On the Origin and Evolution of Thermophily: Reconstruction of Functional Precambrian Enzymes from Ancestors of *Bacillus*

Joanne K. Hobbs,1 Charis Shepherd,2 David J. Saul,3 Nicholas J. Demetras,1 Svend Haaning,1 Colin R. Monk,1 Roy M. Daniel,1 and Vickery L. Arcus1,*

1Department of Biological Sciences, University of Waikato, Hamilton, New Zealand
2Department of Medicine, Dunedin School of Medicine, University of Otago, Dunedin, New Zealand
3ZyGEM Corporation Ltd., Hamilton, New Zealand

*Corresponding author: E-mail: varcus@waikato.ac.nz.
Associate editor: Andrew Roger

Abstract

Thermophily is thought to be a primitive trait, characteristic of early forms of life on Earth, that has been gradually lost over evolutionary time. The genus *Bacillus* provides an ideal model for studying the evolution of thermophily as it is an ancient taxon and its contemporary species inhabit a range of thermal environments. The thermostability of reconstructed ancestral proteins has been used as a proxy for ancient thermal adaptation. The reconstruction of ancestral "enzymes" has the added advantages of demonstrable activity, which acts as an internal control for accurate inference, and providing insights into the evolution of enzymatic catalysis. Here, we report the reconstruction of the structurally complex core metabolic enzyme LeuB (3-isopropylmalate dehydrogenase, E.C. 1.1.1.85) from the last common ancestor (LCA) of *Bacillus* using both maximum likelihood (ML) and Bayesian inference. ML LeuB from the LCA of *Bacillus* shares only 76% sequence identity with its closest contemporary homolog, yet it is fully functional, thermophilic, and exhibits high values for $k_{cat}$, $k_{cat}/K_m$, and $\Delta G^\ddagger$ for unfolding. The Bayesian version of this enzyme is also thermophilic but exhibits anomalous catalytic kinetics. We have determined the 3D structure of the ML enzyme and found that it is more closely aligned with LeuB from deeply branching bacteria, such as *Thermotoga maritima*, than contemporary *Bacillus* species. To investigate the evolution of thermophily, three descendents of LeuB from the LCA of *Bacillus* were also reconstructed. They reveal a fluctuating trend in thermal evolution, with a temporal adaptation toward mesophily followed by a more recent return to thermophily. Structural analysis suggests that the determinants of thermophily in LeuB from the LCA of *Bacillus* and the most recent ancestor are distinct and that thermophily has arisen in this genus at least twice via independent evolutionary paths. Our results add significant fluctuations to the broad trend in thermal adaptation previously proposed and demonstrate that thermophily is not exclusively a primitive trait, as it can be readily gained as well as lost. Our findings also demonstrate that reconstruction of complex functional Precambrian enzymes is possible and can provide empirical access to the evolution of ancient phenotypes and metabolisms.

Key words: ancestral reconstruction, isopropylmalate dehydrogenase, *Bacillus*, thermophily, maximum likelihood, Bayesian.

Introduction

The presence of thermophilic organisms on the branches closest to the root in the universal tree of life led to the long-standing hypothesis that thermophily is a primitive trait exhibited by the first forms of life on Earth (Woese 1987; Pace 1991; Stetter 1996). This theory suggests that contemporary thermophiles are directly descended from ancient thermophilic organisms rather than a result of more recent evolutionary flexibility such that the catabolic and anabolic processes continue to function as the organism adapts to new environments. The evolution of the biophysical properties of the core metabolic enzymes is a central component of this adaptation.

In the study of past evolutionary events that have shaped modern proteins, ancestral sequence reconstruction (ASR) has proved an invaluable tool (reviewed by Harms and Thornton 2010). ASR uses "genetic souvenirs" stored in the sequences of extant proteins, and the phylogenetic relationships between them, to trace their evolutionary relationships and infer the sequences of their ancestors. It has been used to reconstruct a growing number of binding proteins, including hormone receptors (Li et al. 2005; Bridgham et al. 2006, 2009, 2010), visual pigments (Chang et al. 2002; Ugalde et al. 2004; Chinen et al. 2010).
2005; Yokoyama et al. 2008; Field and Matz 2010), carbohydrate-binding proteins (Konno et al. 2007), and elongation factors (Gaucher et al. 2003, 2008). In contrast, few ancestral enzymes have been reconstructed (Malcolm et al. 1990; Stackhouse et al. 1990; Jermann et al. 1995; Chandrasekharan et al. 1996; Zhang and Rosenberg 2002; Thomson et al. 2005; Perez-Jimenez et al. 2011), and those that have are structurally simple and/or evolutionarily young. The demanding structural requirements for enzyme activity mean that errors in inference will likely result in an inactive ancestral enzyme or one that exhibits biologically unrealistic properties. Thus, enzyme activity can both act as an internal control for accurate ancestral inference and provide insight into the evolution of catalysis and metabolism.

There are a number of different methods that can be used to perform ASR. The original maximum parsimony method has now been superseded by the maximum likelihood (ML) method, and all experimental reports of ancestral proteins from the last decade have used this method. However, the accuracy of ancestral sequences reconstructed by ML has been questioned, and the Bayesian approach has been advocated as a more accurate method (Krishnan et al. 2004; Hall 2006; Williams et al. 2006). Krishnan et al. (2004) reported that mitochondrial transfer RNA sequences inferred by the Bayesian method were theoretically more functional than those inferred by ML and warned that similar errors could be seen with reconstructed proteins. In agreement with this, Williams et al. (2006) reported that the ML method overestimated the theoretical thermostability of ancestral proteins compared with the Bayesian method due to an inherent amino acid bias. In contrast to these two studies, Hanson-Smith et al. (2010) compared the accuracy of the ML and Bayesian protein inference methods computationally and reported no increase in accuracy when phylogenetic uncertainties were accounted for by the Bayesian method. Hall (2006) also compared the theoretical accuracies of the two methods and found that DNA sequences inferred by the Bayesian method were more accurate than those inferred by ML but that the opposite was true for inferred protein sequences (Hall 2006). To date, there are no examples of ancestral proteins reconstructed in the laboratory that have been inferred by the
Bayesian method, and the effect of the different inference methods on the accuracy of reconstructed proteins has only been evaluated computationally.

Here, we report the reconstruction of four ancestral sequences of increasing historical age for the structurally complex core metabolic enzyme LeuB. For these four enzymes, we have used the ML method for reconstruction. For comparison, we have also used the Bayesian method for reconstruction in two cases. We have measured the thermodynamic and kinetic properties of these enzymes to trace the origin and evolution of thermophilicity within the Bacillus genus.

**Materials and Methods**

**LeuB Sequences**
The majority of LeuB sequences used in this study were retrieved from GenBank. The Bacillus stearothermophilus leuB sequence was obtained from the B. (Geobacillus) stearothermophilus Genome Sequencing Project (http://www.genome.ou.edu/bstearo.html), and the gene from B. caldovelox was amplified and sequenced from genomic DNA using primers designed against the B. stearothermophilus gene. For B. psychrophilus and B. psychrosaccharolyticus, leuB sequences were determined by genomic-walking polymerase chain reaction as previously described (Cann et al. 1999; Reeves et al. 2000). Full details of the strains used and gene/protein accession numbers can be found in supplementary table S1 (Supplementary Material online).

**Phylogenetic Analysis and Node Age Estimates**
LeuB amino acid sequences were initially aligned using ClustalW2 (Larkin et al. 2007) and then refined manually using Gblocks version 0.3.3 (Drummond et al. 2011). The majority of the alignment was unequivocally resolved by ClustalW2, but some short regions required careful scrutiny and manual adjustment according to the JTT model of amino acid classification (Taylor and Jones 1993). A leuB nucleotide alignment was derived from the amino acid alignment and utilized in tests of positive selection (SLAC, FEL, and REL) performed in Datamonkey (Delport et al. 2010). ProtTest (Abascal et al. 2005) was used to determine the appropriate model of evolution (LG + I + G), and this model was implemented in GARLI version 1.0 (Zwickl 2010) to find the best ML tree. Bootstrapping was also performed in GARLI using 1,024 replicates. Node age estimates were made using the ML branch lengths, and two calibration points taken from the literature (Battistuzzi et al. 2004). These calibration points are taken from a robust prokaryotic phylogeny based on the concatenation of 32 protein sequences, and molecular divergence times estimated from geologic calibration points (Battistuzzi et al. 2004). The two calibration points relevant to this study are the point of divergence of the Bacillus and Clostridium (2,650 My) and the point of divergence of B. subtilis and B. halodurans (950 My). The program r8s version 1.71 (Sanderson 2003) was used to fix the two calibration points and estimate the ages of the remaining nodes; the resulting tree was visualized in FigTree version 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

**Ancestral Reconstruction**
Three different methods of ancestral inference (DNA, codon, and amino acid inference) were performed under the ML criterion in PAML version 4.3 (Yang 2007) using the phylogeny shown in figure 1. For nucleotide inferences in BASEML, the REV/general time reversible (GTR) nucleotide substitution rate model was used as this was chosen as the appropriate model of evolution by jModelTest version 0.1.1 (Posada 2008). For codon and amino acid inference in CODEML, the Jones amino acid rate file was employed. The sequences inferred by the three methods were compiled and any ambiguous sites were resolved using the following criteria: 1) bias was toward an amino acid that appeared at that site in organisms from all temperature ranges, the rationale being that this residue would be unlikely to affect thermophily; 2) the position of the residue on the crystal structure of LeuB from B. coagulans (PDB no. 1V53) was identified and used to guide the selection of the amino acid least likely to affect protein structure; 3) the JTT model of amino acid classification was consulted and used to evaluate the potential phytosiderochemical effects of different amino acids. Bayesian inference was performed in MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003) as described by Hall (2006) using the leuB nucleotide alignment and the appropriate model of evolution (GTR).

**Gene Synthesis, Protein Expression, and Purification**
Contemporary and ancestral leuB genes were optimized for expression in Escherichia coli and synthesized by GENEART (Regensburg, Germany). Following cloning into pPROEX HTb (Invitrogen), recombinant proteins were expressed in E. coli DH5α for ~20 h with 1 mM isopropyl β-D-1-thiogalactopyranoside induction at 37 °C, with the exception of the B. psychrosaccharolyticus and B. subtilis enzymes, which were expressed at 18 °C. Proteins were purified to ≥95% purity by nickel affinity chromatography in Buffer A (50 mM sodium phosphate buffer pH 8.0, 300 mM NaCl) followed by size-exclusion chromatography on a S200 16/60 column (GE Healthcare) in Buffer B (20 mM potassium phosphate buffer pH 7.6) and stored at 4 °C. For purification of the B. psychrosaccharolyticus and B. subtilis enzymes, 5% glycerol (v/v) was included in all buffers to increase shelf life. Protein concentrations were determined using a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific), and extinction coefficients calculated using ProtParam (Gasteiger et al. 2005).

**Enzyme Activity Assays**
LeuB activity was measured by following the reduction of the cofactor, NAD, at 340 nm in a ThermoSpectronic Helios spectrophotometer equipped with a single-cell photometer/effect cuvette holder. The substrate, 3-isopropylmalate (IPM; Wako Pure Chemicals, Japan), was prepared in assay buffer (20 mM potassium phosphate buffer pH 7.6 supplemented with 0.3 M KCl and 0.2 mM MnCl₂) as previously described (Wallon et al. 1997). The Michaelis–Menten constants (kcat) and catalytic rate constants (kcat/Km) for each enzyme were determined at a range of temperatures. Thermoactivity profiles were determined by measuring enzyme activity at 2–3 °C intervals over
a 20–30 °C range in triplicate and calculating the initial rates. Due to substrate expense, these assays were performed using an IPM concentration equivalent to twice the \( K_m \) at the specific test temperature. \( T_{eq} \) was determined as described elsewhere (Lee et al. 2007; Daniel and Danson 2010).

### Thermostability and Unfolding Measurements
Thermostabilities of contemporary and ancestral enzymes were measured by differential scanning calorimetry (DSC) in a TA Instruments Nano DSC. Enzyme solutions of 1 mg/ml and a reference sample of 20 mM potassium phosphate buffer pH 7.6 were ramped at 1 °C/min from 20 to 90 °C. The peak in power input for each enzyme, representing the \( T_{m} \), was found using the accompanying NanoAnalyze software. The free energy of unfolding, \( \Delta G_{N\rightarrow U} \), was determined for each enzyme from urea unfolding rates. Each enzyme was added to a range of urea concentrations, dissolved in assay buffer containing twice the \( K_m \) of IPM, in a 10 mm pathlength cuvette pre-equilibrated to 37 °C. The fluorescence signal at emission 280 nm, excitation 330 nm, was monitored on a Hitachi F-7000 fluorescence spectrophotometer and a single exponential fitted to the data. The free energy of unfolding in the absence of urea was calculated elsewhere (MacDonald and Pozharski 2001).

### Crystallization and Structure Determination
Crystallization was performed using the hanging drop vapor diffusion method at 18 °C. Following extensive screening and optimization of crystallization conditions, crystals of ML ANC4 were obtained overnight in 200 mM diammonium hydrogen citrate pH 5.5, 2 mM MgSO4, 15% (w/v) polyethylene glycol 3350, and 4% (v/v) glycerol using protein at 40 mg/ml in 20 mM N-2-hydroxyethylpiperazine-N-2-ethanesulfonic acid buffer pH 7.6. Data collection was performed on flash-cooled crystals using a Rigaku RUH3R rotating anode, Mar345 image plate detectors with MSC Osmic optics, and Oxford 600 series cryosystems at the University of Auckland, New Zealand. All data were indexed and integrated with MOSFLM (Leslie 2006) and scaled and reduced using SCALA (Evans 2006). The ANC4 structure was determined by molecular replacement using Phaser (McCoy et al. 2005) with a CHAINSAW (Stein 2008) model of the LeuB monomer from B. coagulans (PDB no. 1V53) as the search model. This was followed by iterative cycles of manual model building with COOT (Emsley and Cowtan 2004) and further refinement using REFMAC (Murshudov et al. 1997) and PHENIX (Adams et al. 2010) (simulated annealing Cartesian) with non-crystallographic symmetry restraints. Structural comparisons and analyses were performed using PDBeFold (secondary structure matching) (Krissinel and Henrick 2004) and PDBePISA (Krissinel and Henrick 2007).

### Results

#### Bacillus Phylogeny and Ancestral Reconstruction of LeuB
In this study, we chose the large multidomain core metabolic enzyme LeuB (3-isopropylmalate dehydrogenase, IPMDH, EC 1.1.1.85) as the subject for reconstruction. This dimeric enzyme catalyzes the oxidative decarboxylation of isopropylmalate—the third step in the leucine biosynthesis pathway—in the presence of a divalent cation and the cofactor NAD. LeuB shows moderate sequence conservation across contemporary Bacillus species (32% sequence identity across all species) and, importantly, occupies a genomic region that shows little evidence of recombination (Didelot et al. 2010), which mitigates the risk of spurious phylogenetic inference (Arenas and Posada 2010).

As a basis for ancestral inference, a robust phylogeny of contemporary Bacillus species was determined based on LeuB sequences using the ML criterion (fig. 1). We were particularly interested in the phylogeny of the contemporary Bacillus species in relation to their optimal growth temperature, and LeuB sequences were not available for any psychrophilic species. Therefore, we sequenced the leuB genes from B. psychrosaccharolyticus and B. psychrophilus. Following five iterations, the phylogeny shown in figure 1 had the “best” log likelihood score. Its validity is supported by bootstrap percentages (fig. 1), and its topological congruence with previously reported trees constructed from core genes and 16S ribosomal RNA sequences (Alcaraz et al. 2010). Interestingly, the tree shows that species with similar growth temperatures are not necessarily closely related.

Four ancestral enzymes (ANC1–ANC4) were chosen for ML sequence reconstruction, each positioned progressively deeper in the phylogeny and further back in evolutionary time. ANC1 was considered an interesting subject for reconstruction as it is the last common ancestor (LCA) of contemporary thermophilic species and a contemporary psychrophilic (B. psychrosaccharolyticus). ANC4 represents LeuB from the LCA of Bacillus. The ancestral enzymes were dated using the ML branch lengths, and two calibration points taken from the literature (Battistuzzi et al. 2004). ANC1, 2, 3, and 4 are approximately 670, 820, 850, and 950 My old, respectively. The four ancestral sequences were inferred using the ML criterion and three different methods: DNA, codon, and amino acid inference. The three methods were in agreement for >85% of sites and for a further >10% the codon method (considered the most robust), and one other method was in agreement. Any remaining ambiguous sites were resolved using the criteria detailed in the Materials and Methods, and the average posterior probability at each site was >0.9 for all four reconstructed sequences. The final ancestral sequences differ from all of the contemporary LeuB enzymes by at least 59 amino acids (<84% sequence identity; supplementary table S2, Supplementary Material online), and the oldest enzyme, ANC4, shares only 76% sequence identity (i.e., differs by 88 amino acids) with its closest contemporary descendent.

Alternative versions of ANC2 and ANC4 were reconstructed using the Bayesian method. The average posterior probability for each nucleotide site in the ANC2 and ANC4 sequences was 0.84 and 0.82, respectively. At the amino acid level, the Bayesian versions of ANC2 and ANC4 differ from their ML counterparts by 6.8% and 9.8%, respectively, which equate to 25 and 36 amino acids (supplementary fig. S1, Supplementary Material online). For ANC2, 21 of
25 of these substitutions were unequivocally chosen by the Bayesian method, and the four remaining ambiguous sites had close to equal posterior probabilities for two amino acids that belong to the same JTT model classification group, for example, Lys and Arg. For ANC4, 35 of the 36 substitutions were unequivocally chosen by the Bayesian method; the single remaining site had equal posterior probability for two amino acids.

### Catalytic Activity and Thermoactivity of Contemporary LeuB Enzymes

In order to evaluate the biologically relevant properties of the reconstructed enzymes, catalytic activity and thermoactivity data were determined for LeuB enzymes from representative contemporary psychrophilic, mesophilic, and thermophilic *Bacillus* species (table 1 and fig. 2). Three contemporary LeuB enzymes, and the ancestral LeuB, were overexpressed to a high level in *E. coli*, purified by Ni-affinity chromatography and subsequent size-exclusion chromatography, and tested for activity (supplementary fig. S2, Supplementary Material online). For all enzymes, kinetic constants were first determined at a range of temperatures and then, once the optimum temperature for activity (*T*\textsubscript{opt}) had been identified, repeated at *T*\textsubscript{opt}. The Michaelis constants (*K*\textsubscript{M}), catalytic constants (*k*\textsubscript{cat}), and catalytic efficiencies (*k*\textsubscript{cat}/*K*\textsubscript{M}) shown in table 1 were determined at the respective *T*\textsubscript{opt} values for each enzyme. The kinetic parameters determined for the different contemporary enzymes are similar to each other and comparable with those reported previously (Wallon et al. 1997). The low *k*\textsubscript{cat} value for the *B. psychrosaccharolyticus* enzyme is common for psychrophilic enzymes, even when compared with other enzymes at their respective *T*\textsubscript{opt} values (Georlette et al. 2004). The high value of *K*\textsubscript{M} (NAD) for the *B. subtilis* enzyme may be a result of LeuB4, which is in proximity of the NAD-binding site and is substituted for a Pro in all the other *Bacillus* LeuB enzymes. This may be due to an error in the reported *B. subtilis* sequence and demonstrates the sensitivity of enzymatic parameters to potential sequence errors.

The thermoactivity profiles of the contemporary LeuB enzymes were determined by measuring their initial rate of activity at different temperatures over a 20–30 °C range (fig. 2). Initial rates were used as longer assay durations result in inconsistent *T*\textsubscript{opt} determinations (Daniel and Danson 2010). The thermoactivity data were also fitted to the Equilibrium model, which calculates an additional thermal parameter for each enzyme, *T*\textsubscript{eq} (Daniel and Danson 2010). *T*\textsubscript{eq} has been shown previously to be a better predictor of an organism's growth temperature than stability (Lee et al. 2007). The *T*\textsubscript{opt} and *T*\textsubscript{eq} values for the contemporary enzymes (table 1) correlate with the optimal growth temperatures of their host organisms, which are consistent with the correlations between protein thermostability and growth temperature and *T*\textsubscript{eq} and growth temperature reported previously (Gromiha et al. 1999; Lee et al. 2007). It is also typical for enzymes from psychrophilic and mesophilic organisms to exhibit *T*\textsubscript{opt} values higher than their host's optimal growth temperature (Georlette et al. 2004).

### Catalytic Activity, Thermoactivity, and Thermostability of ML Ancestral Enzymes

Following recombinant expression and purification (supplementary fig. S2, Supplementary Material online), all four ML reconstructed ancestral LeuB enzymes were found to be catalytically active, and their kinetic constants were determined (table 1). All four enzymes exhibited kinetic parameters comparable with their homologs from...
contemporary _Bacillus_ species. Interestingly, their thermoaolvity profiles (fig. 3A), and subsequent _T_\text{opt} and _T_\text{eq} values (table 1), show that they are adapted to very different thermal environments. The common ancestor of the _B. psychrosaccharolyticus_ and _B. caldovelox_ LeuB enzymes, ANC1, is more thermophilic than the _B. caldovelox_ enzyme (_T_\text{opt} of 73 vs. 69 °C) and has a broader temperature profile (figs. 2 and 3A). Going back in evolutionary time, there is then a sharp decline in thermophily to ANC2, which has a narrow psychrophilic/mesophilic activity profile (_T_\text{opt} of 49 °C), followed by a gradual increase in thermophilicity again from ANC2, through ANC3 to ANC4 (_T_\text{opt} of 60 and 70 °C, respectively). This trend in thermal adaptation was also confirmed by determining the midpoint of thermal unfolding (_T_\text{m}) for each enzyme (table 1 and fig. 3B). This is the biophysical parameter typically used to compare the structural stabilities of reconstructed ancestral enzymes (Malcolm et al. 1996; Jermann et al. 1995; Konno et al. 2007; Gaucher et al. 2008). The _T_\text{m} values are lower than the _T_\text{opt} and _T_\text{eq} values because the _T_\text{m} determinations were performed in the absence of substrate and cofactor, which have a stabilizing effect (data not shown).

**Catalytic Activity and Thermoactivity of Bayesian Ancestral Enzymes**

To test whether the temperature trend we observed with the ML ancestral enzymes was robust to the inference method used and compare the accuracy of the two inference methods, we chose to reconstruct the two ancestors most pivotal to the temperature trend (ANC2 and ANC4) using the Bayesian approach.

Both reconstructed Bayesian enzymes were found to be catalytically active, but their _K_\text{M} values for the substrate, and in the case of ANC4 for the cofactor also, were greatly increased compared with their ML counterparts and the contemporary LeuB enzymes. The Pro→Leu substitution, which we suspect may be the cause of the anomalous _K_\text{M} (NAD) for the _B. subtilis_ enzyme, is not present in the Bayesian ANC4 enzyme. Due to their increased Michaelis constants and the low _k_\text{cat} value of Bayesian ANC2, the Bayesian enzymes also exhibit much lower catalytic efficiencies than all the other enzymes (table 1). The kinetic parameters exhibited by the Bayesian enzymes are inconsistent with values for contemporary LeuB enzymes measured by us and others and suggest that in the case of _Bacillus_ LeuB ancestral sequences, the Bayesian versions contain a greater number of sequence errors than their ML counterparts. In addition, the prediction that the ML reconstruction process leads to an overestimation of thermophily for ancestral proteins (Williams et al. 2006) is not borne out by our results. The thermoactivity profile of Bayesian ANC4 (supplementary fig. S3, Supplementary Material online) reveals that it is thermophilic, with a _T_\text{opt} only 2 °C lower than its ML counterpart (table 1). In contrast, the _T_\text{opt} of Bayesian ANC2 is 14 °C higher than that of its ML counterpart (table 1). This suggests that the fluctuations in the temporal temperature trend are less pronounced than implied by the ML enzymes; however, the apparent sequence errors in the inferred Bayesian enzymes (by virtue of their high _K_\text{M} and low _k_\text{cat} and _k_\text{cat}/_K_\text{M} values) lead us to consider their thermoactivities with caution.

**Kinetic Unfolding of Contemporary and Ancestral Enzymes**

As an additional method to investigate the structural stability of the LeuB enzymes and compare the biophysical properties of the reconstructed ancestral enzymes with their contemporary counterparts, the free energy of unfolding was determined for each enzyme from urea unfolding rates (table 1). In general, the _ΔG_\text{N}^\ddagger _{N→U} Values for the contemporary and ancestral enzymes correlate with the _T_\text{opt} and _T_\text{m} data, that is, the most thermophilic enzymes have the highest _ΔG_\text{N}^\ddagger _{N→U} values. This is in agreement with previously published findings (Beadle et al. 1999; Wittung-Stafshede 2004). The exceptions to this trend are the Bayesian ancestors. Bayesian ANC2 and Bayesian ANC4 are thermophilic enzymes but have low _ΔG_\text{N}^\ddagger _{N→U} values (92.9 and 92.8 kJ mol\(^{-1}\), respectively), lower than even the contemporary psychrophilic LeuB (94.9 kJ mol\(^{-1}\)). This implies that although they are thermodynamically stable, they are kinetically unstable and would unfold quickly in a thermophilic environment. This is further evidence that the biophysical
properties of the Bayesian reconstructed enzymes are unreliable.

ML ANC4 fits the trend of being thermophilic and having a high kinetic barrier to unfolding, but its \( \Delta G_{N-U} \) is exceptionally high (110.8 kJ mol\(^{-1} \)). The slow rate of unfolding for ANC4 compared with ANC1, which has a similar \( T_m \) value to ANC4, is illustrated in figure 4A. A high \( \Delta G_{N-U} \) is not the only unusual property exhibited by ANC4; Figure 4B highlights the considerably higher catalytic activity of ANC4 compared with the other ML ancestral enzymes at a range of temperatures. These two properties may reflect the evolution of folding, structure, and function for this enzyme whereby, over evolutionary time, the optimization of folding and activity has been in competition (Fersht 1995).

Three-Dimensional Structure of ML LeuB from the Last Common Ancestor of Bacillus

To investigate the structural evolution of LeuB and compare the structure of an ancestral LeuB with its contemporary homologs, we used X-ray crystallography to determine the 3D structure of ML ANC4 (fig. 5). The data collection and refinement statistics for this structure can be found in supplementary table S3 (Supplementary Material online).
The ANC4 structure shows that, like contemporary LeuB enzymes, the ancestral enzyme is dimeric and has a similar topology and fold to the LeuB structure from *B. coagulans* (PDB no. 1V53). A sequence-independent structural comparison (using root mean square deviation; RMSD) between ANC4 and other structures from the isocitrate/isopropylmalate dehydrogenase-like fold (Murzin et al. 1995) showed that close structural homologs span the prokaryotic tree. Two close structural homologs are noteworthy: LeuB from the deeply branching organisms (Battistuzzi et al. 2004) *Acidithiobacillus ferrooxidans* (RMSD = 1.22 Å, 339 aligned residues) and *Thermotoga maritima* (RMSD = 1.35 Å, 341 aligned residues). Interestingly, ANC4 is more closely aligned with these two structures than the *B. coagulans* structure (RMSD = 1.44 Å, 336 aligned residues). These structural comparisons are reinforced by sequence comparisons for the *Bacillus* LeuB ancestors; the ANC1–ANC4 sequences move progressively closer to the *T. maritima* enzyme (59.8–62.5% sequence identity, respectively) despite the fact that the *T. maritima* sequence was not used in the ASR process. This sequence trend is also seen for ANC1–ANC4 and LeuB from the ancient bacterium *Aquifex aeolicus* (Battistuzzi et al. 2004).

**Discussion**

The genus *Bacillus* is one of the most studied genera of prokaryotes (Sonenshein et al. 2002), and its contemporary members are adapted to a diverse range of environmental temperatures, yet the origin and evolution of this extreme range of thermal adaptation are unknown. One of the aims of this study was to trace the origin and evolution of thermophily within this genus, and in the process, test two previously proposed hypotheses: that since the LCA, there has been a gradual loss of thermophily (Gaucher et al. 2008) and that contemporary thermophily is a primitive trait (Woese 1987; Pace 1991; Stetter 1996).

Initially, the thermophily of ANC4 appears to support the hypothesis that thermophily is a primitive trait; our data suggest that the LCA of *Bacillus* was thermophilic therefore, the origin of thermophily in the *Bacillus* genus is ancient. However, our analysis of the mechanisms of thermophily
exhibited by ANC1 and ANC4 shows that they are distinct, and therefore, the thermophily of ANC1 is new and did not originate from ANC4. The temporal trend of fluctuating thermostability exhibited by the progression from ML ANC1 to ANC4 (fig. 6) suggests that thermophily can be readily gained as well as lost, and the lack of a thermophilic signature passed down by ANC4 shows that thermophily is not strictly a primitive trait. Our data also add significant fluctuations to the broad trend in thermophily proposed by Gaucher et al. (2008), suggesting that the trend they observe is undersampled. There is a good correlation between the growth temperature of an organism and the thermostability of its proteins, both in this study and others previously (e.g., Gromiha et al. 1999); therefore, we are confident that the growth temperature of ancestral organisms can be reliably inferred from the thermostability of ancestral proteins. However, as the Bacillus genus demonstrates, contemporary organisms exhibit a wide range of optimal growth temperatures depending on their habitat; therefore, the thermostability of proteins from a single ancestral organism cannot be taken as an indicator of global temperature (Gaucher et al. 2008). Instead, we hypothesize that the fluctuations in thermophily we have observed likely reflect changes in the microenvironment encountered by the evolving Bacillus species. This hypothesis is supported by the speed at which we have observed changes in thermal adaptation, for example, ML ANC1 and ANC2 are separated by only ~150 my, yet their temperature optima differ by >20°C. Even though our ancestral enzymes span the climatically erratic Neoproterozoic era (Hyde et al. 2000), it is unlikely that the overall temperature of the Earth changed this much in such a short time.

Our data suggest that thermophily has evolved at least twice within the Bacillus genus via independent evolutionary paths. We acknowledge that this hypothesis is based on data from a single metabolic enzyme and intend to reconstruct other enzymes to strengthen our findings. This study also acts as a proof of concept with regard to the successful reconstruction of complex Precambrian enzymes. We have presented experimental evidence to support the validity of our inferred enzymes: they are functional, their kinetic constants are comparable with those of contemporary LeuB enzymes, and the structure of ML ANC4 is more similar to the structure of LeuB from ancient deeply branching bacteria than from contemporary Bacillus species, despite the absence of LeuB sequences from these deeply branching bacteria in the ASR process. ML ANC4 also exhibits

---

**Fig. 7.** Amino acid differences between ML ANC4 and ML ANC2 (A) and between ML ANC2 and ANC1 (B) highlighted on the ANC4 structure. The structure is depicted as a transparent surface with a cartoon representation beneath. Variable positions are shown in blue and surface-exposed residues also show a blue area at the surface. The active site is shown by placing isopropylmalate (orange sticks) and Mg$^{2+}$ (gray sphere) from the closely related Acidithiobacillus ferrooxidans structure (PDB no. 1A05).
some remarkable biophysical properties, in particular its unusually high catalytic constant and kinetic barrier to unfolding (table 1). The high thermoactivity and thermostability of ML ANC4, in addition to its favorable catalytic properties, highlight the potential use of ASR in the development of enzymes for biotechnology.

To our knowledge, this is the first experimental comparison between ancestral proteins reconstructed using the ML and Bayesian methods. At least in the case of LeuB enzymes from Bacillus, it appears that there are a greater number of errors in the inferred sequence using Bayesian methods. The Bayesian ancestors have $k_M$, $k_{cat}$, $K_M$, and $AG_{N→UV}$ values that are inconsistent with all other contemporary and ancestral LeuB enzymes reported here and elsewhere.

We have demonstrated the capacity of ASR to reconstruct functional, and structurally complex, enzymes from the Precambrian age. Our results significantly extend the reach of both the evolutionary time over which ASR of complex active enzymes is possible and the diversity of contemporary sequences from which ASR can be implemented. The reconstruction of other functional Precambrian enzymes will continue our experimental interrogation of the evolution of biophysical processes and adaption to changes in environmental temperature and will potentially allow experimental investigation of the evolution of ancient metabolism and protein-folding mechanisms.

**Supplementary Material**

Supplementary tables S1–S3 and figures S1–S4 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

This work was supported by the Marsden Fund of New Zealand. S.H. was supported by a Carlsberg Foundation postdoctoral fellowship. We would like to thank Moreland Gibbs of Macquarie University, Sydney, Australia, for providing B. psychrophilus genomic DNA. We acknowledge the Biomolecular Interaction Centre at the University of Canterbury, New Zealand, where the DSC was performed. The crystal structure of ML ANC4 has been deposited in the RSCB Protein Data Bank, PDB no. 3U1H.

**References**


Zwickl DJ. 2010. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. [PhD dissertation]. [Austin (TX)]: The University of Texas at Austin.