12. Otwinowski Z, Minor W (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode Methods. *Enzymology 276, Macromolecular, Crystallography, part A*, eds Carter CW, Jr., Sweet RM (Academic, New York), pp 307–326.
13. Matthews BW (1968) Solvent content of protein crystals. *J Mol Biol* 33(2):491–497.
14. Vagin AA, Isupov MN (2001) Spherically averaged phased translation function and its application to the search for molecules and fragments in electron-density maps. *Acta Crystallogr D Biol Crystallogr* 57(Pt 10):1451–1456.
15. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5):760–763.
16. McRee DE (1999) XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* 125(2-3):156–165.
17. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53(Pt 3):240–255.
18. Brünger AT, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54(Pt 5): 905–921.

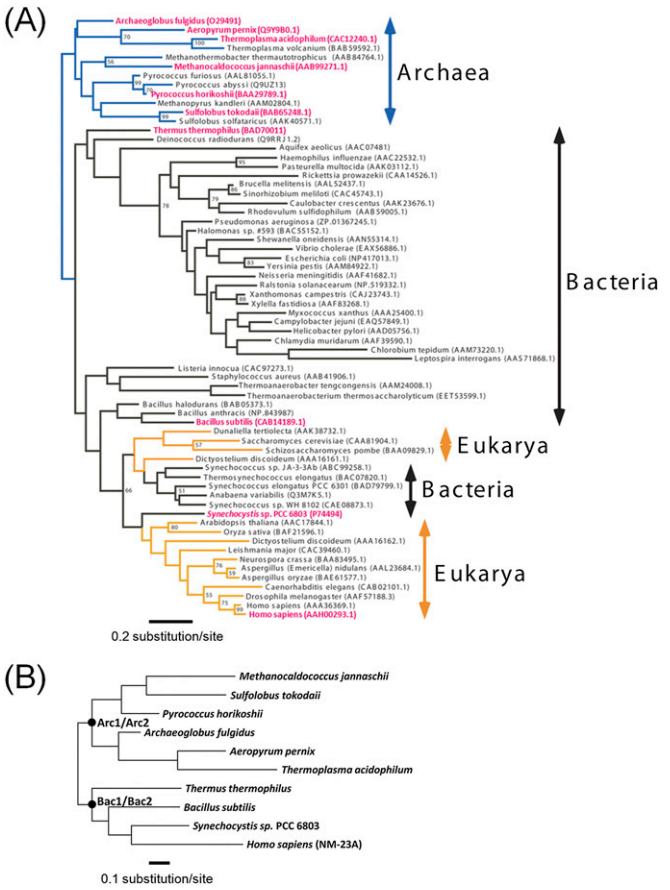

Fig. S1. Phylogenetic tree built with extant NDK sequences. (*A*) An unrooted tree constructed from 66 NDK sequences using the maximum likelihood (ML) method. The number at each node indicates the percentage bootstrap probability associated with the correct positioning of the node. The bootstrap probabilities were obtained from 100 pseudoreplicates generated by TREEFINDER. The accession number for each species obtained from GenBank is shown in parentheses. Blue, black, and yellow lines identify the archaeal, bacterial, and eukaryotic lineages, respectively. The 10 sequences that represent the main branches of the tree are shown in magenta. (*B*) The phylogenetic tree used to infer the amino acid sequences of the ancestral archaeal and bacterial NDKs. The nodes corresponding to the last common archaeal and bacterial ancestors are identified.

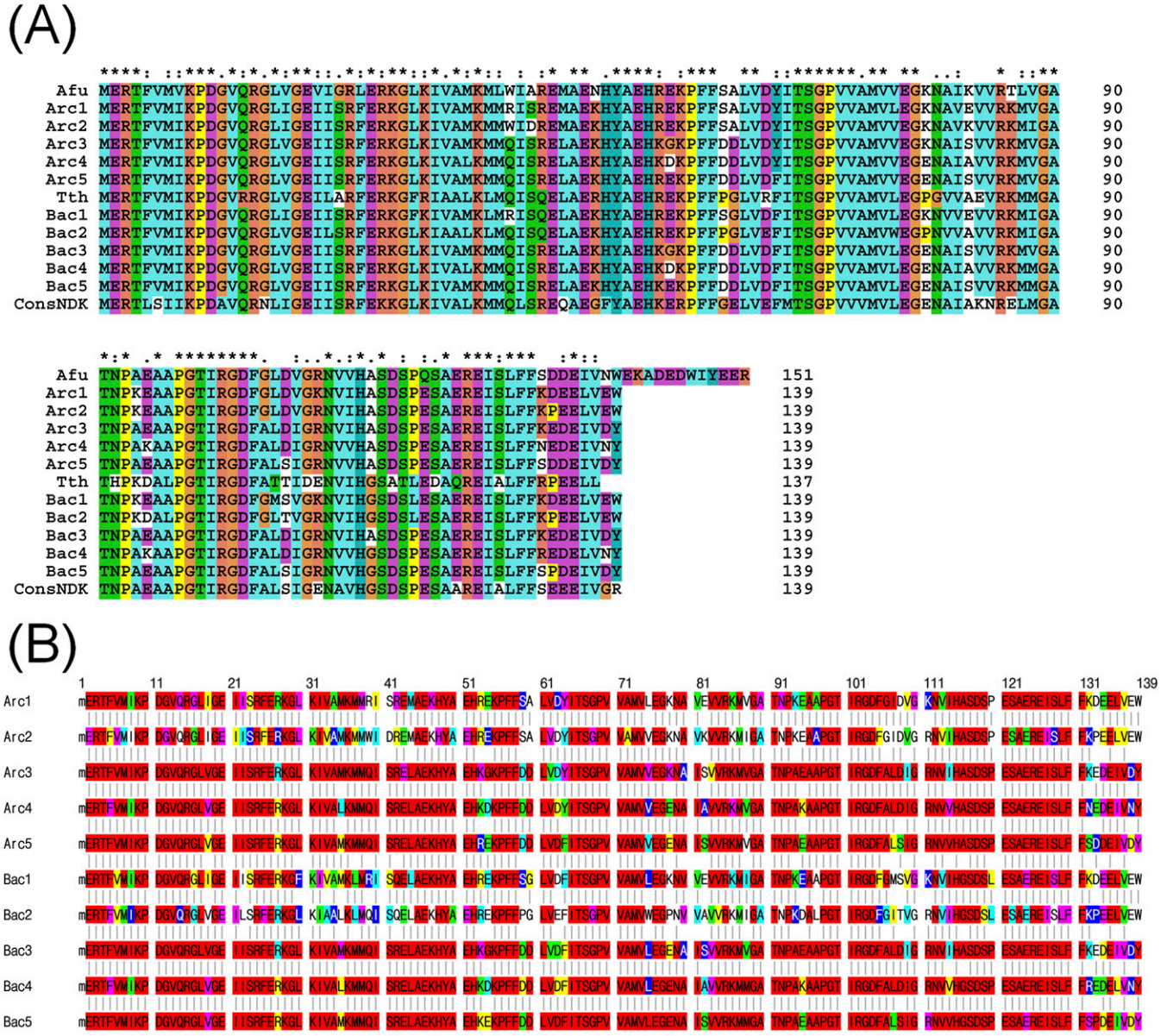

Fig. S2. The amino acid sequences of the ancestral NDKs reconstructed in this study. (*A*) A multiple amino acid sequence alignment for two contemporary NDKs and the ancestral proteins. Afu, *Archaeoglobus fulgidus* NDK; Tth, *Thermus thermophilus* NDK. The sequence of ConsNDK contains the amino acid most frequently found at each position. (*B*) Inferred ancestral amino acid sequences and their site-specific probabilities. The N-terminal methionines were not included in the ancestral sequence analysis. The reliability that the ancestral residues are correctly inferred is expressed as their a posteriori probability (p): red, p > 0.95; magenta, 0.95 ≥ p >0.90; yellow, 0.90 ≥ p >0.80; green, 0.80 ≥ p >0.70; cyan, 0.70 ≥ p >0.60; blue, 0.60 ≥ *p* >0.50; white 0.50 ≥ p.

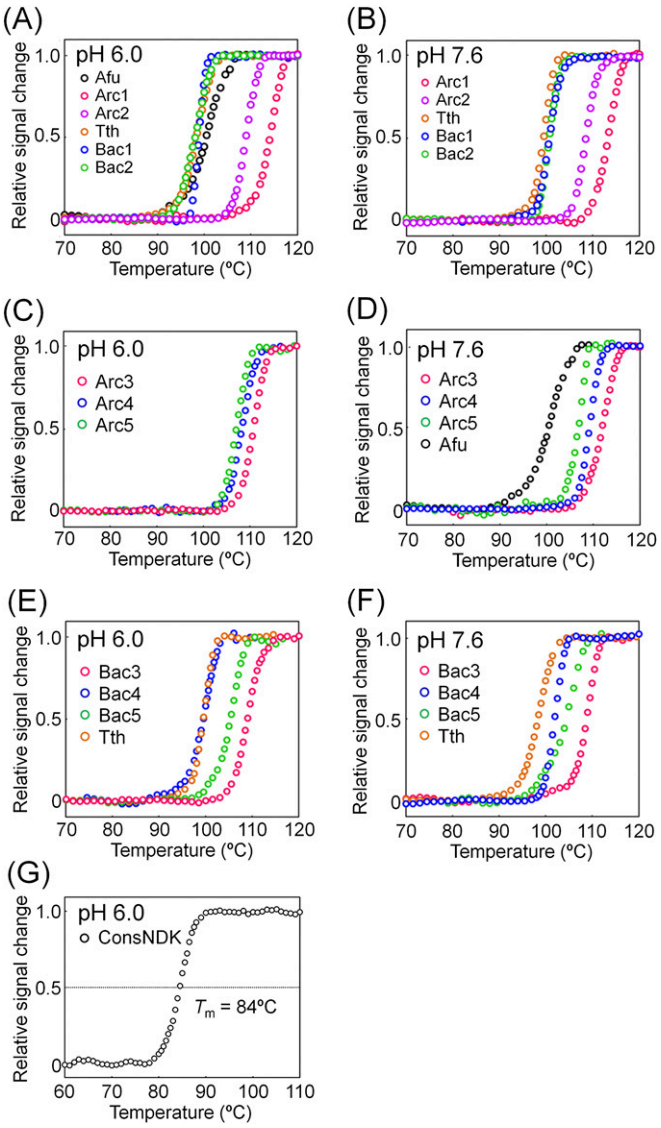

Fig. S3. Unfolding of NDKs monitored at θ_{222}. Each enzyme (20 μM) was dissolved in 20 mM potassium phosphate, pH 6.0 (*A*, *C*, *E*, and *G*) or pH 7.6 (*B*, *D*, and *F*), 50 mM KCl, 1 mM EDTA. The scan rate was 1.0 °C/min. Identical melting profiles were produced within experimental error for duplicated measurements. The plots were normalized with respect to the baselines of the native and denatured states. Afu, *A. fulgidus*; Tth, *T. thermophilus*.

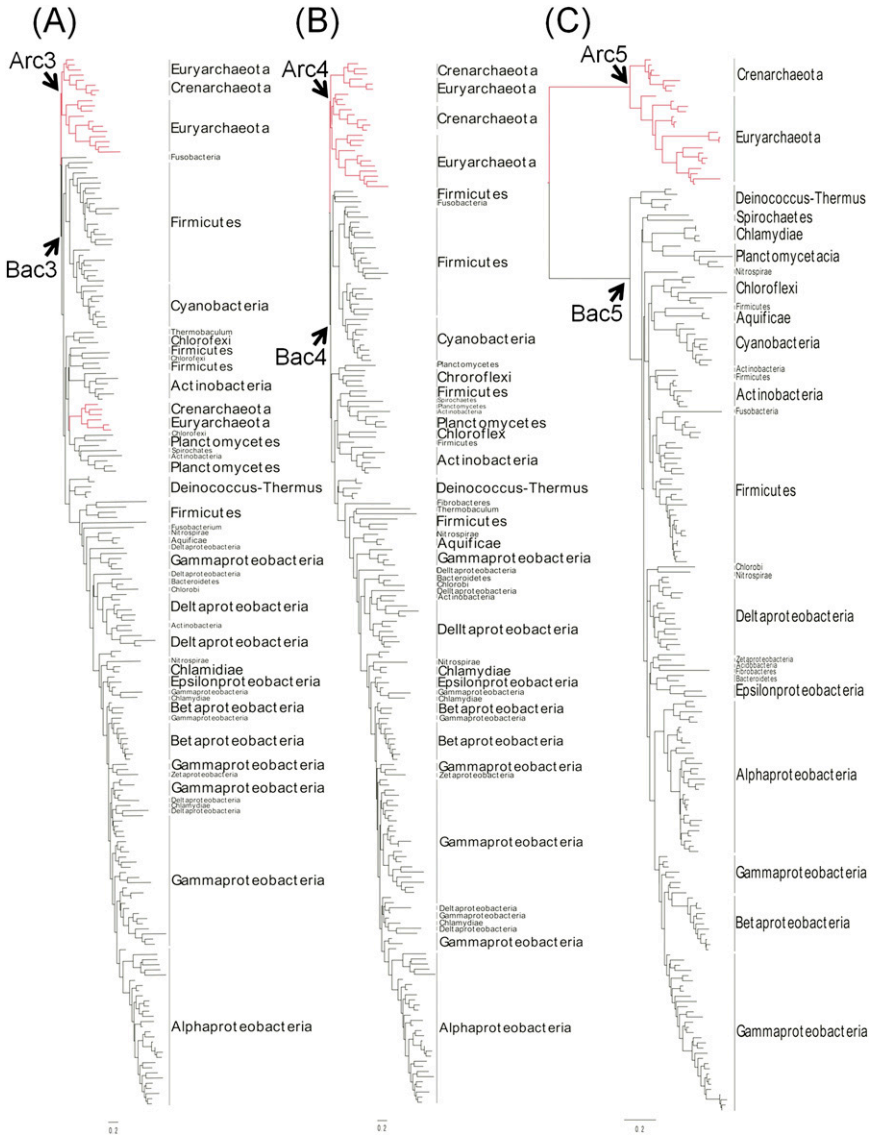

Fig. S4. Phylogenic trees containing the 204 sequences used to reconstruct the ancestral NDK sequences. (*A*) The ML tree built without constraints. (*B*) The ML tree built with the constraint that Archaea and Bacteria form separate monophyletic groups. (*C*) The ML tree built from a tree containing the small-subunit rRNA sequences from the same species.

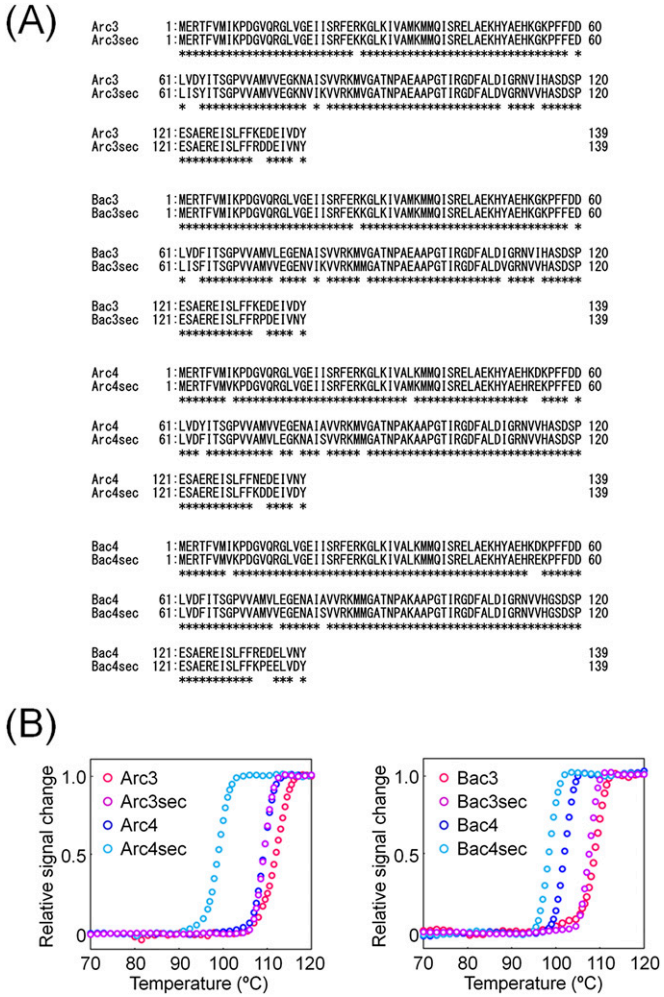

Fig. S5. The proteins containing the "second-best" ancestral residues. (*A*) Pair-wise amino acid sequence alignments for the inferred ancestral NDKs and the corresponding variants that contain the second-best ancestral residues at positions for which the prediction was uncertain (a posteriori probabilities <0.8). (*B*) The effects of uncertainty on the ancestral sequence reconstructions assessed by thermostability measurements. The experiments were performed as described for Fig. S3, with the pH of the solutions being 6.0.

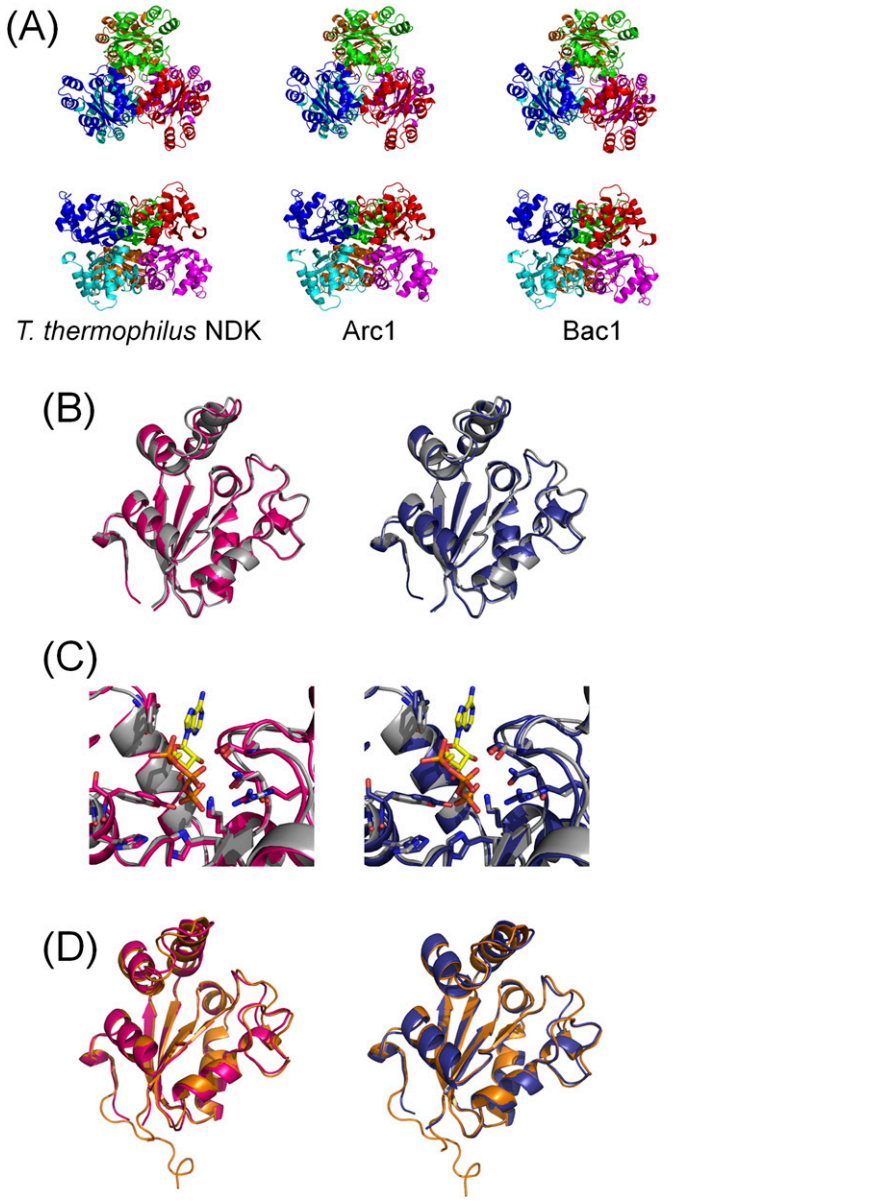

Fig. S6. Crystal structures of *T. thermophilus* and the resurrected ancestral NDKs. (*A*) The hexameric-type structure shown in the figure for Arc1 (*Center*) and Bac1 (*Right*) has been found for extant NDKs, including that from *T. thermophilus* (*Left*). For clarity, each protomer within a hexameric structure is shown in a different color. Upper panels are views along the threefold axis of rotation, and the lower panels are views along the twofold axis of rotation. (*B*) Comparison of the structures of the resurrected ancestral and *T. thermophilus* NDKs. (*Left*) Superpositioned Arc1 (magenta) and *T. thermophilus* NDK (gray; PDB code 1wkj). (*Right*) Superpositioned Bac1 (blue) and *T. thermophilus* NDK (gray). Superpositioning was performed by PDBeFold (www.ebi.ac.uk/msd-srv/ssm/). (*C*) Comparison of the ancestral and *T. thermophilus* NDK active sites. The *T. thermophilus* NDK structure (colored gray; PDB code 1wkl) was solved with an ATP bound in the active site. (ATP is shown as a stick model in yellow and orange.) The crystal structures of Arc1 (*Left*) and Bac1 (*Right*) (colored magenta and blue, respectively) were solved without substrate or product present. The side chains of the active-site residues are shown as stick models. Calculation of the C_α rmsd of the active-site residues K9, H48, Y49, F57, T91, R102, N112, and H115 was performed by the Swiss PDB Viewer (http://spdbv.vital-it.ch/). (*D*) (*Left*) Superpositioned Arc1 (magenta) and *Dictyostelium discoideum* NDK (orange). (*Right*) Superpositioned Bac1 (blue) and *D. discoideum* NDK (orange).

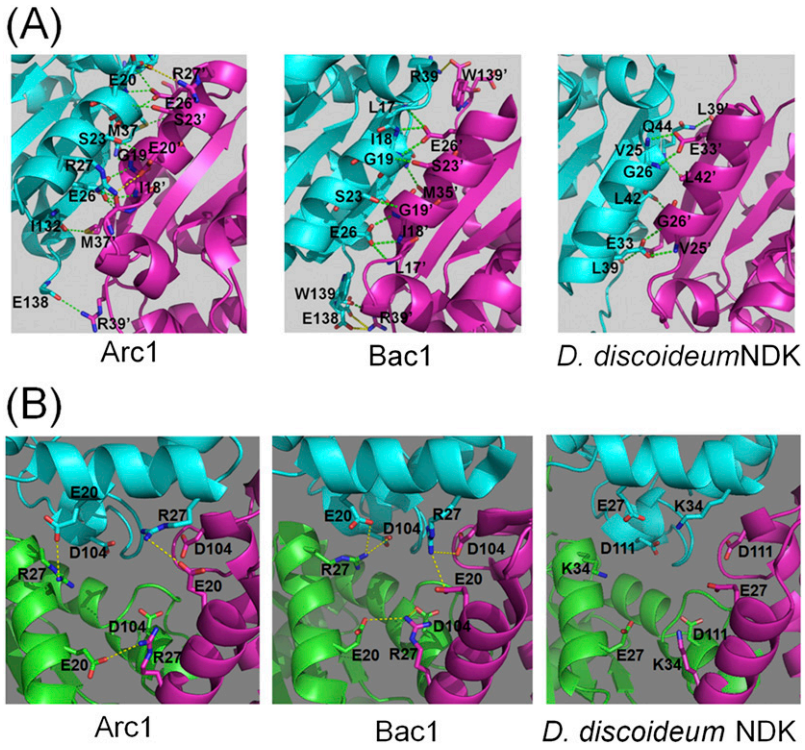

Fig. S7. Comparison of the protomer interfaces. (*A*) Interprotomer polar interactions at the dimer interface. Residues involved in interprotomer electrostatic interactions are shown as stick models. The yellow and green dotted lines indicate the interprotomer ion pairs and hydrogen bonds, respectively. (*B*) The central cavity within the trimer structure of Arc1, Bac1, and *D. discoideum* NDK. The three sets of acidic/basic side chains directed toward the center of each cavity are shown as stick models. Ion-pair separations of ≤6 Å are indicated by the yellow dotted lines.

Table S1. Comparison of factors known to be responsible for a protein's stability

Factor	Arc1	Bac1	*D. discoideum* NDK
Total accessible surface area (ASA) (%)	32,200 Å^2 (100)	35,000 Å^2 (100)	34,500 Å^2 (100)
Polar ASA (%)	15,400 Å^2 (48)	17,000 Å^2 (49)	14,100 Å^2 (41)
Nonpolar ASA (%)	16,800 Å^2 (52)	18,000 Å^2 (51)	20,400 Å^2 (59)
Number of intraprotomer ion pairs (separation <6 Å)	22	20	18
Dimer contact data (per interface)			
Number of nonpolar contacts (separation <5 Å)	22	21	16
Number of hydrogen bonds	14	11	8
Number of ion pairs (separation <6 Å)	2	3	0
Trimer contact data (per trimer)			
Number of nonpolar contacts (separation <5 Å)	6	3	12
Number of hydrogen bonds	12	15	21
Number of ion pairs (separation <6 Å)	9	12	6

Table S2. Crystallographic data collection and refinement statistics

Data collection	Arc1	Bac1
Space group	$P2_13$	$P2_13$
Unit cell dimensions (Å)	$a = b = c = 110.40$	$a = b = c = 109.35$
Resolution (Å)	20–2.40 (2.46–2.40)	20–2.40 (2.46–2.40)
Completeness (%)	100 (97.6)	100 (100)
R_{merge} (%)	7.1 (53.9)	5.8 (22.3)
Refinement		
Resolution (Å)	20–2.40	20–2.40
No. reflections	17715	17316
R/R_{free} (%)	18.7/22.2	18.0/22.1
No. atoms		
Protein	2,190	2,190
Water	42	61
B-factors		
Protein	34.48	25.94
Water	32.40	29.81
Rmsd		
Bond length (Å)	0.025	0.026
Bond angles (°)	1.986	1.983
PDB ID code	3VVT	3VVU

Values in parentheses are for the highest-resolution shell.